



30 anni di informatica archeologica – 3

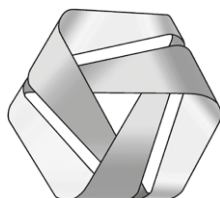
ANALYSE DES DONNÉES AND ARCHAEOLOGY FIFTY YEARS LATER: FROM DATA ANALYSIS TO DATA SCIENCE

François Djindjian and Paola Moscati

with

Lorenzo Cardarelli and Alessandro Di Ludovico





7

FUTURO
ANTERIORE

30 anni di informatica archeologica – 3

François Djindjian and Paola Moscati

with

Lorenzo Cardarelli and Alessandro Di Ludovico

ANALYSE DES DONNÉES
AND ARCHAEOLOGY FIFTY YEARS LATER:
FROM DATA ANALYSIS TO DATA SCIENCE



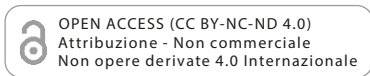
All'Insegna del Giglio

On the cover: detail of the figure on p. 96 and graphic rendering of a bronze mirror from Praeneste (by M. Bellisario). In the center, the logo of the 3rd International Symposium on Computing and Archaeology (Rome 1995).

ISSN 1723-4565

ISBN 978-88-9285-374-4

e-ISBN 978-88-9285-375-1



Futuro Anteriore
giugno 2025

All'Insegna del Giglio s.a.s.
via A. Boito, 50-52
50019 Sesto Fiorentino (FI)
www.insegnadelgiglio.it

Stampato a Sesto Fiorentino (FI)
BDprint, giugno 2025

TABLE OF CONTENTS

Prefazione	
<i>Tito Orlandi</i>	7
1. Introduction	
<i>François Djindjian, Paola Moscati</i>	9
2. The French school of data analysis	
<i>François Djindjian</i>	15
2.1 Introduction	15
2.2 Jean-Paul Benzécri and his community	17
2.3 The statistical methods of the French school of data analysis.	23
2.4 Fields of application.	25
2.5 The key phases of data analysis in France.	27
2.6 Conclusions	28
3. Correspondence analysis: a ‘neglected multivariate method’ in the international panorama?	
<i>Paola Moscati</i>	29
3.1 The statistical roots	29
3.2 The archaeological approach	34
3.3 Theoretical and practical developments of data analysis techniques	39
3.4 A literature-based analysis: the case of «Archeologia e Calcolatori»	41
4. Data analysis and prehistoric archaeology	
<i>François Djindjian</i>	47
4.1 The influence of data analysis on the evolution of archaeological methods	47
4.2 Data analysis as a cognitive process.	63
4.3 Teaching data analysis in archaeology	64
4.4 The post-modern reaction in archaeology.	65
5. Data analysis and ‘classical’ archaeology	
<i>Paola Moscati</i>	67
5.1 The quantitative approach and the data encoding strategy	67
5.2 Quantitative archaeology, data analysis, and ancient pottery	69
5.3 ‘Automatization of Etruscan corpora’: a pioneering application of data analysis in the archaeology of pre-Roman Italy	80

6.	Textual analysis in archaeology: proposals, methods, and applications	
	<i>Alessandro Di Ludovico</i>	89
6.1	Textual analysis: assumptions and needs of archaeological research	89
6.2	Textual analysis and archaeology: premises, beginnings and developments	90
6.3	Textual analysis and the French school of data analysis: the investigation of visual languages	101
6.4	Textual analysis for methodological and meta-archaeological research	104
6.5	Closing remarks	108
7.	The legacy: the Data Science movement	
	<i>Lorenzo Cardarelli</i>	109
7.1	The leading actors: Data Science, Machine Learning and Deep Learning	109
7.2	Not only Deep Learning: the role of manifold learning and dimensionality reduction	113
7.3	A practical example: UMAP and archaeological ceramics.	121
7.4	Conclusions and perspectives.	121
8.	References	123

Tito Orlandi

PREFAZIONE

L'archeologia e la linguistica sono state le discipline più studiate agli inizi delle applicazioni dell'informatica agli studi umanistici. Ma l'archeologia ha questo di particolare, che la rivoluzione informatica è sopravvenuta su di essa in un momento in cui era sottoposta per conto suo ad una crisi ad un tempo di obiettivi e di metodi. L'archeologia aveva già conosciuto fra Sette ed Ottocento una trasformazione da ricerca estetica e nostalgica delle vestigia delle antiche civiltà ad una scienza che intendeva ricavare memorie e informazioni storiche dai monumenti che si venivano cercando e portando alla luce. Alla metà del Novecento da un lato si ampliava l'oggetto delle ricerche al periodo preistorico e alle civiltà o culture che non avevano lasciato documenti scritti o i cui documenti non erano decifrabili; dall'altro si cercava con metodi matematici di trovare un significato alla documentazione materiale estraendone rapporti di differenza e di somiglianza. Si deve ricordare che Binford aveva proposto le sue teorizzazioni di analisi matematiche dei dati del tutto indipendentemente dall'utilizzo di macchine informatiche, o computer, per eseguire i calcoli.

L'evoluzione degli studi in informatica umanistica, e in particolare nelle applicazioni all'archeologia, presenta delle caratteristiche contraddittorie e in qualche modo misteriose. Da un lato i più recenti interessi di ricerca appaiono dimenticare, anzi negare quelle che pochi decenni prima sembravano le linee conduttrici della riflessione su quei temi. Dall'altro, nello svolgimento delle discussioni che accompagnano i lavori e le tecniche concrete, è forte l'impressione che ritornino sempre gli stessi temi principali, e gli stessi problemi, che non si riesce a superare o comunque a inquadrare in modo soddisfacente.

L'influenza del progresso della tecnica, con la sempre più ampia disponibilità di macchine e potenza di calcolo, coniugata con la facilità di connessione e comunicazione in rete, invita a ripetere i procedimenti tradizionali in modo automatico e dunque moltiplicato ad un livello prima non immaginabile. Tuttavia sembra si voglia evitare l'automazione dei ragionamenti, che potrebbe aiutare nella valutazione dei risultati delle analisi ottenute dai cosiddetti big data. L'analisi statistica, oggetto della ricapitolazione storica che è al centro dell'attenzione degli autori di questo libro, sembra stare a metà fra le due strade, in quanto cerca di indirizzare le operazioni statistiche nel modo più conforme ad ottenere la comprensione storica degli oggetti presi in considerazione.

Paola Moscati è stata un precursore, a contatto con i grandi che soprattutto in Francia ed Inghilterra guidavano le ricerche sull'analisi statistica dei dati, e poi un tenace promotore di quelle generali ricerche in Italia, soprattutto con la rivista «Archeologia e Calcolatori». A François Djindjian va il merito di aver diffuso per primo i metodi dell'*analyses des données* nella ricerca archeologica, con particolare riferimento agli studi preistorici. Con questo libro, insieme ad esperti collaboratori, assolvono al benemerito compito di ripercorrerne lo sviluppo a futura memoria.

François Djindjian, Paola Moscati

1. INTRODUCTION

Le langage mathématique et l'outil informatique pénètrent aujourd'hui dans toutes les disciplines. Engouement sincère mais peu éclairé; poudre aux yeux; progrès méthodologique...: il est impossible de tout trier et de tout juger! D'ailleurs la parabole de l'ivraie et du bon grain nous avertit qu'il serait prématuré d'arracher du champ de la science, les herbes folles. Cependant entre les modèles mathématiques qui conservent les formules et l'enchaînement déductif des théories pures pour les ajuster non sans peine au réel, et les spectaculaires démonstrations de banques d'images appelées d'un trait de stylet sur un écran cathodique, on sait que nous avons choisi la voie de l'analyse des données (BENZÉCRI 1977b, 193).

The impact of data analysis on archaeology, which transformed scholarship more than 50 years ago, has become a topic of historiographical examination today. Most of the pioneers of this scientific journey are now retired, and many have unfortunately passed away. That is why this book is also valuable for future scholarly research, as its authors have actively participated in this ongoing evolution. Indeed, they represent four generations of archaeologists who have embraced an interdisciplinary quantitative approach to enhance their research practices, in the period spanning from the 1970s to the present.

The concept of data analysis is broad and somewhat vaguely defined. In archaeology, it is generally accepted that data analysis was first introduced in 1966, simultaneously in the UK, influenced by SOKAL and SNEATH's (1963) book *Numerical Taxonomy* (see in particular HODSON, SNEATH, DORAN 1966; DORAN, HODSON 1966), and in the US, with the application of factor analysis techniques (BINFORD, BINFORD 1966). In 1975, the first synthesis of data analysis in archaeology was published in *Mathematics and Computers in Archaeology* (DORAN, HODSON 1975). Thanks to the growing use of computers, multivariate data analysis techniques were already well-established, particularly in the area of automatic classification methods. This represented a technical revolution in the quantitative approach to typology, the study of cultural facies, and chronological seriation.

In contrast, the term *Analyse des Données* refers to a more stringent concept. It is not merely the French-language translation of 'data analysis'; it identifies a school of thought and research promoted by the mathematician Jean-Paul Benzécri. Since the 1970s, this approach has gained a dominant position in France (*fig. 1*). Benzécri challenged the probabilistic foundations on which the statistical procedures commonly used in the Anglo-Saxon world were based, as well as the hypothetico-deductive methods typically

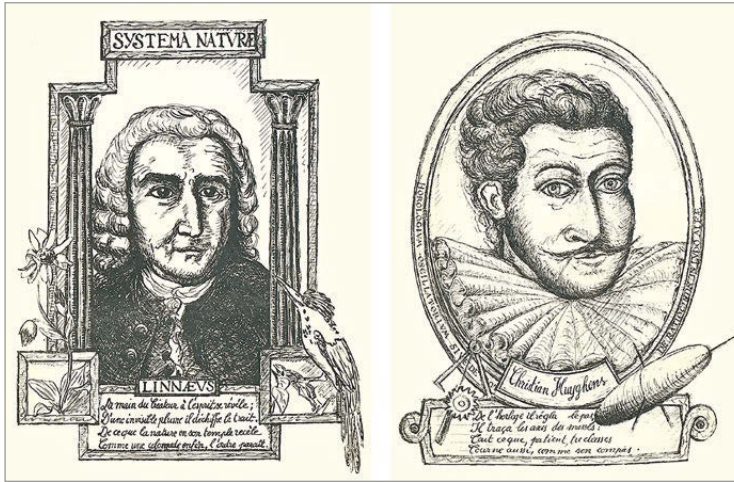


fig. 1 – The figures of Linnaeus and Huygens introduce Benzécri’s two-volume *Analyse des Données* (1973), evoking Taxonomy and Correspondence Analysis, respectively.

employed in research. The five methodological principles enunciated by BENZÉCRI (1980) are often referenced in the history of applied statistics in the social sciences and humanities. They testify to «the international movement back to the sources of statistics – ‘let the data speak’ – and to sometimes fierce criticisms of an abusive formalization» (SAPORTA 2019). We quote them below:

- 1) *Statistique n’est pas probabilité.*
- 2) *Le modèle doit suivre les données et non l’inverse.*
- 3) *Il convient de traiter simultanément des informations concernant le plus grand nombre possible de dimensions.*
- 4) *Pour l’analyse des faits complexes et notamment des faits sociaux, l’ordinateur est indispensable.*
- 5) *Utiliser un ordinateur implique d’abandonner toutes techniques conçues avant l’avènement du calcul automatique.*

In a nutshell, Benzécri asserted the priority of data, or of a data-driven approach – anticipating, according to some, the significant role that Big Data would later play and the transition from classical statistics to Data Science. With the goal «to make patterns emerge from data» (BENZÉCRI 1969), he emphasized the strategic importance of computers in processing vast amounts of data simultaneously to conduct exploratory data analysis and summarize their key characteristics. To achieve this, the adoption of computers and new adapted techniques was essential. As a result, factor analysis, particularly correspondence analysis, emerged among the French community of data analysts, becoming the hallmark of the school’s work. Data encoding, visualization, and computer programs were recognized as crucial steps in the analytical process.

Benzécri and his team adopted a geometric perspective and adapted tools from linear algebra to introduce a new approach to summarizing associations in contingency tables. Indeed, correspondence analysis is a dimensionality reduction method designed for

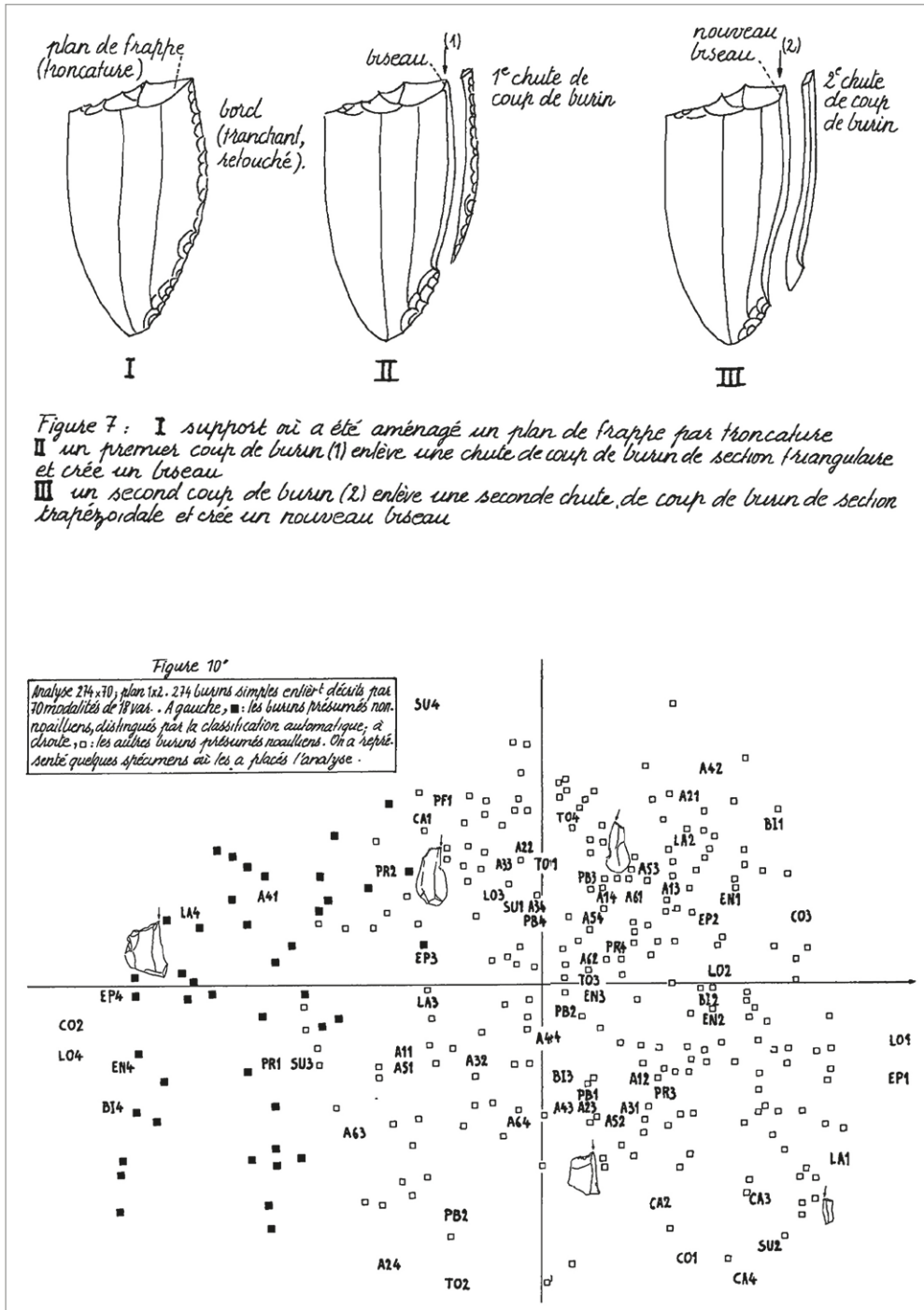


fig. 2 – Selected illustrations from the article ‘Typologie de l’outillage préhistorique en pierre taillée. Application à la définition du type burin de Noailles’ (BENZÉCRI 1977c).

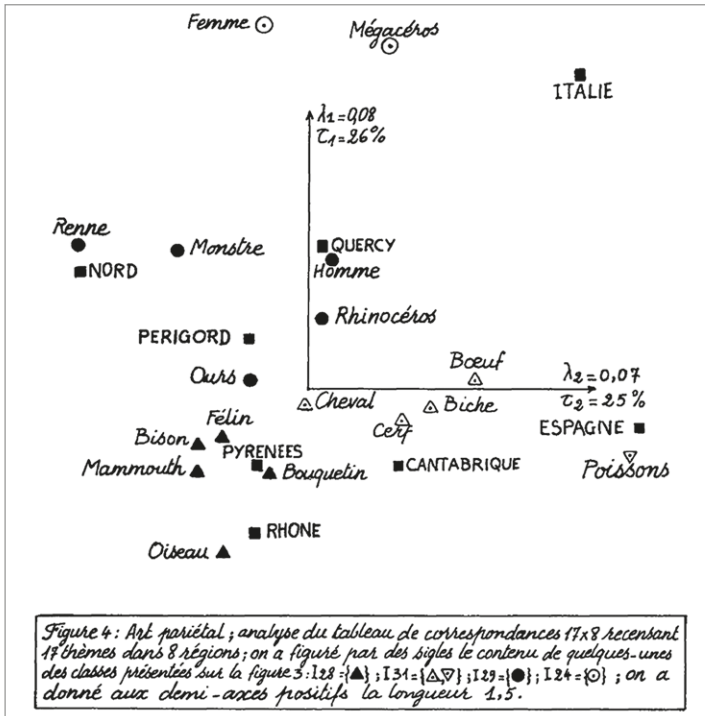


fig. 3 – An illustration from the article ‘Analyse de données sur l’art préhistorique’ (ROUX, ROBERT, JAMBU 1976).

analyzing large datasets and quantifying categorical data. It allows for a unique graphical interpretation of tabular data, enabling users to visualize the relationships between rows and columns of a data matrix. The totality of information is made visible through a geometric representation within a reduced-dimensional space (CIBOIS 1983). This framework is oriented toward users and practitioners, placing particular emphasis on the effective display of results and the development of tools and software to implement these methods in practice.

The most basic form of correspondence analysis, simple correspondence analysis, is applied to a two-way contingency table. Multiple correspondence analysis, in turn, can analyze many categorical variables simultaneously. Textual correspondence analysis is another development characteristic of the French school, reflecting its founder’s abiding interest in the relationship between mathematics and linguistics (BENZÉCRI 1966; BENZÉCRI *et al.* 1981), which lies at the intersection of several disciplines: linguistics, discourse analysis, statistics, and computer science. In this regard, it is worth noting that Benzécri first presented correspondence analysis orally in 1962-1963, during six lectures delivered at the Collège de France as part of the ‘Cours Peccot’ entitled ‘Statistique et structure des langues naturelles; essai de synthèse mathématique’, and that his first course at the University of Rennes focused on mathematical linguistics.

The growing interest in the above-mentioned techniques within archaeology – a disciplinary area that Benzécri closely investigated (BENZÉCRI 1997b, 1997c) (figs. 2-3) –

originates from their ability to create typologies without imposing a priori assumptions on the results. Additionally, the process of data encoding, especially the use of a standardized technical vocabulary, is crucial for managing the heterogeneity of descriptive data related to archaeological materials. Despite employing different methods and perspectives, these goals align in the wider context of automating documentary systems and developing automated documentation techniques. It is worthy of note that the topic of encoding procedures to analyze archaeological artifacts was first introduced in France by Jean-Claude Gardin in the 1950s, who stressed the necessity of adopting a uniform and standardized description system (GARDIN 1958, 1959; GARDIN, LÉVY 1962).

In line with the objectives of the Series – to retrace past experiences, understand the present, and anticipate future developments – this book focuses on the evolution of *Analyse des Données* over the past fifty years, beginning with the publication of Jean-Paul Benzécri’s seminal two-volume work in 1973, which established the theoretical foundation and formal framework for this research field.

The 1980s marked a significant turning point for correspondence analysis techniques. During this decade, these methods began to spread beyond European academic circles, gaining greater international visibility. This expansion was closely linked to advances in computing and programming. The emergence of personal computers made statistical software more accessible, encouraging the development of user-friendly tools, many of which are referenced in this book. As a result, correspondence analysis evolved from a niche method rooted in a distinctively French statistical philosophy into a globally recognized technique, particularly within the social and behavioral sciences. A major milestone during this period was the establishment of the CARME (Correspondence Analysis and Related Methods) conference series in 1991, which created an interdisciplinary environment for methodological innovation and academic exchange (*fig. 4*).

In archaeology, the 1990s saw a decisive consolidation of correspondence analysis and, in general, of exploratory data analysis techniques, both methodologically and practically. A growing body of literature – from introductory texts to advanced treatises – broadened the reach of these methods. Its integration into mainstream statistical software packages, such as SAS, SPSS, and SPAD, further normalized its use.

Starting in 1995, the landscape of data analysis in archaeology began to change. The rising popularity of Geographic Information Systems (GIS) started to overshadow traditional statistical methods. Concurrently, Bayesian statistics experienced a resurgence, especially in the calibration of radiocarbon dating. The 2000s ushered in a ‘3D revolution’ characterized by the adoption of digital photogrammetry and virtual reality techniques. In the following decade, Artificial Intelligence reemerged, showcasing advanced Neural Network models such as Self-Organizing Maps (SOMs), which are useful for data mining and text analysis. Although these models are less effective for the nuanced interpretive needs of archaeologists, they are considered a significant legacy of Benzécri’s school of data analysis.

Despite these technological shifts, correspondence analysis has maintained a stable and valuable role in data analysis. This is supported by the widespread availability of

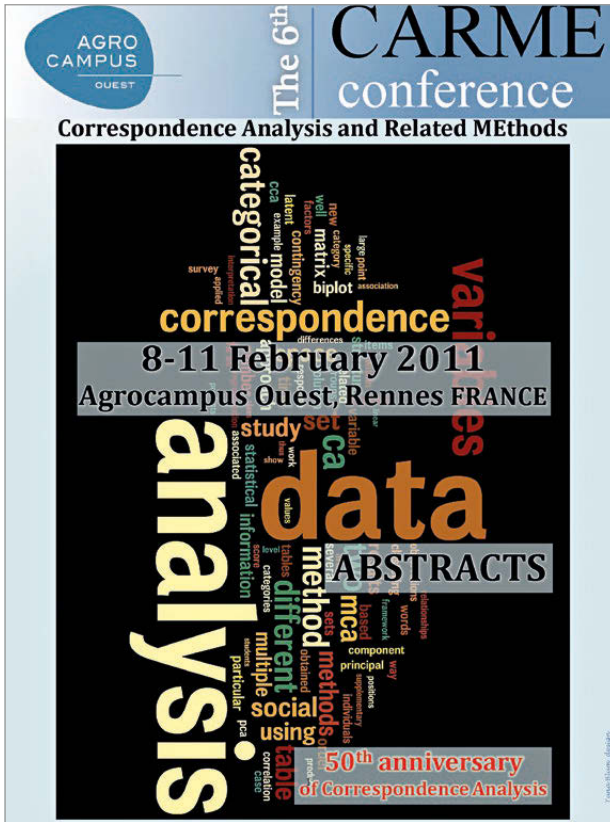


fig. 4 – The 6th edition of the CARME conference, dedicated to the 50th anniversary of *Analyse des Données*, commemorating the beginning of Benzécri’s teaching at the Faculty of Science, University of Rennes.

statistical programming languages like R and their integration with other multivariate statistical techniques. This accessibility continues to promote the application of correspondence analysis across various disciplines, ensuring its ongoing relevance in archaeological research and beyond.

The present book follows this evolutionary trajectory and consists of seven chapters. François Djindjian reconstructs the most salient moments in the development of the French school of data analysis, highlighting its impact on quantitative applications in prehistoric archaeology, a field that has greatly benefited from unsupervised classification methods and dimensionality reduction techniques. Paola Moscati explores the reception and dissemination of multivariate analysis techniques, with a particular focus on the application of correspondence analysis in classical archaeology and its distinctive features. The final two chapters, authored by Alessandro Di Ludovico and Lorenzo Cardarelli respectively, address two contemporary topics with historical roots: Textual Data Analysis and the transition from Data Analysis to Data Science.

The volume concludes with a comprehensive bibliographical section that reflects the scope of this ‘quantitative’ movement, which has profoundly influenced archaeological scholarship over the past 50 years.

François Djindjian

2. THE FRENCH SCHOOL OF DATA ANALYSIS

Le terme même d'analyse des correspondances remonte à l'automne de 1962, et le premier exposé de la méthode sous ce titre fut donné par J. P. Benzécri au Collège de France dans une leçon du cours Peccot de l'hiver 1963 (BENZÉCRI 1977a, 9).

2.1 INTRODUCTION

The French school of data analysis illuminated the world of statistics for nearly 30 years between 1965 and 1995.

Under the generic term of ‘data analysis’, we can classify a variety of techniques, including the family of factor analyses (psychometric factor analysis, principal component analysis, and correspondence analysis), the family of cluster analysis (hierarchical ascending classifications, non-hierarchical classifications, and segmentations), and the family of non-metric multidimensional scaling, not to mention canonical analysis and discriminant analysis.

These methods were developed from the beginning of the 19th century onwards: principal component analysis by Karl Pearson in 1901 (PEARSON 1901; HOTELLING 1933); canonical analysis by Harold Hotelling in 1936 (HOTELLING 1936); discriminant analysis by Ronald Fisher in 1936, in his famous study of irises (FISHER 1936); ‘factor analysis’ by the psychometrician Louis Thurstone in 1947 (THURSTONE 1947); cluster analysis by Robert Sokal and Peter Sneath in 1963 (SOKAL, SNEATH 1963); and multidimensional scaling by Roger Shepard and Joseph Kruskal in 1962 and 1964, at the Bell Telephone Laboratories (SHEPARD 1962; KRUSKAL 1964).

However, it is thanks to the development of computer science, and in particular to processing power, memory size, and algorithm programming, that these techniques have been at the heart of a revolution in the humanities and social sciences since the 1970s.

The French school of data analysis was begun, first and foremost, by Jean-Paul Benzécri, professor at the University of Rennes between 1960 and 1965 (*fig. 1*). A former student at the École Normale Supérieure, in 1955 he obtained a PhD in mathematics through a joint program between the École Normale Supérieure and Princeton University, under the supervision of the mathematician Henri Paul Cartan. In 1965, at the University of Rennes – where an IBM 1620 had been installed two years earlier (*fig. 2*) – he supervised Brigitte Cordier-Escofier’s thesis, which marked the birth of correspondence analysis based on the chi-square metric with the property of distributional equivalence.

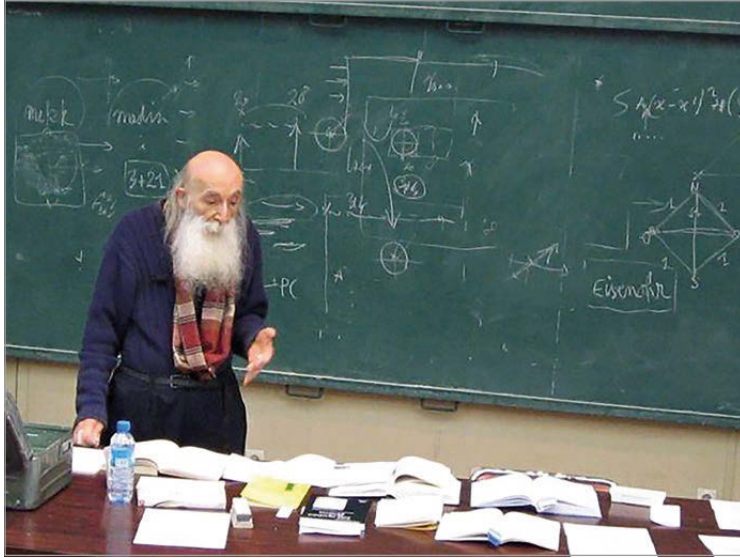


fig. 1 – Jean-Paul Benzécri during a lecture at the Institut National Agronomique Paris-Grignon (INA-PG), in October 2006.



fig. 2 – The IBM 1620 installed at the Computing Center of the University of Rennes in 1963 and nicknamed ‘Caroline’. As recalled by J. ANDRÉ (2004, 230), during an initial demonstration, «Benzécri, qui n’avait jamais pratiqué Fortran, réussit à écrire en deux heures un programme de résolution de systèmes linéaires par la méthode de Gauss».

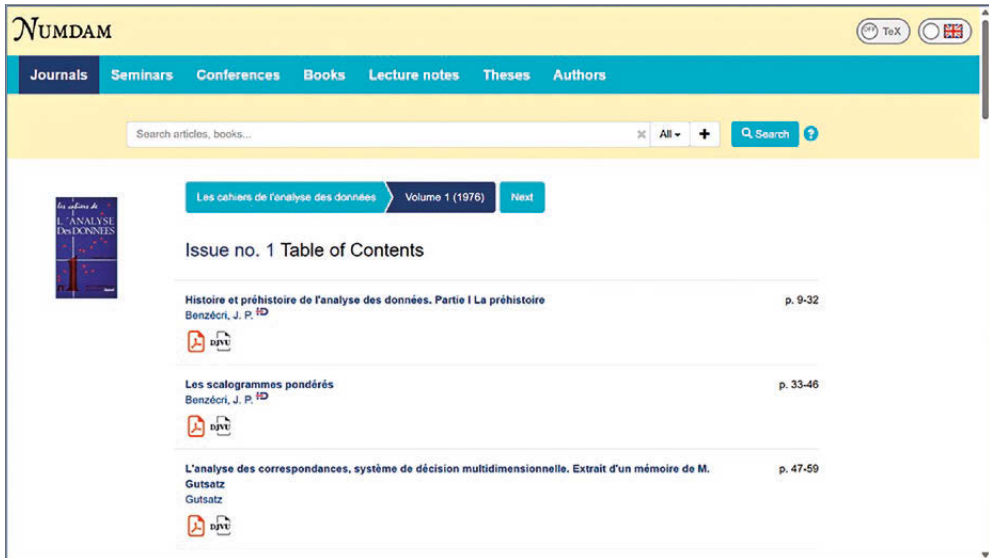


fig. 3 – «Les Cahiers de l'analyse des données», available for consultation on the Numdam platform, the French digital mathematics library (<https://www.numdam.org/item/CAD/>).

Benzécri then held the chair of statistics at the Institut de Statistique (ISUP) of the University Pierre et Marie Curie (Paris VI) from 1967 until 1992. The first Parisian thesis was defended by Maurice Roux in 1968 (ROUX 1968). In the 1970s, the DEA (Diplôme d'Études Approfondies) program enrolled approximately 200 students, and 40 theses were defended each year at the University of Paris VI. The publication of the two volumes on data analysis (Taxonomy and Correspondence Analysis) in 1973 was a critical juncture in this story (BENZÉCRI 1973).

The laboratory used 60% of the computing hours available on the University's computer. In addition, researchers had access to the computers of the Centre Inter-Régional de Calcul Électronique (CIRCE) at the Centre National de la Recherche Scientifique (CNRS) in Orsay, with the so-called 'Saturday researchers'. In 1976, the journal «Les Cahiers de l'analyse des données» was created by Dunod and ran until 1997 (fig. 3).

2.2 JEAN-PAUL BENZÉCRI AND HIS COMMUNITY

Gravitating around Jean-Paul and Françoise Benzécri (who was always present) was a team of close collaborators who contributed to the development of methods and software, including multiple correspondence analysis with complete disjunctive coding and the joint use of correspondence analysis and hierarchical ascending classification: Maurice Roux, Pierre Cazes, Michel Jambu, Jean-Pierre Fénelon, Marie-Odile Lebeaux, but also Henry Rouanet and Ludovic Lebart.



fig. 4 – The Bell Laboratories Building, erected in 1925 (New York City, 463 West Street).

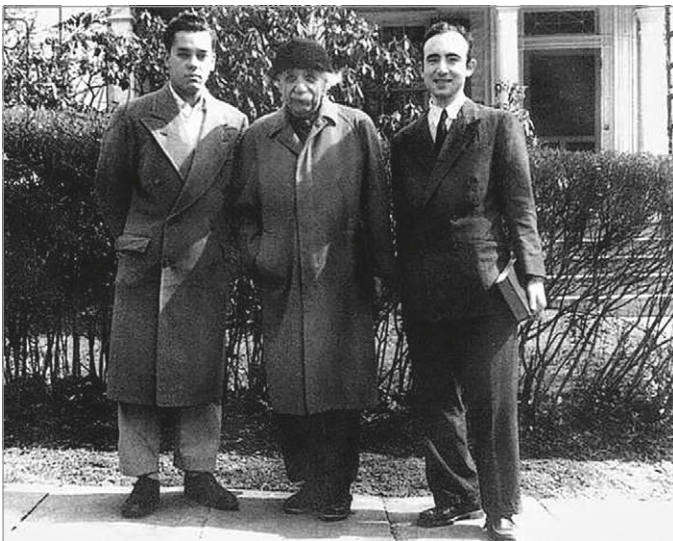


fig. 5 – Albert Einstein flanked by Jean-Paul Benzécri and José Francisco Cadilla. The photo was taken in 1955 in Princeton, shortly before Einstein's death, when Benzécri and Cadilla were there as students (© Estate of Gerald Bloncourt. All Rights Reserved 2025/Bridgeman Images).

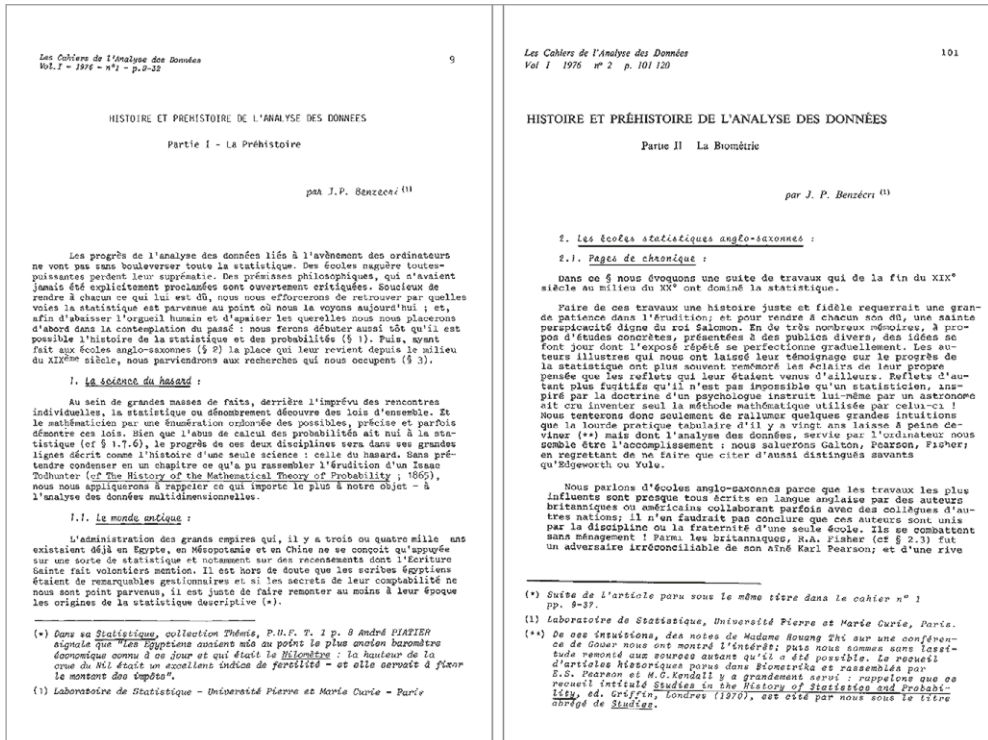


fig. 6 – The first two articles on the ‘Histoire et préhistoire de l’analyse des données’, published in «Les Cahiers de l’analyse des données» in 1976. See https://www.numdam.org/item/CAD_1977__2_1_41_0/ for a systematic index of subsequent articles.

All the DEA students recall the lectures given by Benzécri – who sported the beard of a prophet and used to dress in dark velvet pants and gardening boots – in the crowded amphitheater at the University of Paris VI, and the students who queued to see him in his small office after the class, to ask a question or to seek help.

Benzécri’s divergence from the Anglo-Saxon approach, notwithstanding his Princeton background, was probably the result of his unsuccessful visit to Bell Labs in 1965 (figs. 4-5) to present correspondence analysis to the multidimensional scaling team, who were only impressed by the extravagances of their visitor (as Henry Rouanet later reported: ROUANET 2008). Benzécri was also upset by Mark Hill’s article published in 1974 under the apparently promising title of ‘Correspondence Analysis: A Neglected Multivariate Method’ (HILL 1974). Benzécri felt that Hill’s article presented correspondence analysis in a biased manner, positioning it as a filiation of Guttman’s scalogram method (GUTTMAN 1941) and Torgerson’s scaling approach (TORGERSON 1958).

In addition, Hill positioned correspondence analysis as a scaling method rather than as a method for processing contingency tables. Benzécri’s reaction was to write several articles in «Les Cahiers de l’analyse des données» in 1976 and 1977, under the title

‘Histoire et préhistoire de l’analyse des données’ (*fig. 6*), which were later merged into a book published in 1982 (BENZÉCRI 1976a-d, 1977a, 1982).

Regarding Hill’s article, here are some sharp remarks by Benzécri:

«Le terme même d’analyse des correspondances remonte à l’automne de 1962, et le premier exposé de la méthode sous ce titre fut donné par J. P. Benzécri au Collège de France dans une leçon du cours Peccot de l’hiver 1963. En nous référant au terme même, nous évitons de nous prononcer d’abord, quant à la définition des facteurs issus d’un tableau rectangulaire de nombre positifs, sur des questions de priorité qu’un article récent pourrait soulever, mais que nous préférons réduire à leur juste proportion sinon à leur solution définitive par un exposé chronologique, où seront scrupuleusement notées les rencontres successives de l’analyse des correspondances avec les travaux d’autres écoles» (BENZÉCRI 1977a, 9).

In addition to being a high-level mathematician, the inventor of correspondence analysis, and the promoter of the dissemination of software packages for all data analysis methods (ADDAD, SPAD) through his team, Benzécri was also the epistemologist who championed the ability of multidimensional analysis to discover structures in data through the miracle of factorization. Benzécri’s attitude was to separate statistics from probabilities (as a geometrical method) and to consider the formalization upstream of the data set unnecessary – such as the agronomist’s experimental design, the sociologist’s sampling procedure or questionnaire, or the economist’s model, to which the archaeologist might add the formalization of the description of artifacts.

Benzécri’s reliance on students to collect data, without possessing the necessary erudition to do so, gradually alienated practitioners and statisticians. Nevertheless, he maintained the same enthusiasm for data analysis, which he practiced to bring out the right factors and structures. It is interesting to note that this confidence in the ability of the algorithm to process all the data in order to find the right result is also evident in contemporary Artificial Intelligence (AI) approaches, as seen with archaeologists using Bayesian methods for radiocarbon dating. The alternative approach assumes that input noise will always produce output noise, in other words, anarchic data will produce incomprehensible mixtures of structures, and that only input formalization will consistently be more efficient than the best algorithm.

There were, therefore, those who accompanied Benzécri for a shorter or more extended period and who continued their professional career in other institutions, such as:

- Brigitte Escofier-Cordier at the University of Rennes (ESCOFIER 1969; ESCOFIER, PAGÈS 1984, 2008) (*fig. 7*);
- Henry Rouanet and Brigitte Le Roux at the University of Paris V (ROUANET, LE ROUX 1993);
- Philippe Cibois at the University of Paris V and then at the University of Saint-Quentin (CIBOIS 1983, 1996) (*fig. 8*);
- Ludovic Lebart at the Centre de recherche pour l’étude et l’observation des conditions de vie (CREDOC) and then at the École Nationale Supérieure des Télécommunications



fig. 7 – Brigitte Escofier and the first page of her *L'analyse factorielle des correspondances*, published in 1969 by the Institut de Statistique de l'Université de Paris (ESCOFIER 1969).



fig. 8 – Two books that contributed to the dissemination of French data analysis: *L'analyse factorielle*, published in the series *Que sais-je?* (CIBOIS 1983) and *Qu'est-ce que l'Analyse des Données*, published in the Collection Lefonen (FÉNELON 1981).

(ENST) (LEBART 1978; ; LEBART, MORINEAU, PIRON 1995; LEBART, MORINEAU, TABARD 1977) (fig. 9);

- Pierre Cazes at the University of Paris-Dauphine;
- Michel Jambu at the Centre National d'Études des Télécommunications (CNET) and then at the University of Nice (JAMBU, LEBEAUX 1983);
- Maurice Roux at the Institute Agro Montpellier and then at the University of Aix-Marseille.

The French school of data analysis also includes the community of statisticians who, while not claiming to be part of Benzécri's school, contributed to its international reputation through their work:

- Edwin Diday, at the University of Paris-Dauphine (Ceremade laboratory), author of a K-means non-hierarchical classification (DIDAY 1980);

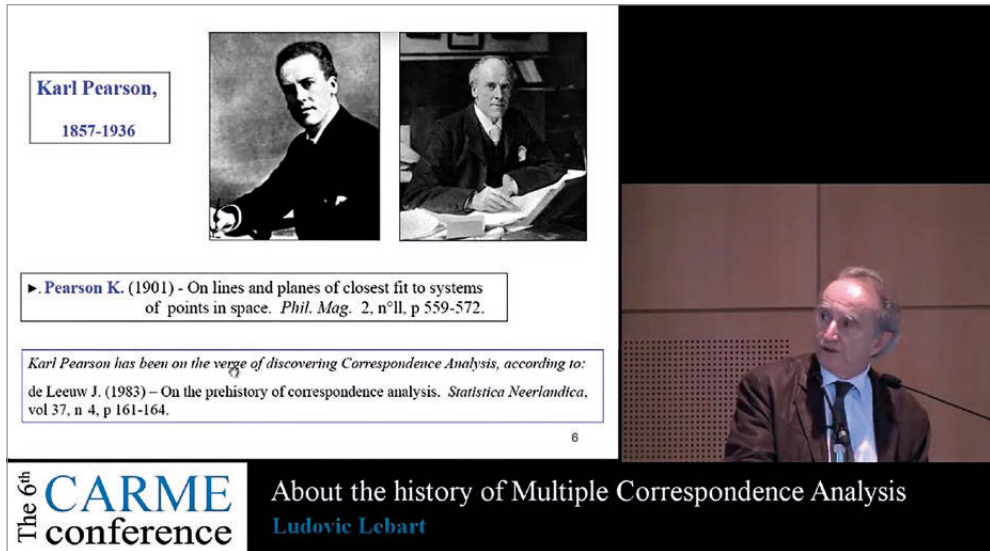


fig. 9 – Ludovic Lebart illustrates the history of *Analyse des Données* at the 6th CARME (Correspondence Analysis and Related Methods) conference (Rennes 2011).

- Jean-Pierre Pagès at the Commissariat à l'Énergie Atomique (CEA) (CAILLIEZ, PAGÈS 1976);
- Michel Tenenhaus at the École des Hautes Études Commerciales (HEC) (TENENHAUS 1998);
- Michel Volle at the Institut National de la Statistique et des Études Économiques (INSEE) (VOLLE 1997);
- Israël César Lerman at the University of Rennes (IRISA Laboratory), with his work on automatic classification (LERMAN 1970, 1981);
- Henri Leredde at the University of Paris XIII, with his work on seriation (LERMAN, LEREDDE 2022);
- Jean-Pierre Nakache at the Institut National de la Santé et de la Recherche Médicale (INSERM);
- Guy Der Megreditchian at the National Meteorology;
- Jérôme Pagès at the University of Rennes (Agrocampus);
- Jean-Marie Bourroche, with Gilbert Saporta (BERTIER, BOUROCHE 1981; BOUROCHE, SAPORTA 2006; SAPORTA 2006), founded the consulting company Coref before Saporta became a professor at the Conservatoire National des Arts et Métiers (CNAM). Coref promotes the use of multidimensional scaling applied to quantitative marketing and credit scoring studies, through the Disqual program.

The ADDAD (FÉNELON 1981), SPAD (LEBART, MORINEAU 1982), and Tri-deux (CIBOIS 1996) packages have largely popularized the use of French-style data analysis,

which the R library has gradually replaced. The SAS and SPSS software packages have also included French programs in their library.

Interest in data analysis was less pronounced among the *École des Hautes Études en Sciences Sociales* (EHESS) teams, such as Marc Barbut and Bernard Monjardet, who were associated with the journal «*Mathématiques et sciences humaines*», active between 1962 and 2012, as well as with the *Laboratoire d'Informatique pour les Sciences Humaines* (LISH) in Marseille and Paris. This laboratory, which supported many researchers (including Mario Borillo, Alain Guénoche, Jacques Virbel, Wenceslas Fernandez de la Vega, Philippe Cibois), was disbanded by the CNRS in the early 1980s (BORILLO 1978).

The spread of correspondence analysis in the Anglo-Saxon world can certainly be attributed to the sustained efforts of Michael Greenacre, who defended his thesis under the supervision of Benzécri in 1973 (GREENACRE 1984).

2.3 THE STATISTICAL METHODS OF THE FRENCH SCHOOL OF DATA ANALYSIS

It is important to emphasize the differences in impact between the software packages of the French school of data analysis, which had limited distribution outside France, and American packages, which achieved worldwide distribution (for example SAS or SPSS).

The methods of the factorial approach are principal component analysis (applied to a table of measurements), correspondence analysis (applied to a contingency table) and multiple correspondence analysis (applied to a complete disjunctive table and its associated complement, the Burt table).

Aids to the interpretation of factors – such as the histogram of eigenvalues, the inertia of individuals and variables, the contributions of individuals and variables to the construction of factorial axes, and square cosines – are essential elements not only for analyzing factorial axes but above all for improving the readability of the structures that have emerged by redoing the analysis after refining the data table. In this iterative process, the function of the main elements and the function of additional elements are very important in the optimization of the structure (elimination of aberrant individuals, redefinition of ambiguous variables, hierarchization of variables to identify the right level of relevance of the description, and removal of variables and individuals with too great a weight, etc.).

Over time and with applications, factor analysis programs have been enriched with new features to help researchers in exploring data. A notable example is Philippe Cibois' *Tri-deux* program (CIBOIS 1996) and Ludovic Lebart's *DtmVic* (LEBART, PIRON 2012) (*fig. 10*).

The clustering methods include hierarchical ascending classification, which incorporates similarity indices and cluster aggregation criteria of the numerical taxonomy, while also introducing variance aggregation (in close association with the chi-square distance of the correspondence analysis) and reciprocal neighbor clustering, but also non-hierarchical classification (aggregation around mobile centers).

Here again, cluster interpretation aids enable a statistical description of each cluster, revealing the role that the variables play in the characterization of the clusters. Additionally, individuals can be treated as additional elements, which entails the implementation of a mechanism for classifying an individual into the cluster that is closest to him. It is thus possible to classify the same individuals who were used to construct the clusters, in order to calculate an accurate ranking index, which helps determine the optimal cut-off level of the clustering tree.



Site personnel de Philippe Cibois

Page d'accueil

Logiciel Trideux

Présentation

Aide en ligne

Chargement

Textes en ligne

Textes pédagogiques

Livres en ligne

Blog Quanti

Articles

Autres textes

Textes de travail

Archives

Le logiciel Trideux

Il s'agit d'un logiciel libre (**précisions**) de dépouillement d'enquête, librement reproductible. Il fonctionne sous windows (de windows 98 à windows 10). L'objet de ce logiciel est de fournir aux chercheurs, enseignants, étudiants et à toute personne en ayant l'utilisation, un outil simple d'usage et gratuit pour dépouiller des enquêtes en utilisant des techniques simples comme les tris croisés ou plus complexes comme l'analyse factorielle, la classification automatique, des méthodes post-factorielles ou la régression sur données d'enquête.

Les données peuvent exister préalablement à l'utilisation de Trideux : le cas le plus habituel est constitué de données individuelles où à chaque ligne correspond un individu statistique et à chaque position une réponse de cet individu à une question. Ces données peuvent avoir été saisies avec Excel (puis transférées directement ou en utilisant le format Excel .CSV) ou venir d'autres logiciels en Ascii.

Avec Alex ALBER, actuellement maître de conférences à l'université de Tours, a été mis au point à partir de 2005 une version initialement destinée à l'enseignement à l'UVSQ (Université de Versailles St-Quentin) sous la forme d'une version 5 de Trideux. Elle n'est plus diffusée car Alex Alber a mis au point une version Cloud de Trideux, c'est à dire que le programme n'a plus à être installé sur l'ordinateur de l'utilisateur mais qu'il est utilisable directement sur Internet. Trideux.cloud est accessible [à cette adresse](#).

DTMVIC : Data and Text Mining : Visualization, Inference, Classification



DtmVic
Ludovic Lebart
DR. C.N.R.S. IRI
TELECOM Paris
Contact:
lebart@lpsat.com



Standing on the shoulders of giants... (Bernard de Chartres, 1158)

Download some DTMVICs here

<p>Activités</p> <ul style="list-style-type: none"> • Recherche • Publ. (livres) • Publ. (articles) 	<p>Activities</p> <ul style="list-style-type: none"> • Research • Publ. (books) • Publ. (articles) 	<p>Actividades</p> <ul style="list-style-type: none"> • Investigación • Publ. (libros) • Publ. (artículos) 	<p>Atividades</p> <ul style="list-style-type: none"> • Pesquisa • Publ. (livros) • Publ.(artigos) 	<p>Attività</p> <ul style="list-style-type: none"> • Ricerca • Pubbl.(libri) • Pubbl.(articoli)
<p>Logiciel DtmVic</p> <ul style="list-style-type: none"> • Télécharger • Tutoriels • Manuel Dtm-Vic 	<p>Software DtmVic</p> <ul style="list-style-type: none"> • Download • Tutoriels • User's Guide 	<p>Software DtmVic</p> <ul style="list-style-type: none"> • Descargar • Tutoriales • Manual del usuario 	<p>Software DtmVic</p> <ul style="list-style-type: none"> • Programa • Tutoriais • User's Guide (Eng.) 	<p>Software DtmVic</p> <ul style="list-style-type: none"> • Scaricare • Tutoriali • User's Guide (Eng.)
<p>Python (language)</p> <ul style="list-style-type: none"> • Python (textual data) • Python (PCA) • Python (ACF) • Python (AC) 	<p>Python (language)</p> <ul style="list-style-type: none"> • Python (textual data) • Python (PCA) • Python (CA) 	<p>Python (language)</p> <ul style="list-style-type: none"> • Python (datos textuales) • Python (ACP) • Python (AC) 	<p>Python (language)</p> <ul style="list-style-type: none"> • Python (dados textuais) • Python (PCA) • Python (CA) 	<p>Python (linguaggio)</p> <ul style="list-style-type: none"> • Python (dati testuali) • Python (PCA) • Python (CA)
<p>R (language)</p> <ul style="list-style-type: none"> • R (ACP, AC, ALM) 	<p>R (language)</p> <ul style="list-style-type: none"> • R (PCA, CA, MSt) 	<p>R (language)</p> <ul style="list-style-type: none"> • R (ACP, AC, ALM) 	<p>R (language)</p> <ul style="list-style-type: none"> • R (PCA, CA, MSt) 	<p>R (linguaggio)</p> <ul style="list-style-type: none"> • R (ACP, AC, MSt)
<p>Textométrie et poésie</p> <ul style="list-style-type: none"> • Biaisens / Shakespeare... 	<p>Textometrics and poetry</p> <ul style="list-style-type: none"> • Biaisens / Shakespeare... 	<p>Análisis de datos textuales y poesia</p> <ul style="list-style-type: none"> • Biaisens / Shakespeare... 	<p>Análise de dados textuais e poesia</p> <ul style="list-style-type: none"> • Biaisens / Shakespeare... 	<p>Analisi dei dati testuali e poesia</p> <ul style="list-style-type: none"> • Biaisens / Shakespeare...

fig. 10 – Trideux and DtmVic software designed by Philip Cibois and Ludovic Lebart, respectively (<https://ciboispagesperso.fr/Trideux.html> and <http://www.dtmvic.com/index.html>).

The combined use of factor analysis and cluster analysis facilitates their interpretation through many functions, such as projecting the inertia ellipses of the clusters on the factorial axes.

Other methods have more specialized applications, such as canonical analysis, multiple regression, discriminant factor analysis, discriminant analysis on qualitative variables (Disqual), principal component analysis on qualitative variables, algorithms implementing the log-linear model, Procrustes analysis, and generalized canonical analysis. These methods reveal the richness of the corpus of multidimensional statistical techniques. Moreover, it is important to specify that the methods cited are those in common use and that each statistician has been able to develop or improve them, particularly between 1975 and 1995, when dissemination was all the more restricted because open access software was not available.

It is interesting to note the existence of algorithms that introduce additional constraints between individuals or between variables in both factorization and classification. The first example is the aggregation with contiguity constraints, which is essential in spatialized applications (LEBART 1978) and has been applied in intrasite spatial analysis in archaeology. The second example is the aggregation with chronological constraints, which we proposed in the toposeparation method (DJINDJIAN 1985b).

Finally, Artificial Intelligence (AI) emerged in this field from the 1990s onwards via multilayer neural networks with supervised models (i.e., discriminant analyses) and self-organized models (i.e., cluster analysis), such as those proposed by Kohonen's SOM algorithms (KOHONEN 1989), whose primary advantage is the speed of execution in processing big data.

2.4 FIELDS OF APPLICATION

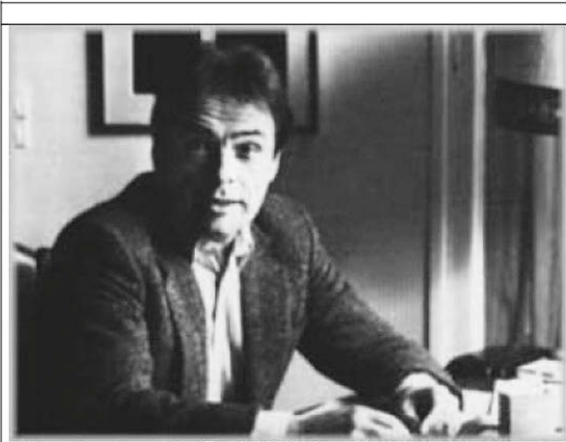
Reading «Les Cahiers de l'analyse des données» reveals the vast field of application of data analysis in the following domains:

- humanities and social sciences (textual analyses, linguistics, survey questionnaires, opinion surveys, psychometry, sociology, politics, archaeology);
- economics (econometrics, quantitative marketing);
- earth sciences (geology, geography, paleontology, ecology, systematics of living beings, paleoclimatology, paleo-environmental analysis);
- biology (pharmacological analyses, biological analyses, genomic analyses);
- agronomy;
- exact sciences (particle physics);
- computer science (AI, robotics);
- business sector (bank credit scoring, insurance risk scoring and all the applications of data mining);
- music (music score analysis).

L'@alyse des données : histoire, bilan, projets, ..., perspective

Jean-Paul BENZÉCRI

In memoriam : Pierre BOURDIEU



Pierre Bourdieu, né le 1^{er} août 1930 à Denguin (Pyrénées-Atlantiques),
décédé à Paris le 23 janvier 2002.

Il y a un demi-siècle, Pierre Bourdieu et moi-même étions élèves à l'Ecole Normale Supérieure de la rue d'Ulm ; lui dans la section des Lettres, et moi en Sciences. Or, le savoir ne se sectionne pas ! C'est pourquoi, depuis lors, nous ne nous sommes jamais longtemps perdus de vue, l'un l'autre...

La dernière lettre reçue de Bourdieu me posait une question trop difficile : c'était, ou presque : qu'est-ce que l'Analyse des Données ? Je lui fis une réponse évasive.

Maintenant que mon ami n'est plus de ce monde, je lui dois de rassembler mes esprits, sinon pour lui répondre, au moins pour attester que je n'ai pas de réponse à donner qui me satisfasse.

fig. 11 – Benzécri's tribute in memoriam to Pierre Bourdieu, published in 2006 in the journal «MODULAD» (Le Monde des Utilisateurs de L'Analyse de Données).

This is why, in many disciplines, the use of data analysis has become unavoidable, and it would be presumptuous to believe that all the applications can be listed. Thus, for example, Pierre Bourdieu made it the preferred method of building social spaces (BOURDIEU *et al.* 1978) (fig. 11). The author of this text still considers that data analysis is at the heart of epistemology and the methods of archaeology, an aspect that did not escape Benzécri's judgment:

«Il semble aujourd'hui qu'après d'autres disciplines l'archéologie préhistorique a rencontré l'analyse des données. A la vérité il est impossible d'assigner une date

à cette rencontre... d'ailleurs depuis 25 ans et plus, chaque méthode statistique nouvelle relevant de près ou de loin de ce que nous appelons analyse des données, n'a guère tardé à s'essayer à la préhistoire. Mais il est symptomatique qu'ait paru en 1975 un ouvrage britannique qui sous le titre de *Mathématiques et ordinateur en archéologie*, traite en fait pour plus de la moitié d'analyse des données; cependant que dans un mémoire de maîtrise ès lettres soutenu à Paris en Juillet 1976 F. Djindjian, appliquant l'analyse des données à l'étude de *l'outillage de pierre taillée*, a obtenu des résultats auxquels le IX^e Congrès de l'Union Internationale des Sciences Préhistoriques et Protohistoriques (Nice 13-18 Septembre 1976) a fait un chaleureux accueil; accueil mérité selon nous par le progrès méthodologique décisif que constitue l'application conjuguée de l'analyse des correspondances, du codage sous forme disjonctive complète et de la classification par agrégation suivant la variance avec la distance du χ^2 » (BENZÉCRI 1977b, 193).

2.5 THE KEY PHASES OF DATA ANALYSIS IN FRANCE

It is possible to distinguish several periods in the development of data analysis in France.

- The first period was that of the pioneers (1965-1975) who developed the methods, the first software programs, and university training.
- The second period is that of the exploration of the field of applications of data analysis (1975-1985) with publications, theses, specific methods (for instance, in archaeology spatial analysis, toposeparation or Procrustes analysis), and the first seminars in application disciplines.
- The third period (1985-1995) was that of the integration of these applications into the sciences concerned (first syntheses, university training integrated into the curriculum, proliferation of master's and thesis courses).
- The fourth period (1995 to the present day) is that of the trivialization of the use of data analysis, also marked by a decrease in mathematical innovation on new algorithms, and in the humanities and social sciences by the scientific regression of postmodernism. The data analysis algorithms are then integrated into computer methods that work for the user, as black boxes, such as Harris matrix processing programs in archaeology. However, the reactivation of research is regularly manifested to meet new needs marked by the big data constraints (data mining, AI, 3D, Internet).

2.6 CONCLUSIONS

In this short text on the French school of data analysis, it is important to emphasize that the researchers designing the algorithms and writing the software programs, as well as the researchers in the application disciplines, felt that they had participated in

an epistemological revolution in the humanities and social sciences over the course of thirty years.

Those who have mastered statistical techniques and have succeeded in the essential step of formalizing the problem – defining the population, selecting the variables, and putting them in a datatable – have obtained spectacular results. Conversely, those who merely processed a data table to ‘do science’ and obtain a factorial map generally obtained disappointing results. Good results depend on many iterations, which are needed to make a structure appear and to improve its sharpness, in a learning process that is reminiscent of the algorithms used in AI.

Paola Moscati

3. CORRESPONDENCE ANALYSIS: A 'NEGLECTED MULTIVARIATE METHOD' IN THE INTERNATIONAL PANORAMA?

Nous parlons d'écoles anglo-saxonnes parce que les travaux les plus influents sont presque tous écrits en langue anglaise par des auteurs britanniques ou américains collaborant parfois avec des collègues d'autres nations; il n'en faudrait pas conclure que ces auteurs sont unis par la discipline ou la fraternité d'une seule école... et d'une rive à l'autre de l'Atlantique les diverses doctrines de l'analyse factorielle ont garde de former un harmonieux concert. Mais les chercheurs de ces écoles rivales bénéficient des travaux les unes des autres; ils ignorent au contraire ce qu'on produit dans d'autres langues en dehors de leurs cercles (BENZÉCRI 1976b, 101-102).

3.1 THE STATISTICAL ROOTS

The title of this chapter references an article by Mark Hill published in the 1970s in the «Journal of the Royal Statistical Society», which aimed to illustrate the statistical technique known as 'correspondence analysis' (HILL 1974). Hill defines this technique as «an analogue of principal components analysis, which is appropriate to discrete rather than to continuous variates», and provides a brief overview of its development. The method was formerly proposed by Hermann HIRSCHFELD (1935), but since Ronald FISHER (1940) did not mention this text, the latter has often been regarded as the inventor of the technique known as contingency table analysis. In Hill's interpretation, correspondence analysis is algebraically equivalent to Fisher's contingency table analysis. However, as it usually applies to both incidence data and contingency tables, Hill asserts that the term was derived from Jean-Paul Benzécri's *analyse factorielle des correspondances*, implicitly acknowledging that the technique, while having its roots in the 1930s, was developed in France during the late 1960s (*fig. 1*).

Four decades later, in their comprehensive historical discussion of correspondence analysis, Eric BEH and Rosaria LOMBARDO (2012) also argued that the seed of correspondence analysis germinated in France during the 1960s, although its 'genealogy' is complex. For the authors, the crucial groundwork necessary for this seed to develop was laid by English statisticians, in particular Pearson and Fisher, who played a pivotal role in measuring the association between categorical variables and in paving the way for modern correspondence analysis. In his 1901 and 1904

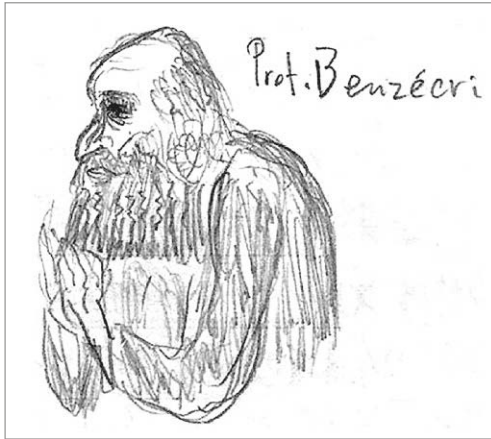


fig. 1 – Sketch of J.-P. Benzécri by S. Iwatsubo, made during the 2nd International Symposium on Data Analysis and Informatics (ISDAI), held at Versailles, France, in the autumn of 1979 (IWATSUBO 2023).

papers, Pearson employed linear regression to calculate the correlation coefficient for a two-way contingency table (PEARSON 1901, 1904). According to Jan DE LEEUW (1983), he was on the verge of discovering correspondence analysis. Later, in 1940, Fisher introduced discriminant functions, which reduce the number of dimensions necessary to analyze a multidimensional space while maximizing the informational content of the data. This approach was applied to the cross-classification of the eye and hair colors of Scottish children (FISHER 1940).

Two other influential scholars are frequently referenced for their pioneering contributions: Louis GUTTMAN (1941) and Cyril BURT (1950). Early correspondence analysis, initially centered on two-way contingency tables, was significantly advanced in 1941 by Guttman, who extended the method to multidimensional scenarios by introducing dual or optimal scaling for handling multi-way contingency tables. In the following decades, researchers explored these ideas in innovative ways, notably Burt, who, in 1950, introduced a novel method for organizing and analyzing data using a special type of table. The ‘Burt matrix’ organizes contingency tables in a ‘super-diagonal’ format to examine patterns and relationships across multiple variables simultaneously.

In the history of correspondence analysis there is broad agreement that, while its theoretical origins can be traced back to France in the 1960s, its introduction to the English-speaking research community during the 1970s was a slow process. The method started to appear in English statistical literature in the early 1970s, following Benzécri’s 1969 article in the proceedings of the international conference on *Methodologies of Pattern Recognition* (University of Hawaii, Honolulu, 1968), published by the Academic Press (BENZÉCRI 1969). His approach reframed statistical associations in spatial terms, as systematically articulated in the two-volume treatise (BENZÉCRI 1973), which established its theoretical foundations. However, the essentially French character of the method, both in language and conceptual framing, along with the lack of English translations, limited its accessibility and broader dissemination.

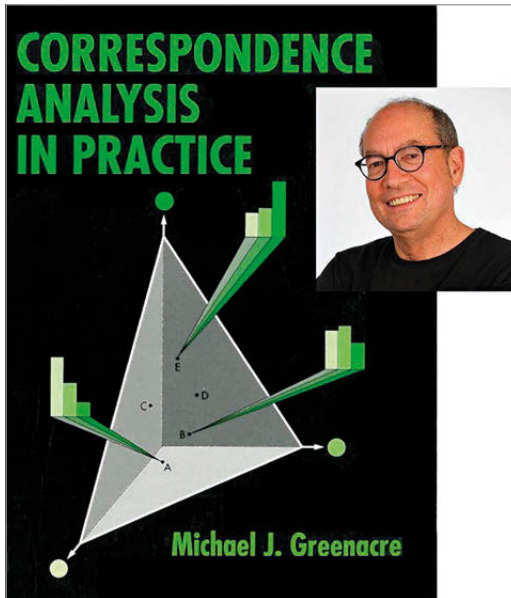


fig. 2 – The popular book by Michael Greenacre, in its second edition, which helped to spread correspondence analysis in the Anglo-Saxon world.

The initial substantial exposure to English-speaking audiences is frequently credited to Hill's influential publications. However, the full extent and originality of Benzécri's contributions to multidimensional correspondence analysis are not always evident in Hill's work (BEAUDOUIN 2016). In fact, the gradual evolution of correspondence analysis during this period was primarily driven by researchers familiar with Benzécri's ground-breaking work, which centered on a geometric approach to contingency table analysis rooted in linear algebra. In the 1980s, the method gained significant recognition beyond Europe, marking a turning point in its adoption, particularly in the UK and the USA. This rise in popularity was driven by influential texts by Michael GREENACRE (1984) and Ludovic LEBART, Alain MORINEAU, and Kenneth WARWICK (1984), which played a crucial role in shaping the method's theoretical and applied framework, as well as by researchers like Leo GOODMAN (1970), who linked correspondence analysis to categorical data modelling.

Greenacre introduces his text by recalling the period of his doctoral studies in France:

«When I arrived in Paris, in July 1973, I had little idea what lay in store for me. After a traditional statistical education and with a Masters degree in multivariate analysis tucked under my arm, I embarked on a course which was to shake up most of my previous ideas on statistics as well as on life itself. It is impossible for anyone to spend two years as a student in Paris and, likewise, impossible for any statistician to spend two years in contact with the revolutionary Jean-Paul Benzécri without being radically affected as a consequence. It has taken me some time since those years of doctoral study in France to fully comprehend all that I learnt. This is an ongoing process, of course, and this book represents a milestone

of 10 years' personal experience of the statistical method the French call *analyse des correspondances*, which has been obviously translated as “correspondence analysis”» (GREENACRE 1984, v).

He then illustrates one of the reasons for the origin of his book (*fig. 2*), which is enlightening for our chronological and historical understanding:

«My reasons for writing this book were twofold. First, in 1980 I was invited to give a paper on correspondence analysis at an international conference on multidimensional graphical methods, called “Looking at Multivariate Data”, in Sheffield, England. There was considerable interest in my talk and I realized then, more than ever, the tremendous communication gap between Benzécri's group in France and the Anglo-American statistical school. I felt that it was almost my duty to undertake the writing of a book which would explain this important facet of French research to English speaking statisticians, using not only a language but a mathematical style familiar to them» (GREENACRE 1984, v).

The importance of Greenacre's book in bridging language barriers is also highlighted by Shizuhiko Nishisato, a member of the Japan school of Chikio Hayashi, whose career continued at the University of North Carolina, with a final move to the University of Toronto. Referring to the years when contacts among scholars mainly took place by correspondence or during international conferences, in the collective book *Modern Quantification Theory* (NISHISATO *et al.* 2021), the authors recall the meetings organized between 1970 and 1980 by French researchers at INRIA (Institut National de Recherche en Informatique et en Automatique) and other universities:

«Many of us in the English-speaking countries were for the first time exposed to the enormous amount of rich and informative French work on quantification theory. Those days, French researchers were in the frontier of quantification research, and it was typical to hear such a comment from French researchers as “it was already discussed in Benzécri's lectures” on papers presented by English speakers. At one conference, Brigitte Escofier presented a talk and Nishisato commented that “we investigated the same topic ten years ago in Toronto.” Guess what! Guttman, who was seated next to Nishisato, said “that was an excellent comment,” and shook his hands firmly. This episode would tell us how dominant the French group was then» (NISHISATO *et al.* 2021, 9).

Advances in computing and programming technology undoubtedly influenced the wider adoption of correspondence analysis. The first FORTRAN program for research purposes was written in 1964 by Brigitte Escofier-Cordier, at Rennes University on an IBM 1620, but it was not very portable (ESCOFIER 2003). For this reason, the Association pour le Développement et la Diffusion de l'Analyse des Données (ADDAD) was created. Starting in the 1980s, in addition to using high-level statistical packages like SPSS, SAS, or GENSTAT, LEBART and MORINEAU (1982) developed software tools like SPAD (Système Portable pour l'Analyse des Données), while GREENACRE (1986)



fig. 3 – Jan de Leeuw in an interview during the 6th CARME (Correspondence Analysis and Related Methods) conference (Rennes 2011) on the history of correspondence analysis.

introduced SIMCA (SIMple Correspondence Analysis). These programs simplified the implementation of correspondence analysis, making it accessible to a broader audience.

In this context, mention should be made of a group of researchers led by de Leeuw at the University of Leiden, who, between 1970 and 1990, produced many theoretical papers and computer programs on nonlinear multivariate analysis (*fig. 3*). In line with de Leeuw's point of view, the absence of models is a strong characteristic of the Dutch school, comparable to the theoretical approach of Benzécri's school – but without the economic resources available to Benzécri in France or Hayashi in Japan (DE LEEUW, MAIR, GROENEN 2016). This approach prefers to rely on geometrical considerations to present findings, with no reference to probabilistic models. The research group published their results under the pseudonym Albert GIFI (1980, 1990), first in Dutch and then in English, in a collective volume organized around a series of computer programs written in FORTRAN for correspondence analysis, principal component analysis, and canonical analysis, then embedded in the SPSS system.

In addition to the four research groups we have mentioned so far – described by de Leeuw as early as 1976 as the Japanese Quantification Theory Group, the French Correspondence Analysis Group, the Dutch Homogeneity Analysis Group, and the Canadian Group of Optimal/Dual Scaling – other countries have participated in the great adventure of multivariate analysis of categorical data. As with other research areas, it is interesting to reflect upon the epistemological and contextual factors that influenced the inclusion and transmission of new methods and their adoption across scientific communities before the advent of the Internet and the definitive establishment of the English language as the medium for the dissemination of research data. While Benzécri's geometric paradigm offered a rich theoretical foundation, its regional confinement raises questions about the inclusiveness and adaptability of methodological advances within a global scientific discourse.

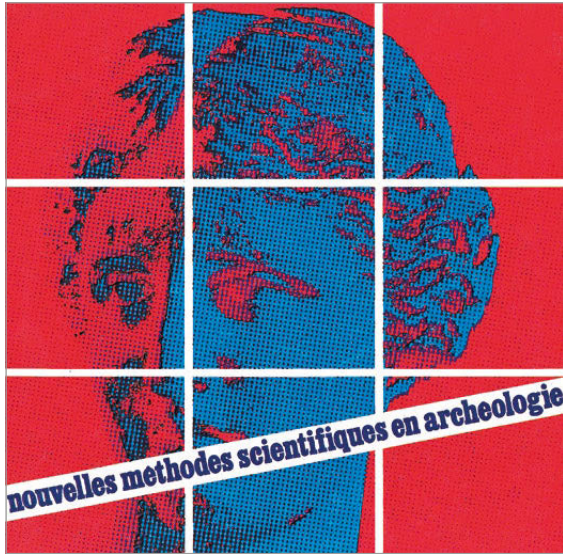


fig. 4 – The cover of the dossier on the automatic data processing in archaeology published in «Les dossiers de l'archéologie» (DJINDJIAN, LEREDDE 1980).

3.2 THE ARCHAEOLOGICAL APPROACH

The reconstruction of the historical roots of correspondence analysis in archaeology is certainly more straightforward, as its diffusion is widely recognized as having been started by the dossier published by François DJINDJIAN and Henri LEREDDE in 1980 (*fig. 4*). Roy Hodson's reconstruction is remarkably clear. In 1988, while presenting a series of statistical programs developed specifically for archaeology with Paul Tyers at University College London, Hodson describes correspondence analysis as follows:

«This procedure is related to Principal Components Analysis but summarises data using a Chi-square related metric rather than Euclidean distance (Greenacre 1984). Thus, when the units of analysis are assemblages, the size of assemblages as well as proportionate counts of variables is taken into account. A distinct attraction of this approach is the close relationship preserved between units and variables when the final result is presented. Following Benzécri (Benzécri 1973) the method has been much favoured in France (*cf.* Leredde & Djindjian 1980). IASTATS by no means exploits the full potential of this approach and simply provides a starting-point» (HODSON, TYERS 1988, 38).

Hodson's interest in data reduction methods is well known, as is his collaboration with David Kendall, who is widely recognized as the founding father of British probability and, in archaeology, as the originator of a new mathematical approach to seriation (KENDALL 1970). This is highlighted in an interview we conducted approximately ten years ago, excerpts of which can be found in the Virtual Museum of Archaeological Computing (MOSCATI 2018). Moreover, in a lively presidential address at the British Association for the Advancement of Science meeting in Liverpool in 1982 (later published in 1985



fig. 5 – James Doran during the unpublished interview we conducted in 2013 on the application of quantitative methods and Artificial Intelligence in archaeology.

in the article *Mathematical Statistics in the Humanities, and Some Related Problems in Astronomy*), Kendall recalled that he had spent the last twenty years of his research studies «acting as an informal advisor to practitioners of the humanities» (KENDALL 1985, 393).

Less widely acknowledged, however, is that Hodson's familiarity with Roger Shepard's and Joseph Kruskal's methods is also linked to his collaboration with James Doran (*fig. 5*). In an unpublished interview we conducted with him in the same years, Doran referred to his first encounter with Hodson:

«When I first encountered Roy Hodson, he was working at the Münsingen La Tène cemetery. He was deploying cluster analysis in order to analyse the artifacts, as well as k-means analysis. He was certainly getting advice from Peter Sneath, who was not an archaeologist, he was a numerical taxonomist [SOKAL, SNEATH 1963]. What I brought into their activity, really, was non-metric scaling, which had been developed in the United States by Roger Shepard and Joseph Kruskal, originally as a way of interpreting proximity data in order to establish the dimensions of conceptualization of human subjects. But it occurred to me that this kind of scaling technique could be used to examine sets of archaeological artifacts. So I wrote a program to implement the relevant algorithm in Algol 60. I talked with Roy and we ran one of his datasets. We published a paper in *Nature* based on those results [DORAN, HODSON 1966]. I think it was the first actual use of non-metric multidimensional scaling in archaeology, several years before it was widely promoted at the Mamaia conference. An important factor was that non-metric scaling did not use what you might call mathematical shortcuts in order to find the most appropriate distribution of items in space. Instead, it used a search procedure of a type I was familiar with from Artificial Intelligence work. It therefore in a way went beyond purely mathematical approaches and into a more computer science AI-oriented method».

Incidentally, in the broader framework of the history of the data analysis movement in the second half of the 20th century, it is no coincidence that many authors (e.g., MURTAGH 2008; BEAUDOUIN 2016; HOLMES, JOSSE 2017) noted that Benzécri was

well acquainted with Shepard's work (SHEPARD 1962) and spent some time visiting him at Bell Telephone Laboratories in the mid-1960s. Moreover, Kendall experimented with the multidimensional scaling (MDSCAL) program developed by Shepard and Kruskal (KRUSKAL 1964) for psychological research to construct the chronological ordering of the La Tène cemetery at Münsingen-Rain (KENDALL 1969; for a critical assessment, see CLARKE, RIVETT 1978).

Let us return to archaeology. The Djindjian-Leredde dossier, as well as the article by DJINDJIAN and VIGNERON (1980), represented a significant moment in the dissemination of innovative data analysis methods within the archaeological community – methods that had already been statistically discussed in the previous decade (DJINDJIAN 1977a, b; DELPORTE, MAZIÈRE, DJINDJIAN 1977). Another frequently cited contribution to the popularization of correspondence analysis in archaeology is the article by Erik BØLVIKEN *et al.* (1982) from the University of Tromsø, Norway, published in the widely distributed journal «World Archaeology». This article compares different archaeological assemblages from excavations in northern Norway using correspondence analysis. Bølviken's article is undoubtedly a methodological reference tool, although it seems somewhat isolated from the broader archaeological discourse, as the author states that, to his knowledge, the examples presented in Hill's paper are the only published applications of this method.

Following the major public events of the 1970s, which were promoted in France by Jean-Claude GARDIN (1970) and in Romania by Roy HODSON, David KENDALL, and Petre TAUTU (1971), in the 1980s a select number of regularly convened conferences became the primary channels for the dissemination and critical assessment of quantitative methods (MOSCATI 2015). Two particularly notable examples stand out for their continuity and influence: the 'Computer Applications and Quantitative Methods in Archaeology' (CAA) conferences, launched in 1973, and the meetings organized by Commission IV, 'Archaeological Methods and Theory', of the Union Internationale des Sciences Préhistoriques et Protohistoriques (UISPP), established in 1976 (*fig. 6*).

The 1980s also saw an increased adoption of correspondence analysis in archaeological research. This trend was closely linked to the broader diffusion of techniques for quantifying multivariate categorical data and the effectiveness of exploratory data analysis in addressing complex archaeological questions. With specific reference to correspondence analysis, it proved especially valuable due to its capacity to simultaneously analyze both quantitative and qualitative variables, as well as both individuals (Q-method) and variables (R-method). This methodological flexibility made it an attractive tool for archaeologists seeking to enhance data interpretation and pattern recognition within their datasets.

Early handbooks dedicated to statistical applications in archaeology began to include discussions on correspondence analysis, although not all explicitly credited its development to the French school of data analysis led by Benzécri. In the UK, the handbook by DORAN and HODSON (1975) served as the primary reference for the quantitative approach to archaeological research for a long period. Clive ORTON (1980) was among the first to write a book on the subject, *Mathematics in Archaeology*, whose highly

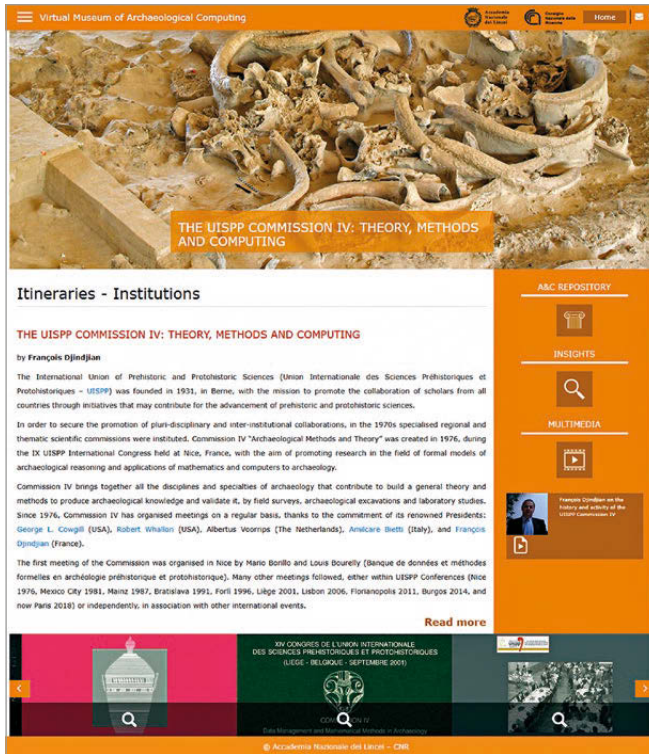


fig. 6 – The multimedia cultural itinerary of the Virtual Museum of Archaeological Computing dedicated to ‘The UISPP Commission IV: Theory, Methods and Computing’ (DJINDJIAN 2017).

original content structure addressed key archaeological questions (What is it? How old is it? Where does it come from? What was it for?). Correspondence analysis does not appear among the methods of multivariate analysis; however, a few years later, Stephen SHENNAN (1988) dedicated a short section to it. Apart from a superficial reference to Djindjian’s work, Shennan’s introduction to the technique is certainly influenced by Bølviken’s research. Shennan himself admits that «the relative slowness of its adoption by Anglo-American quantitative archaeologists must be ascribed to a certain lack of communication between them and the rather different French archaeological world» (283).

A considerable number of contributions written for an archaeological readership also came from non-Anglophone countries, reflecting the method’s international reach. In Germany, Peter Ihm’s introduction to statistical applications in archaeology included a detailed discussion of the distinction between Q- and R-methods (IHM 1978; see also 1983). In Italy – where the use of quantitative graphs began between the end of the 1950s and the beginning of the 1960s (for a review, see GUIDI 2015) – Amilcare Bietti’s (*fig. 7*) key handbook on mathematical techniques for archaeological data analyses (BIETTI 1982), based on a series of lectures given at the Accademia Nazionale dei Lincei in 1978 and 1980, dedicated a section to multivariate analyses, including correspondence analysis and its application to the study of the Paleolithic site of La Ferrassie (Dordogne). Drawing inspiration from the teachings of Bietti and Djindjian, I devoted a chapter to



fig. 7 – Amilcare Bietti, Italian physicist and expert in prehistoric archaeology, during an interview at the Istituto Nazionale di Fisica Nucleare (INFN) (<https://lamediateca.infn.it/mediateca/view.php?v=50>).

correspondence analysis in a book that broadly explored the applications of computers in archaeological research (MOSCATI 1987a). This chapter underscored the potential of factor analyses for the study of artifacts produced in historical contexts (see Chap. 5 below). In Scandinavia, after Bølviken's contribution, Torsten Madsen played a significant role in promoting the method, both in a monograph (MADSEN 1985) and a collective volume on multivariate numerical approaches in Scandinavian archaeology (MADSEN 1988). Correspondence analysis was particularly prominent in the latter, appearing in eight out of eleven articles. The role of André Gob in Belgium is also significant, particularly within the DIAPRE group, responsible for founding the journal «Archeo-LOG» (GOB 1986, 1988). It should also be noted that data analysis spread to Poland and elsewhere in Eastern Europe (SCHILD 1979; BURDUKIEWICZ 1981; DOLUKHANOV, KOZŁOWSKI, KOZŁOWSKI 1980).

These early contributions established correspondence analysis as an important tool in archaeological research, demonstrating its potential for structuring and interpreting complex datasets. In France, the most accurate synthesis of data analysis techniques was produced by DJINDJIAN (1991) in his book *Méthodes pour l'archéologie*. In England, the task was undertaken by Mike BAXTER (1994) in his book *Exploratory Multivariate Analysis in Archaeology*. Baxter devotes two chapters to correspondence analysis ('Correspondence Analysis - The main ideas' and 'Correspondence Analysis - Extensions'), providing a comprehensive bibliographic apparatus, in which linguistic barriers seem to be partially overcome, and making intelligent use of previous bibliographical accounts (see, e.g., DJINDJIAN 1989).

Baxter's analysis is meticulous, and his categorization of the bibliography highlights the main areas of application of correspondence analysis: a) Analysis of chemical com-

positions of artifacts; b) Typological/morphological analysis; c) Assemblage comparison; d) Spatial analysis; e) Expository and review articles and edited collections. Adopting a bibliometric analysis, the author shows that only 10% of the references precede 1980. French and Scandinavian scholars account for about two-thirds of the articles. British usage increased after Ringrose's first detailed application, which was also the first to appear in the CAA proceedings (RINGROSE 1988), one year before the already cited article by DJINDJIAN (1989), 'Fifteen years of Contributions of the French School of Data Analysis to Quantitative Archaeology'. Few papers provide a technical explanation of the methodology, and Baxter asserts that, in the absence of user-friendly statistical packages, correspondence analysis is still considered a 'black box' process. Assemblage comparisons dominate its applications, and chronological seriation is a common objective; typological and morphological studies of archaeological artifacts – particularly lithic productions and funerary vessels – follow. Spatial analysis, however, is still under-represented.

3.3 THEORETICAL AND PRACTICAL DEVELOPMENTS OF DATA ANALYSIS TECHNIQUES

The 1990s marked the beginning of the era of the 'middleman', as advocated by Albertus VOORRIPS (1996). This figure emerges «inside the culture of archaeology, has learned some of the language and culture of information science, alienates himself more or less from his archaeological culture and then functions as a channel through which information and goods are exchanged between both cultures». The language of the 'middleman' facilitates interaction between different disciplines; in the realm of archaeological computing, it consists of mathematical models applied to archaeology.

Following the successful period of 'quantitative archaeology', a systematic review of the literature on computer applications in archaeology reveals a contraction in mathematical and statistical data analysis techniques, or rather, their natural integration into archaeological methods. This might seem to be at odds with the copious production of handbooks in the early 1990s; however, this output was a result of the expansion of the preceding years and the fact that it largely originated from the Anglo-Saxon academic sphere, where university courses and seminars have trained researchers from the humanities to incorporate statistical methods into archaeology.

Within this panorama, a notable contribution is Djindjian's handbook. With a strong theoretical engagement, it delineates the successive steps of the archaeological process, encapsulated in what he describes as the *revanche de l'archéologue* in response to technological advances. Consequently, the *Analyse des Données*, in which the French scholar was trained and which he helped to popularize, is not limited to a single chapter; instead, it permeates several sections of the volume, encompassing typological studies of groups of artifacts, spatial analysis of emerging structures, and chronological ordering of archaeological deposits.

To the best of my knowledge, in the 1990s, only one collective volume specifically devoted to correspondence analysis was published: the book edited by Johannes MÜLLER

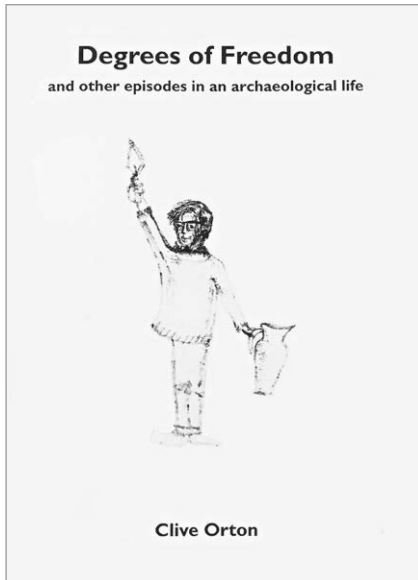


fig. 8 – Cover of the autobiographic limited edition volume published by Clive Orton in 2017 and focusing on his ‘archaeological life’ (ORTON 2017).

and Andreas ZIMMERMANN (1997), commemorating the seventieth birthday of Peter Ihm. Statistical techniques continued to be employed, facilitated by the dissemination of more accessible personal computers and specialized software. However, applications often focused on narrowly defined areas, typically limited to those with the oldest traditions: the analysis of the chemical composition of archaeological materials to ascertain their provenance and production workshops; typometric and morphometric studies aimed at classifying objects; and seriation as a method of chronological ordering. Another significant area of investigation was spatial analysis, which enjoyed a revival (see the pioneering book by HODDER and ORTON 1976), linked to GIS applications in archaeological research (FOTHERINGHAM, ROGERSON 1994).

The growing reliance on Information and Communication Technology tools for documentary research and for contextual analysis of ancient cities and landscapes has obscured the statistical approach, which has diminished in prominence with regard to its role as the focus of targeted theoretical discussions. As highlighted at the 1993 UISPP Commission IV meeting in Australia, the debate – spanning processual, post-processual, and cognitive archaeology – significantly influenced methodological approaches and the transition from data processing to information processing (COWGILL 1994; JOHNSON 1994; READ 1994; RENFREW 1994). In the interplay between archaeology and computer science, this cognitive shift does not merely reflect the study of past ways of thought inferred from material remains, as envisioned by Colin RENFREW (1982). Instead, it embodies the *pensée réfléchie*, or the logicist analysis of interpretative constructs in archaeology, a concept clearly illustrated in the journal «Archeologia e Calcolatori» by GARDIN (1996).

Two key aspects emerge within the statistical approach: the application of sampling techniques to assist in data selection and the effort to integrate different methodological

approaches to interpreting archaeological data. Notably, Orton (*fig. 8*) articulates these expectations by stating: «If we are to make progress with the application of quantitative methods in archaeology in the 21st century, we need to pay more attention to the quality of our data, and to ensuring the compatibility of data from different sources... Otherwise we run the risk of building on insecure foundations» (ORTON 2010, 259). This emphasis on data integration has also led to the increasing adoption of Bayesian statistical methods, which provide a formal framework for updating initial hypotheses based on new observations. By transitioning from prior to posterior probabilities, Bayesian methods of inference enable the iterative reassessment of data in the light of new knowledge (BUCK, CAVANAGH, LITTON 1996; VOORRIPS 1996; and, more recently, COWGILL 2002).

3.4 A LITERATURE-BASED ANALYSIS: THE CASE OF «ARCHEOLOGIA E CALCOLATORI»

From the last decade of the 20th century into the first quarter of the 21st century, the evolution of data analysis methods in archaeology can be traced through the journal «*Archeologia e Calcolatori*», one of the earliest scholarly publications dedicated entirely to computer applications in archaeology and among the most long-lived in the field. It provides insights into the historical development of archaeological computing as an independent discipline and serves as a key resource for understanding the advancement of digital archaeology. Furthermore, it stands out as a rare example of a publication that reflects the international panorama with its commitment to multilingualism.

The quantitative distribution of article topics from 1990 onwards illustrates the persistence of certain themes, the emergence of new research directions, and the decline of others, highlighting a continuously evolving research area influenced by technological advances. The sample is homogeneous in nature, enabling readers to trace the history of computer applications by examining the practices across various fields over time (*figs. 9-10*). Following a significant increase in applications in the first decade of publication, there has been remarkable consistency in the use of mathematics and statistics since the journal was founded thirty-five years ago. The number of articles focusing on the use of statistical techniques – where these methods play a central role in the analysis rather than merely serving as a complementary research tool – are 120 in total.

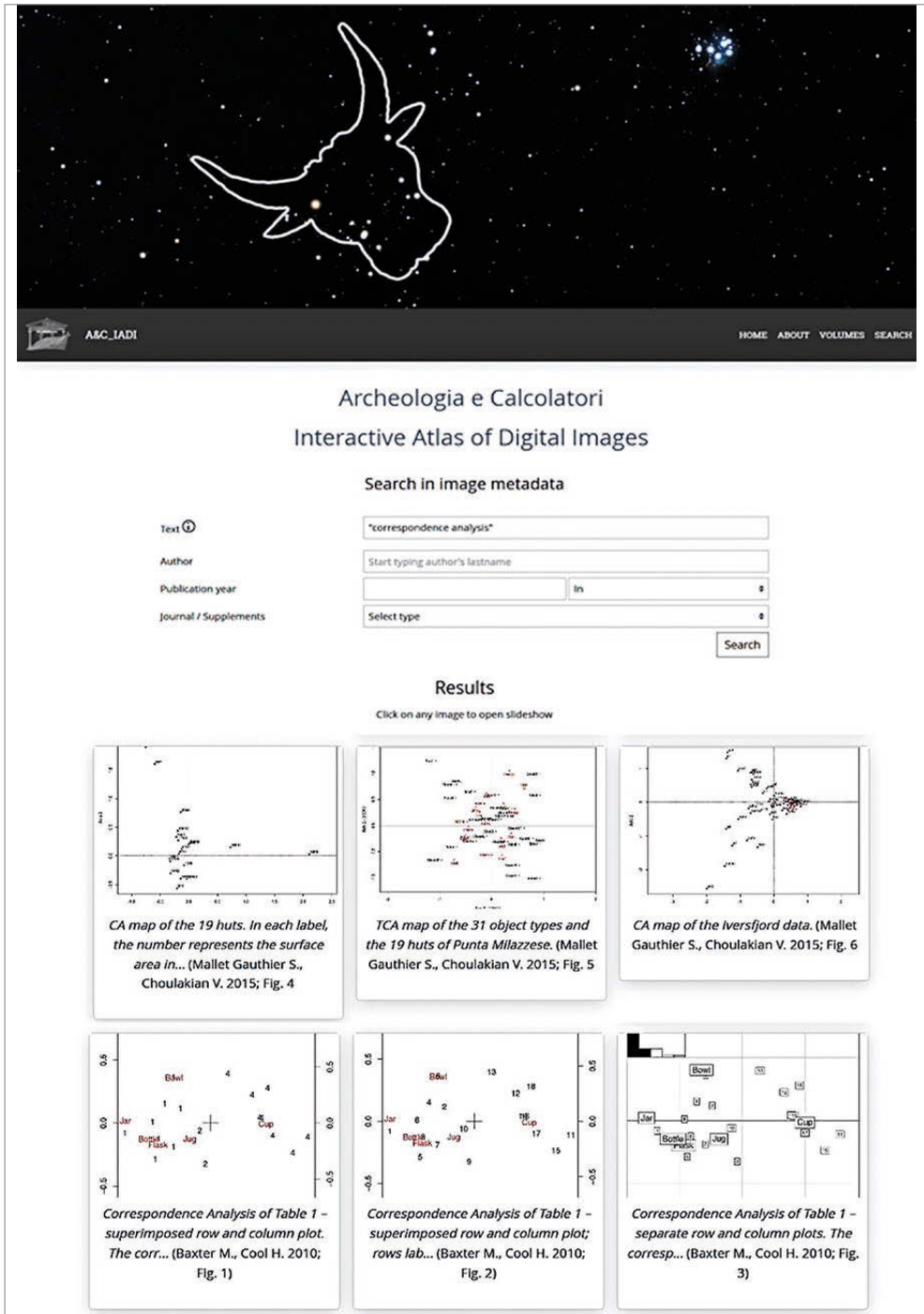
Of these articles, 55 were published before 2000 (46%), 31 from 2000 to 2009 (26%), 23 from 2010 to 2019 (19%), and 11 from 2020 to 2024 (9%). Statistical methods are therefore continuing to attract the attention of archaeologists, although this result contrasts with a different prediction from a recent study leveraging various Machine Learning techniques. The case study concerned the articles published in «*Archeologia e Calcolatori*» in the decade from 2011 to 2020, using the publications in the CAA conference proceedings and journal as a benchmark. Titles and abstracts were analyzed through a supervised classification approach into the subfields of computer science, based on the ACM's taxonomy, and by applying topic modelling techniques, Named Entity Recognition, and geotagging techniques. By exploring digital archaeological



fig. 9 – The homepage of the renovated website of the journal «Archeologia e Calcolatori» (<https://www.archcalc.cnr.it/>). The Search page offers the option of advanced queries of the published articles and connects to the Interactive Atlas of Digital Images (A&C_IADI).

topics within the broader context of Heritage Science (fig. 11), the study anticipated a decline in statistical applications (CARVALE *et al.* 2023).


Cross-referencing the aforementioned frequency distribution data with the most represented sectors of archaeological research shows that quantitative data analysis is the primary subject of a fair number of theoretical and methodological articles (14%), including a specific review of applications (MOSCATI 2000, 2003). This evidence increases when articles that address data encoding issues (e.g., ORLANDI 1993; CAMIZ 2004) are included, or those discussing new software tools, such as R, that facilitate dissemination and education (e.g., BAXTER, COOL 2010; ALBERTI 2013). In terms of application areas, apart from specific epigraphic study areas (MARCHESINI 2004, and a synthesis in ROSSI 2020), multivariate analyses are associated with the categories ‘Classification of Archaeological Finds’ in 39% of cases, ‘Survey and Excavations’ in 24%, and ‘Archaeometry’ in 16%. The fields identified by Djindjian over 35 years ago are thus still relevant for data analysis: typology, spatial analysis, seriation, culture pattern studies, and provenance studies.



The screenshot displays the homepage of the 'Archeologia e Calcolatori' Interactive Atlas of Digital Images. At the top, there is a header with a starry background and a white outline of a longhorn. Below the header, the site title 'Archeologia e Calcolatori' and subtitle 'Interactive Atlas of Digital Images' are centered. A search bar is present with the text 'correspondence analysis' entered. To the left of the search bar are filters for 'Text', 'Author', 'Publication year', and 'Journal / Supplements'. Below the search bar, a 'Search' button is visible. The 'Results' section contains a link to 'Click on any image to open slideshow' and six thumbnail images of correspondence analysis plots. Each thumbnail includes a caption with the plot's title and reference information.

Archeologia e Calcolatori
Interactive Atlas of Digital Images

Search in image metadata

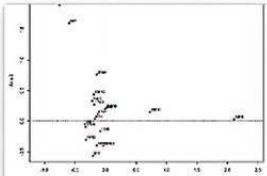
Text 
Author
Publication year
Journal / Supplements

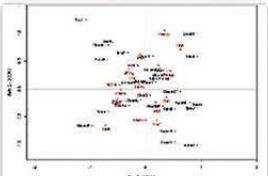
Search in image metadata

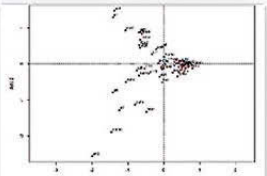
Select type

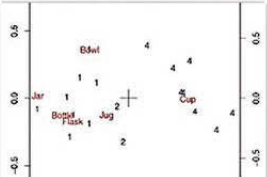
Results

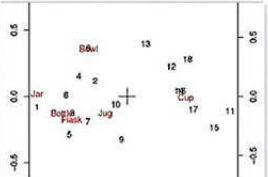
Click on any image to open slideshow


CA map of the 19 huts. In each label, the number represents the surface area in... (Mallet Gauthier S., Choulakian V. 2015; Fig. 4)


TCA map of the 31 object types and the 19 huts of Punta Milazzese. (Mallet Gauthier S., Choulakian V. 2015; Fig. 5)


CA map of the Iversfjord data. (Mallet Gauthier S., Choulakian V. 2015; Fig. 6)


Correspondence Analysis of Table 1 – superimposed row and column plot. The corr... (Baxter M., Cool H. 2010; Fig. 1)


Correspondence Analysis of Table 1 – superimposed row and column plot; rows lab... (Baxter M., Cool H. 2010; Fig. 2)

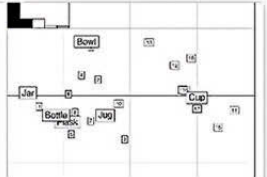

Correspondence Analysis of Table 1 – separate row and column plots. The corresp... (Baxter M., Cool H. 2010; Fig. 3)

fig. 10 – A&C_IADI - Interactive Atlas of Digital Images (<https://iadi.archcalc.cnr.it/>) is an open access portal that showcases 30 years of the journal's history through its images.

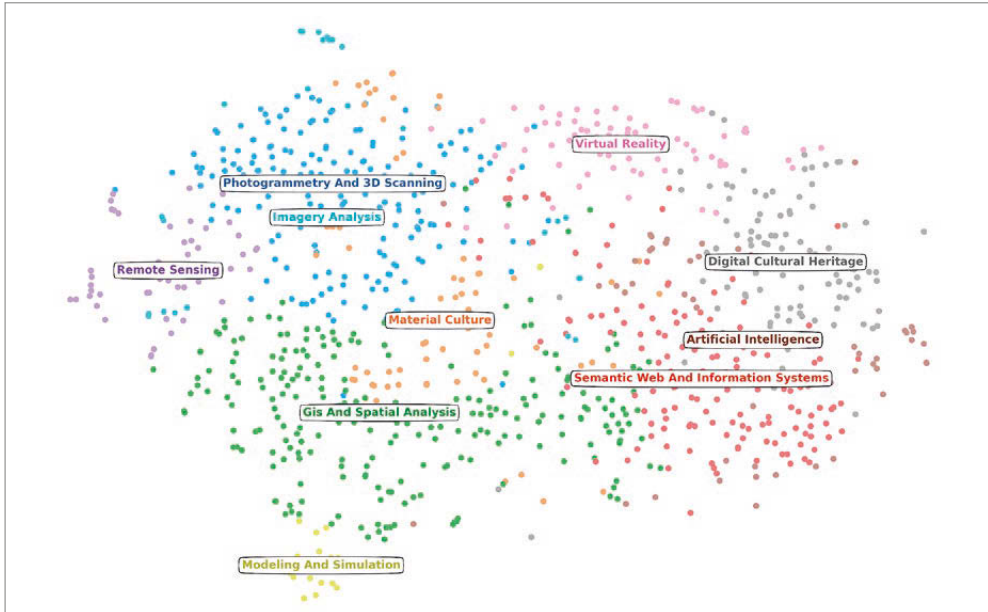


fig. 11 – Emerging topics in Digital Archaeology and Heritage Science domains: results of the topic modelling as applied to the journal «Archeologia e Calcolatori» and the Proceedings of the CAA conferences (CARVALE *et al.* 2023).

Focusing on specific mathematically oriented topics, the application of correspondence analysis – one of the techniques whose application has strongly characterized the history of «Archeologia e Calcolatori» – is explored through various geometric approaches in the articles published in the journal. These include, besides Simple Correspondence Analysis and Multiple Correspondence Analysis, Detrended Correspondence Analysis (e.g., BAXTER, COOL 2010), Taxicab Correspondence Analysis (e.g., MALLET GAUTHIER, CHOULAKIAN 2015), and Procrustes Analysis (e.g., DI LUDOVICO, CAMIZ 2020, 2022). These methods are not used in isolation but are invariably associated with other exploratory multivariate techniques, particularly cluster analysis, principal component analysis, and discriminant analysis.

Three principal archaeological themes appear to dominate. First, the classification of archaeological artifacts based on their typology and provenance remains a central focus, encompassing a diverse range of artifacts, from Mesopotamian cylinder seals (ROVA 1995; see Chap. 6 for references), to Attic pottery, as well as to corpora of bronze and stone artifacts (see Chap. 5 for references). Second, spatial analysis investigates the distribution of sites across landscapes and of artifacts within them, contributing to the reconstruction of ancient settlement patterns and demographic dynamics, involving the assessment of artifact density, spatial clustering, and associations with specific economic activities (BIETTI 1993; DJINDJIAN 1999). The applications cover a variety of chronological and spatial contexts (*figs.* 12-13), ranging from the analysis of pottery dumps in a Neolithic layer at

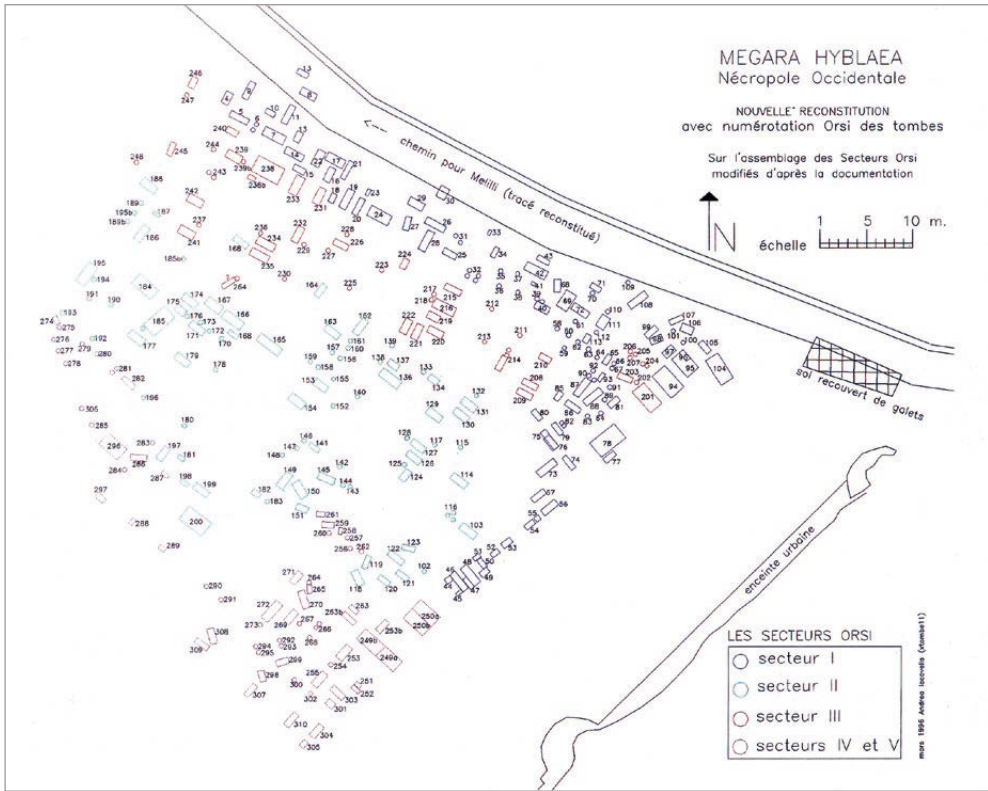


fig. 12 – The plan of the western necropolis of Megara Hyblaea. The numbering of the tombs follows the system adopted by P. Orsi and F.S. Cavallari during excavations at the end of the 19th century (IACOVELLA 1997).

the site of Chalain, Jura, France (GILIGNY 2008), to the study of graves distribution in the archaic western necropolis of Megara Hyblaea (IACOVELLA, AUDA 1994; IACOVELLA 1996), to the mapping of historical archaeological sites in the Lagoon of Venice, Italy (CANAL, CAVAZZONI 1990), as well as to a recent and innovative educational application focused on quantifying the rate of accidental ceramic discard at the excavation site of Siponto, Northern Apulia, Italy (RAGNO 2023). Finally, archaeometric studies – where the application of principal component analysis consistently remains the most widely applied method – focus on specific material categories, including vegetal and faunal remains, lithic assemblages, and ceramic artifacts.

This continuity over time confirms Baxter's findings in his review of the literature prior to 1994, which unfortunately lacked complete coverage of Mediterranean countries (BAXTER 1994, 279-284). However, in «Archeologia e Calcolatori», in all likelihood due to its distinctive chronological perspective, the classification of archaeological finds places particular emphasis on the study of figurative representations, while research on ancient settlement systems highlights the role of cities and cemeteries.

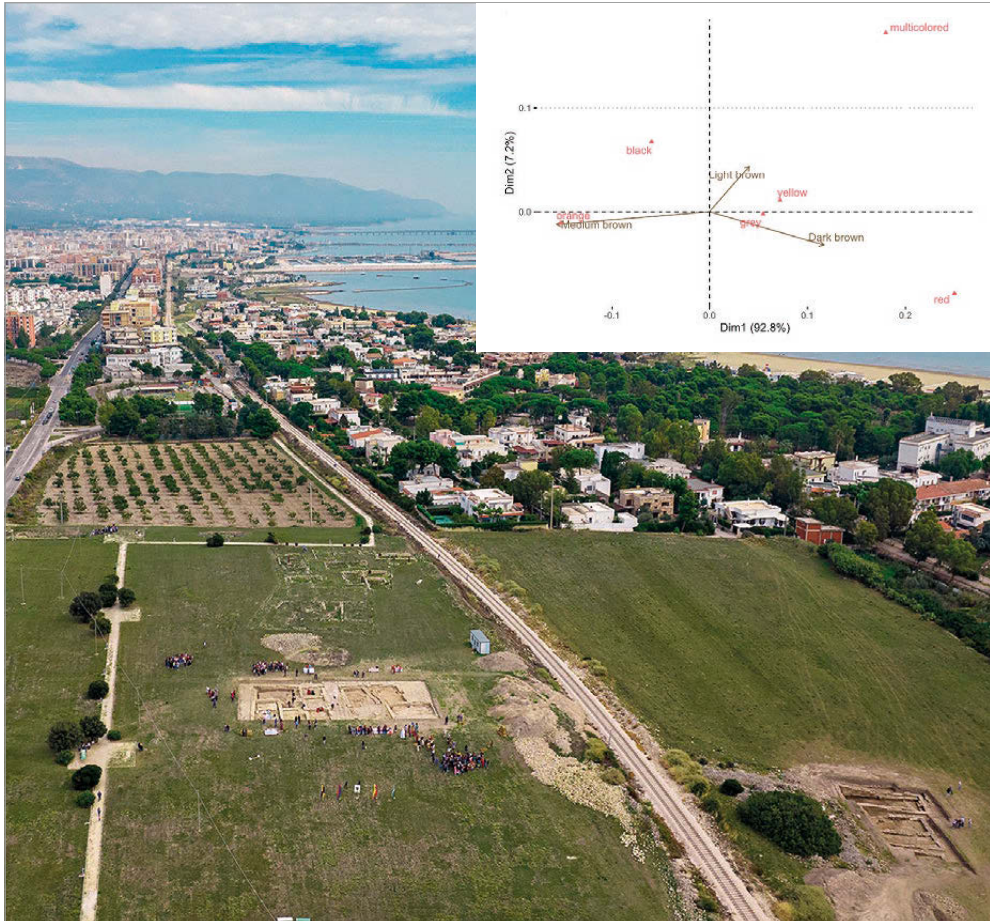


fig. 13 – Excavations at the archaeological site of Siponto, Italy, the focus of a recent application of correspondence analysis to assess the influence of soil color on the most discarded sherds.

In conclusion, the journal's systematic review underscores that data analysis – primarily applied within the domain of prehistoric archaeology, though its adoption has expanded across various areas of historical archaeology – remains a crucial component of the discipline as it adapts to new digital tools and methodologies in the emerging era of Data Science (see Chap. 7). In the new millennium, the widespread dissemination of R has significantly enhanced the application of correspondence analysis in the humanities by providing a free, open-source platform that includes specialized packages designed to simplify both analysis and graphic outputs. The extensive documentation and active user community surrounding R have further improved the accessibility of statistical methods. Additionally, R's flexibility and graphic interface facilitate integration with other quantitative techniques, thereby fostering interdisciplinary research in fields such as archaeology, history, and linguistics.

François Djindjian

4. DATA ANALYSIS AND PREHISTORIC ARCHAEOLOGY

A main impression is that a positive change is going on, even if slowly, in communications between mathematicians-statisticians and archaeologists-historians. Difficulties today will be reduced tomorrow, by the coming of a new generation of archaeologists, trained to be better prepared for joint tasks. Teaching is a key (MOBERG 1971, 561).

4.1 THE INFLUENCE OF DATA ANALYSIS ON THE EVOLUTION OF ARCHAEOLOGICAL METHODS

Data analysis techniques have significantly contributed to almost all archaeological methods since the 1970s. By enabling archaeologists to avoid the detailed computations of elementary or classical statistics, these techniques have allowed them to build a more formalized methodology progressively.

The following pages are devoted to the historiography of data analysis applications in prehistoric and protohistoric archaeology. These applications are described according to the archaeological issues addressed, highlighting the role of the scholars who have participated in this methodological and cognitive revolution over time. Emphasis has been placed here on earlier periods, the era of the pioneers (1966-1972), the era of the great development (1975-1985), and the era of its apogee (1985-1995). Of course, the story continues to this day. However, the use of data analysis has become commonplace, and published articles integrate the results of these analyses into scientific discourse in the same way as they do with histograms, radiocarbon dating, or physico-chemical composition data.

Typometry

Empirical typologies, matrix analysis, and attribute association tests (SPAULDING 1953) all resulted in the first cluster analysis applications (HODSON, SNEATH, DORAN 1966). From 1966 to 1975, DORAN and HODSON (1975) tested different Q-mode data analysis techniques: multidimensional scaling, principal component analysis, k-means analysis, and discriminant analysis. The fundamental role of the variables was emphasized by DJINDJIAN (1976a, 1977a) who introduced the simultaneous use of the R+Q mode correspondence analysis and a variance-based ascending hierarchical cluster analysis, on tables of raw data transformed by disjunctive coding (0, 1). The search for

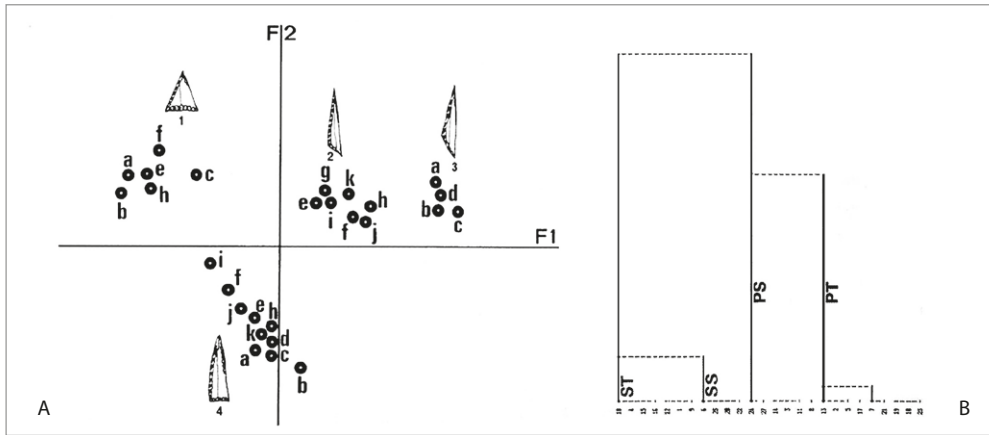


fig. 1 – Mesolithic points found in the Paris Basin (DECORMEILLE, HINOUT 1982).

an impossible exhaustive description of artifacts was abandoned in favor of the definition of a minimum homogeneous intrinsic description (DJINDJIAN 1980b), improving the emergence and stability of the clustering structures, and allowing extrinsic and intrinsic interpretations of the clusters.

The biggest difficulty to be overcome was formalizing the description of the artifacts, the quantification of the measures and the coding of the data, particularly with a mix of quantitative and qualitative variables when transformed by a complete disjunctive coding (0, 1). The high inertia of the size effect produced a parasite structure in the first axes, hiding the searched structures pushed back in the far axes. The parasite structure is a Gutmann effect produced by the correlation of the modalities of quantitative measures coded (0, 1) due to the presence of a size effect. Thus, the question to be solved was the separation of the size effect from the other structures by adequate coding.

A pioneering case was the typology of Noailles burins (DJINDJIAN 1977a): the results of the analysis showed the differentiation between burins on a truncation and Noailles burins, the size effect from small to large Noailles burins, and, more interestingly, the different manufacturing sequences of the Noailles burins.

Most of the other applications were not successful for the reasons explained above, as in the case of the Bronze Age axes (VUAILLAT, MASSONIE 1974), or the unsuccessful simulation of distinguishing teapots from coffee pots (CAHEN, MARTIN 1972), due not to the algorithm's weakness, as suggested by the authors, but to the description and its coding.

However, the difficulties to be overcome were too great for archaeologists without any training in how the algorithms and statistics worked, and after the initial enthusiasm, most of them decided to give up. The results were even more disappointing for those who directly used cluster analysis, whose tree also varied with the choice of similarity indices or distances and aggregation criteria, which masked the effect of the encoding, resulting in parasitic chain effects.

The ‘debitage’ has been less studied (see next section) despite its interest as revealed by the work of GUILBAUD (1987) on the Aurignacian and Chatelperonian debitage of Saint-Cézaire. The lower Palaeolithic industry has been studied by DIVE (1985).

An alternative and easier way was to process the shape separately and to find the appropriate description and coding to achieve efficient clustering. The application of separate processing of different intrinsic descriptors, with their respective coding, was proposed (*fig. 1*) by DECORMELLE and HINOUT (1982) in a concept of multiple typological analysis (DJINDJIAN 1991).

Shape analysis and manufacturing process

Shape analysis (or pattern recognition) concerned archaeology, for example for the clustering of Acheulean handaxes, Mesolithic points, or ceramics in archaeology, as well as paleoanthropology for morpho-cranial measures of human skulls, and more extensively paleontology.

In the UK, beginning in 1975, archaeologists such as WILCOCK and SHENNAN (1975) tested the efficiency of the mosaic method and of the sliced method, while in France ratio-based methods were initially tested (DJINDJIAN, DE CROISSET 1976; GUENOCHÉ, TCHERNIA 1978; MOHEN 1980; GILIGNY 1990).

Different descriptions and codes were evaluated (DJINDJIAN 1985a), which have shown the main influence of the following three factors:

- Variance, as well as the discriminant efficiency of coding, for a given type of shape;
- Ability of data analysis techniques (principal component analysis, correspondence analysis), applied to a given coding, to separate the size factors from the shape factors;
- Need to adapt a coding iteratively, instead of searching for an optimal coding.

The first question to be asked is whether the size effect is part of the clustering, or if it is a parasite structure to be eliminated either by coding (for example a ratio), by processing (the metrics), or by factorization (a size factor). The slice method on raw measures or on ratio measures also proved to be the most efficient and most open coding (*fig. 2*). The first use of automated digital measure tools (KAMPFMEYER 1986) already revealed the future of 3D automated recording more than twenty years after. It has now become easy, through digital stereophotogrammetry, to quickly build a collection of artifacts, and to conduct pattern recognition of them.

The description of the manufacturing processes of the lithic industry is essential in ancient and recent prehistory (otherwise incorrectly referred to as the *chaîne opératoire*). The complete formalization of the description and quantification of these manufacturing processes was achieved by the implementation of a graph, the Petri net, which models the sequence of manufacturing tasks (DJINDJIAN 2013) and leads to the construction of a Burt table, which can be processed by data analysis (*fig. 3*).

The reference application of this approach is the technical evolution of burin manufacturing in the Early and Middle Upper Palaeolithic of Europe (DJINDJIAN 1985c, 1993-94).

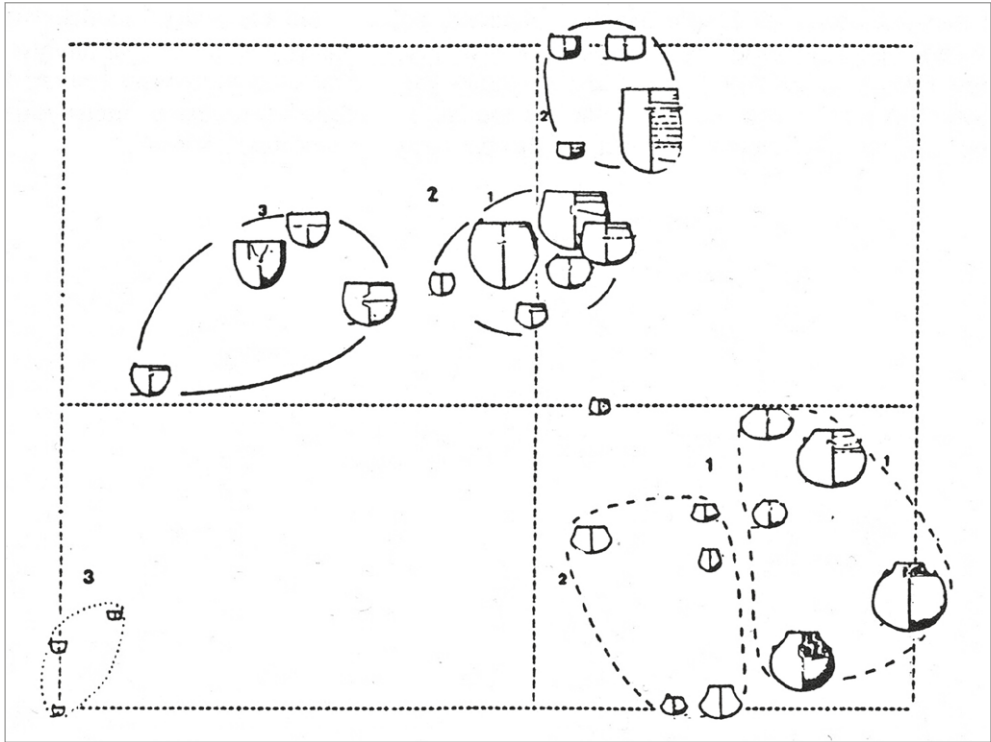


fig. 2 – Shape analysis of ceramics from the Chalcolithic settlement of Bousargues in the Hérault (GILIGNY 1990).

Intrasite spatial analysis

The earliest methods of intrasite spatial analysis in archaeology were the one-dimensional and two-dimensional techniques of quantitative ecology as single spatial distribution analysis (e.g., Nearest Neighbor Analysis - NNA test) and association tests between pairs of spatial distributions (e.g., permutation test). These methods were not well-suited to identifying dwelling structures and working activities, which are intrinsically multidimensional, based on the concept of a local multivariate density measure.

Initial attempts were made by JOHNSON (1977) and GRAHAM (1980), who proposed an R-mode multivariate analysis applied to an NNA-like similarity matrix. Later, KINTIGH and AMMERMAN (1982) used cluster analysis directly on the (x, y) coordinate matrix of each spatial distribution, but the superposition of clusters did not provide a true multivariate intrasite spatial analysis. HESSE (1984) and BOUCHET (1986) used correspondence analysis directly on the counting cell matrix.

Subsequently, WHALLON (1984) developed an 'unconstrained clustering' method based on a cluster analysis applied to a multivariate local density matrix, after having smoothed the spatial distributions. An improvement to this approach was proposed by DJINDJIAN (1988), who combined the use of correspondence analysis (as R+Q mode

HODSON 1966), principal component analysis (HODSON 1969), canonical analysis (GRAHAM, ROE 1970), and constellation analysis (AZOURY, HODSON 1973) are all Q-mode techniques.

Correspondence analysis, when associated with hierarchical cluster analysis, has been applied as R+Q mode analysis with the chi-square metric, which is well adapted to abundance tables. This combination enables the determination of the role of different types within structures, such as partitions (facies) and seriation (evolution).

Applications of correspondence analysis have been numerous for the Palaeolithic, Mesolithic, and Neolithic periods in Western Europe and the Middle East: DJINDJIAN (1976, 1977b, 1980b), HOURS (1976), LAPLACE and MERINO (1979), BERGOUGNAN and MOHEN (1980), DECORMEILLE and HINOUT (1982), BØLVIKEN *et al.* (1982), SLACHMUYLDER (1985), DJINDJIAN (1986a, 1988, 1993a, 1993b, 1996), and BOSSELIN and DJINDJIAN (1988, 1997, 1999). Principal component analysis has also been applied by DOLUKHANOV, KOZŁOWSKI and KOZŁOWSKI (1980), CALLOW and WEBB (1981), and BIETTI, BURANI, ZANELLO (1985).

In cases where it is better to avoid the step of building a typology, the method has been extended to R+Q mode analysis on attribute or Burt tables (DJINDJIAN 1980a), and integrated into a general methodology for intrinsic and extrinsic interpretations of culture facies (DJINDJIAN 1980b, 1985c).

Finally, the use of elementary elements in data analysis – i.e. the elements (individuals or variables) projected into the multidimensional space without participating in its

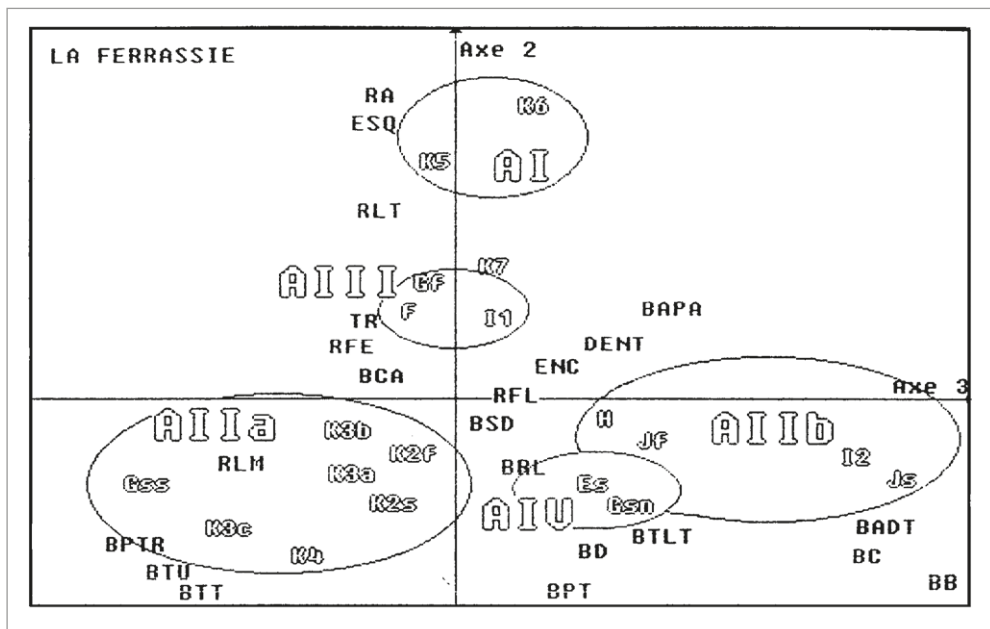


fig. 4 – Identification of the cultural facies of the Aurignacian through a technological description obtained by cognitive iteration of a typological list (DJINDJIAN 1993a).

construction, thus performing a factorial regression – allows an iterative learning process to be undertaken on both the variables. This process improves both the suitability of the types or subtypes, enabling them to be redefined, and the individuals, allowing them to be clustered (*fig. 4*).

The results of these applications have been an updating of the chrono-stratigraphy and the sequence of the cultural facies of the Upper and Final Palaeolithic in Europe, as well as the Mesolithic, and the explanation of the variability of the lithic industry of the Middle Palaeolithic, revealing the main factors at the origin of that variability.

Seriation

Ordering techniques, whether based on a direct (0, 1 tables) or an indirect (similarity table) processing, have been numerous – from BRAINERD (1951) and ROBINSON (1951) to ÉSTER (1981) – showing the interest of archaeologists (and other scientists) in seriation algorithms.

Data analysis techniques have been introduced in seriation by KENDALL (1971) with a multidimensional scaling algorithm applied to a contingency table modified through a special ‘circular product’ transformation, which revealed a horseshoe structure: the Horshu program.

In parallel, the reciprocal averaging method was successfully applied by GOLDMANN (1973) in Germany, WILKINSON (1971) in the UK, and LEGOUX (1980) in France. However, the theoretical similarity with the travelling salesman problem (WILKINSON 1971), well-known in Operational Research, did not lead to real developments.

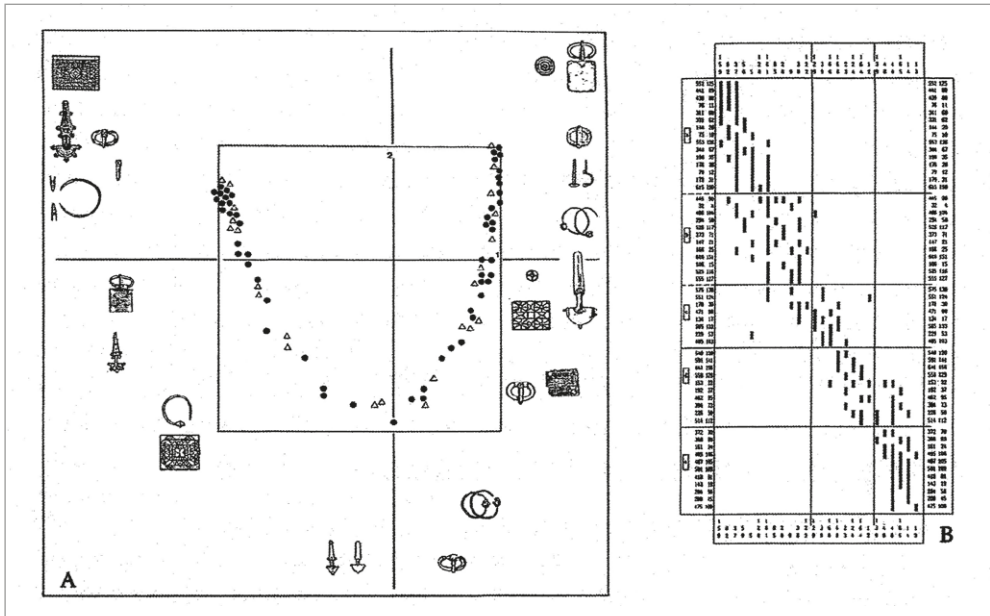


fig. 5 – Seriation: the Visigothic cemetery of Duraton, Spain (CIÉZAR 1990).

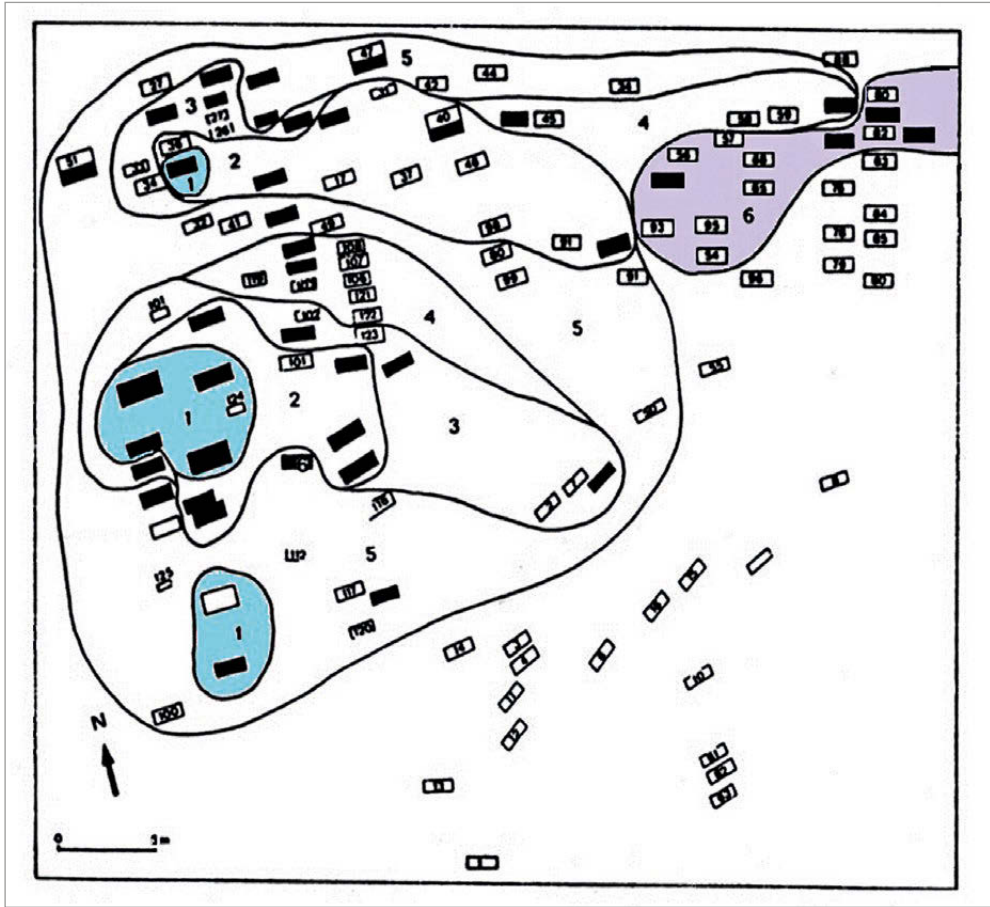


fig. 6 – Toposeriation: Merovingian cemetery of Mézières (DJINDJIAN 1985b).

Correspondence analysis has been introduced due to two of its main features (DJINDJIAN 1976):

- Evidence of a parabolic ‘Gutmann effect’, the equivalent of the horseshoe structure, for both observations (e.g., graves, layers) and variables (e.g., types, attributes), which enables seriation and the reorganization of the table based on the coordinates of the first axis.
- Demonstration of the identical results obtained using correspondence analysis and the reciprocal averaging method, as pointed out by Hill in 1974.

Moreover, correspondence analysis is used as an interactive serial analysis, for the elimination of observation outliers, aberration variables, and non-chronological factors (fig. 5), as well as for revealing complex seriation patterns (DJINDJIAN 1980b, 1985b).

A toposeriation method has also been developed (DJINDJIAN 1985b), simultaneously using correspondence analysis and a k-means-like cluster analysis with chronological

constraints on the spatial distribution of cemetery graves. The results reveal the topographical evolution of the cemetery over time around several poles of foundation graves (*fig. 6*).

Numerous studies of seriation using correspondence analysis have been published: LEREDDE and PERIN (1980), STEHLI and ZIMMERMANN (1980), COURBIN (1983), AUDOUZE (1984), SLACHMUYLDER (1985), GILIGNY (1990), CIÉZAR (1990), PION (1997), MARION (2002), etc.

The seriation method has been applied to the graves of Bronze Age and Iron Age necropolises in Europe, and to Visigothic and Merovingian cemeteries. It is the primary method for a precise periodization of the Bronze Age, the Iron Age, and the period of the Great Invasions. It has also been directly applied to artifacts to reveal the evolution of the style of decoration or manufacturing technology.

A special seriation package using correspondence analysis was developed by SCOLLAR, WELDNER, HERZOG (1985).

Physico-chemical characterization

Physico-chemical characterization involves both unprocessed raw materials (obsidian, flint, amber, hard stone for polished axes, grindstones, building materials, volcanic ash etc.), for which the analysis is carried out on the identified outcrops, and transformed

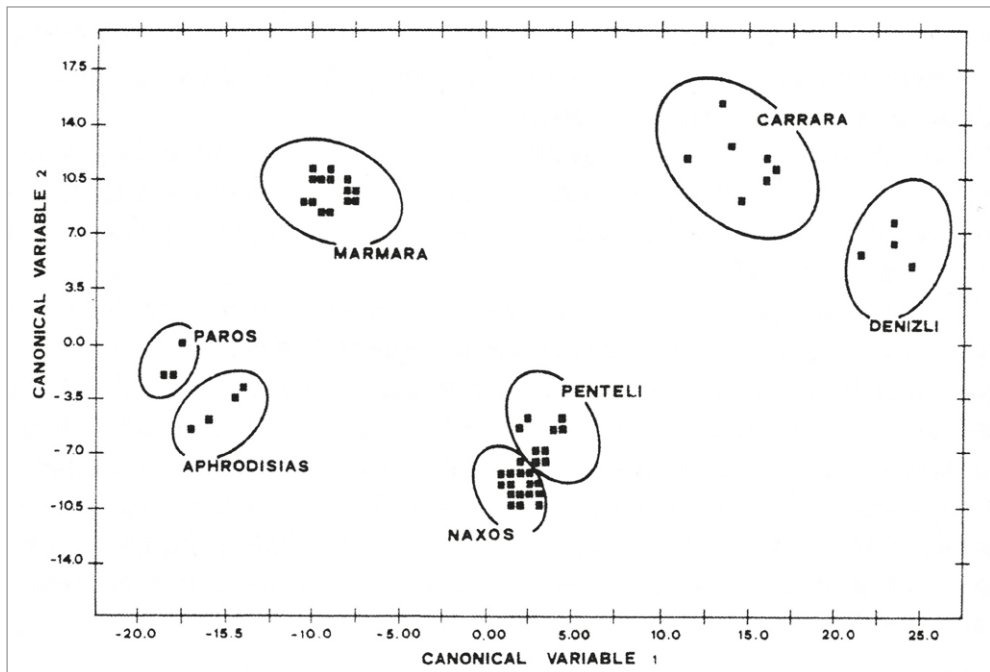


fig. 7 – Physico-chemical characterization of Mediterranean marbles (MELLO, MONNA, ODDONE 1988).

materials (ceramics, metal alloys, precious metals, glass, stained glass, stucco, mortars and plasters), for which the analysis must be able to distinguish the original raw material (clay, ore, silica) from its treatment (firing, reduction, mixtures) and its inputs (temper, flux, dye, glazes, etc.).

In addition to microscopic, petrographic, and metallographic studies, physico-chemical characterization methods use various spectroscopy techniques (X-ray fluorescence, mass spectrometry, atomic absorption, atomic emission, Pixe, etc.), neutron activation, and isotopic analyses. These analyses make it possible to characterize a sample by its atomic composition, made up of major elements (e.g., copper, tin, arsenic, lead for bronze), trace elements (present in used and reduced ores), and isotopes. It is this table of measures that is processed by data analysis (*fig. 7*). The objective is to identify by discrimination the sources of raw materials (flint, obsidian, amber and hard stone deposits, marble quarries, volcanic eruptions), the origin of copper, lead, tin, silver, and gold ingots (by isotopes), and glass frits, and, finally, the manufacturing centers (potters' workshops, forges).

To this end, data analysis has been widely used: DESBAT and PICON (1986) for potters' workshops, BERTHOULD (1979) for the metallurgy of ancient Iran, GALE (GALE, STOS-GALE 1982; GALE 1991) for the isotopic analysis of copper ingots in the Mediterranean, MELO, MONNA, ODDONE (1988) for Mediterranean marbles, WIDEMAN *et al.* (1979) for Gallic amphorae workshops, CAUVIN *et al.* (1998) for obsidian in the eastern Mediterranean, BAXTER (1989) for glass, etc.

Landscape analysis

Territory analysis (formerly called intersite spatial analysis) is not limited to the study of the transformation of a landscape by humans but more generally concerns the exploitation of a territory by humans in relation to animal and plant species. From the 1990s onwards, it found a remarkable tool, the Geographic Information System (GIS), to record and, above all, visualize a territory on successive maps according to different variables of physical geography, various resources, farming, settlement location, etc. However, this inherently multidimensional frame of reference has not sufficiently motivated archaeologists to supplement graphic visualizations with data analyses. One of the rare examples is the study on the location of the ancient Punic harbor of Carthage (VITALI *et al.* 1992), which reconstructs the lines of the lagoon shore over time using systematic coring.

Harris matrix processing

The Harris matrix method is unquestionably the reference method for reconstructing the stratigraphy of anthropogenic deposits, such as those studied in urban archaeology. However, the method produces such a large number of archaeological units and stratigraphic relationships that it is essential to have a tool capable of reducing them into a synthesis graph. It is a data analysis algorithm that has enabled this automation, based on a graph slightly modified to make it compatible with graph theory (DESACHY,

DJINDJIAN 1990), and which can be found implemented in the software *Le Stratifiant* (DESACHY 2008). A similar algorithm was designed for the correlation of incomplete stratigraphic sequences of geological origin (DJINDJIAN, VITALI 1992).

Archaeology of built heritage

The archaeology of built heritage adopts the previous stratigraphic concept and applies it to the various phases of construction, addition, and destruction observed on a building. In this context, it is also possible to implement an analytical treatment of building data, but to my knowledge, no study has been carried out except for a study on the stuccoes decorated with floral motifs on the intrados of an arch of the Bust mosque in Afghanistan (HARDY-GUILBERT, DJINDJIAN 1980). The data analysis revealed that the construction of this decoration was the work of two teams employing distinct strategies for the installation of the stuccoes in the diamond-shaped mesh of the intrados based on the models of their arrangement. The first team, working on the left arch, followed a progressive distribution from bottom to top, while the second team, responsible for the right arch, employed a uniform distribution.

Prehistoric art

Palaeolithic cave art has provided a large corpus of animal and human representations, as well as signs. Their first inventory was published in 1965 in *Préhistoire de l'art occidental* by A. Leroi-Gourhan, using punched cards. These inventory tables were processed by data analysis as early as 1976 (ROUX, ROBERT, JAMBU 1976) and then re-examined by Georges SAUVET and Suzanne SAUVET (1979). Statistical studies on the style of representations have been carried out in an attempt to trace a sort of history of Palaeolithic art – for example, in the analysis of female statuettes (BOURILLON, FRITZ, SAUVET 2012), cut-out contours with horse heads from the Middle Magdalenian (BUISSON *et al.* 1996), horses (APELLANIZ, GOMEZ 1999), fishes (CITERNE 2003), horses and bison (PETROGNANI 2010).

The analysis of bestiaries from cave art and mobile art within well-dated contexts (stratigraphy, painted caves dated by C14) has made it possible to identify different types of bestiaries relevant to the Aurignacian/Gravettian, the Solutrean and the Magdalenian periods, and to propose a new explanation of the function of cave art within the circulation spaces of hunter-gatherer groups (DJINDJIAN 2004, 2013a).

Paleoanthropology

Physical anthropology has a long tradition of morpho-cranial measurements, dating back to the end of the 19th century, in particular within the framework of the Société d'Anthropologie de Paris. These measurements have been carried out on the collections of skulls from all periods and all continents concerned in numerous laboratories worldwide (e.g., Musée de l'Homme, Harvard, etc.). The first attempt to apply data analysis to craniometry was undertaken by W. Howells (HOWELLS 1966, 1973), who

was trained in statistics by H. Hottelling. A conference proceedings volume presented the results of these research studies in the early 1980s: *Multivariate statistical methods in physical anthropology: a review of recent advances and current developments* (VAN VARK, HOWELLS 1984).

In France, the first correspondence analysis on primate skulls was conducted by N. Petit-Maire in her PhD dissertation (PETIT-MAIRE *et al.* 1972). Since this pioneering work, data analysis has been in permanent use, as evidenced by articles published from the 1980s to the present. However, physical anthropology has increasingly focused on Pleistocene and Miocene paleoanthropology, abandoning the study of Holocene human populations since the 1980s.

Although statistical methods have been facilitated by the availability of craniometry data – over whose definition and relevance there is general consensus – there are nevertheless difficulties that make the practice delicate, such as sexual dimorphism and allometry in a context of great rarity, and frequent fragmentation of available fossils. While principal component analysis remains the most widespread technique, it is important to note that when applied to raw measurements, its first axis will always be a size axis.

Over the past twenty years, the 3D revolution has significantly changed the practice of cranial morphology studies by improving the availability of fossil data (tomographic databases and 3D meshes), stereophotogrammetry software for 3D acquisition, and 3D image processing software (particularly on 3D meshes). Additionally, digital morphometry software has made it possible to place and manage anatomical landmarks. Procrustes analysis, which seeks optimal superimposition of 3D figures through rota-

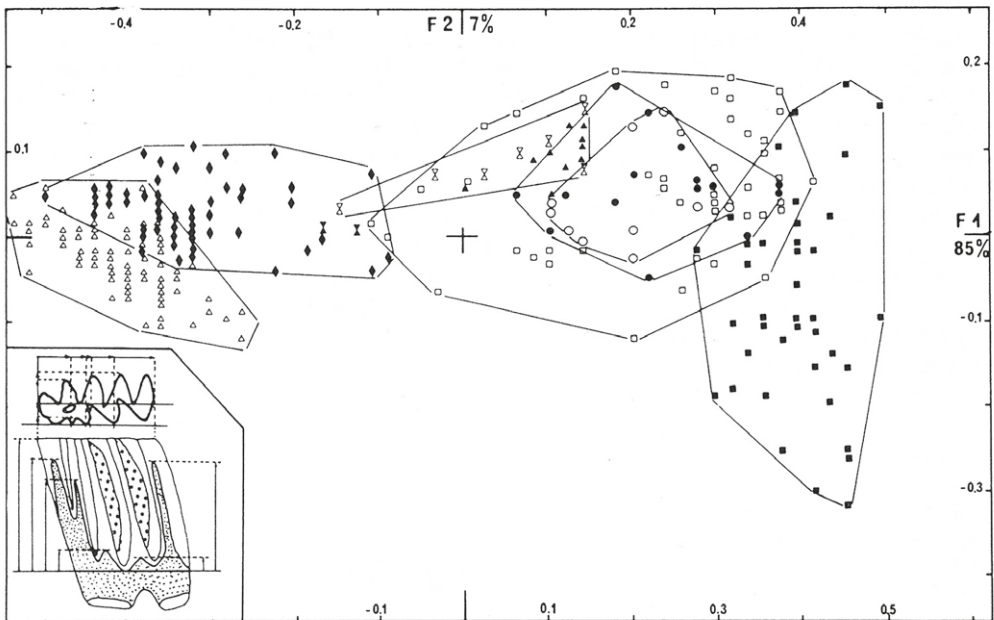


fig. 8 – Taxonomy of *Mimomys* micromammals (CHALINE, LAURIN 1984).

tion, translation and homothety, based on reference points, has become central to this work, alongside data analysis techniques such as principal component analysis and classification (BOOKSTEIN 1997; DUMONCEL 2017).

Paleontology, paleogenetics, paleoenvironmental studies, paleoclimatology

Paleontology presents the same need for the application of data analysis in addressing problems related to the systematics of species and their phylogenetic evolution (*fig. 8*). Numerous studies have been carried out, including on the genus *Equus* (EISENMANN, TURLOT 1978), rodents (CHALINE, LAURIN 1984; DENYS 1985), and insectivores with the genus *Crocidura* (ABI-BOUTROS, BELLIER 1977). Cladistic studies, which look for evolutionary traits in phylogenetic construction, are the formal equivalent of seriations in archaeology. Correspondence analysis is a useful tool for revealing an evolution through the parabolic Gutmann effect and optimizing it by selecting for these evolutionary traits. It is easy to understand that the partition/seriation duality is found in archaeology as well as in the systematics of species.



fig. 9 – Paleogenetic geography of Europe (CAVALLI-SFORZA, CAVALLI-SFORZA 1993): a) Factor 1; b) Factor 3.

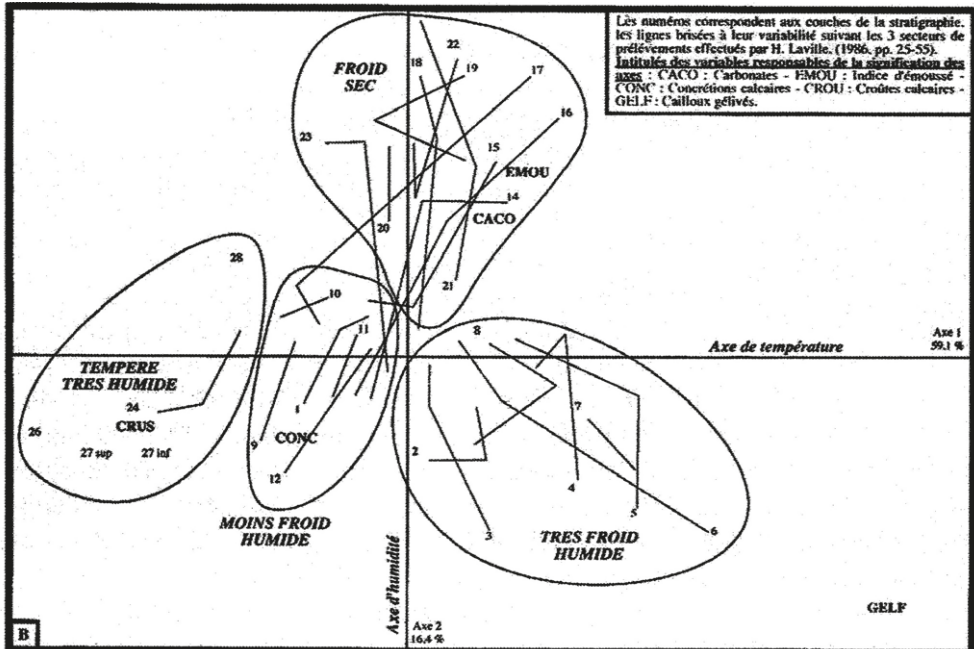


fig. 10 – Analysis of the stratigraphic sedimentation of the Riera Cave, Asturias (BOSSÉLIN, DJINDJIAN 1999).

Paleogenetics is a young discipline that has a significant need for genomic data processing techniques to identify haplogroups (sets of haplotypes that are series of alleles located at precise positions in a chromosome and characteristic of a mutation), and to study their distribution in time and space. Using a principal component analysis on 95 genetic markers, the work of L. Cavalli-Sforza in population genetics (AMMERMANN, CAVALLI-SFORZA 1984; CAVALLI-SFORZA 2000) has highlighted on the first factor the Neolithic colonization of Europe from the Near East, on the second factor the carriers of the Uralic languages of north-eastern Europe, on the third factor the arrival of Yamna nomads from the Don/Volga region, on the fourth factor Greek expansion in the Mediterranean, and on the fifth factor the Basque particularity (fig. 9). It is interesting to note that the paleogenetic approach is similar to the typological approach of material culture and the morpho-cranial measurements approach of physical anthropology: haplogroups, sub-haplogroups and markers are the equivalents of types and subtypes and measurements, whose monitoring in time and space seeks to identify population migrations. In all three cases, nothing is sought regarding the function and real meaning of these 'objects' in the functioning and evolution of human societies.

Paleoenvironmental studies are particularly concerned with the application of data analysis, whether from Quaternary sedimentation (in loess sequences or rock shelters), archaeozoology (composition of mammal, micromammal, or bird faunas), archaeobotany (charcoal, macroremains), palynology (especially pollen in lake sediments, maars

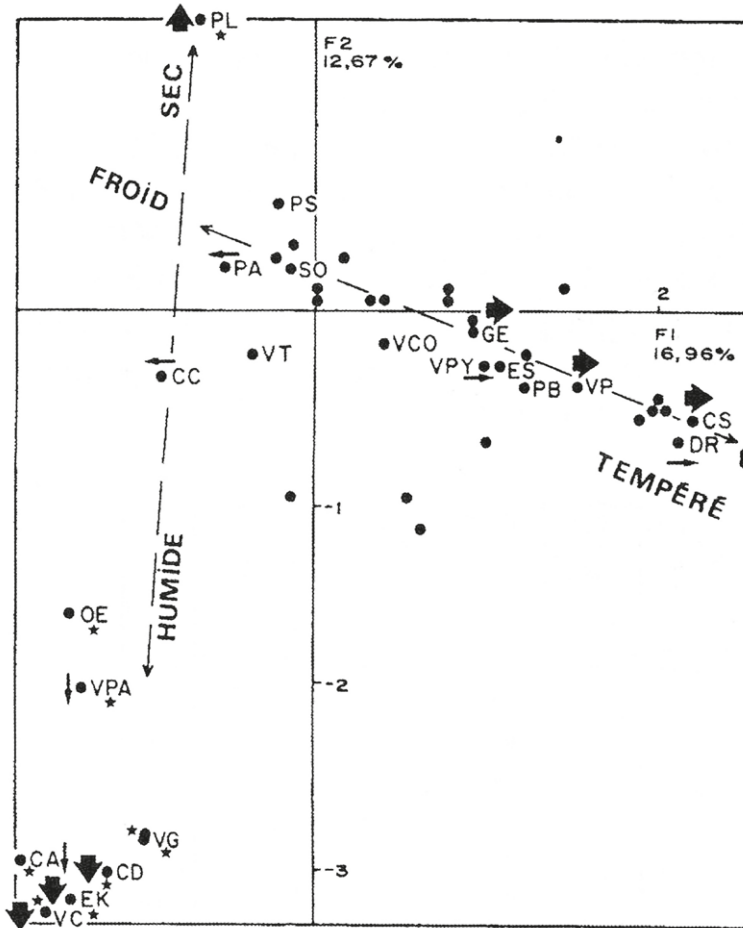


fig. 11 – Climatic and hygrometric gradients of the malacological assemblages of Achenheim in Alsace (LAURIN, ROUSSEAU 1985).

and peat bogs), and malacology (fluvial molluscs, shells of marine sediments, and in particular benthic foraminifera).

Samples are obtained by coring for natural sites and by sampling in stratigraphy or boreholes for archaeological sites (fig. 10). The data are the result of species determinations (at best) except for geoarchaeological studies (for which the standardized descriptors have been gradually refined: fragmentation and surface condition of rocks, grain size studies, geochemistry, etc.). Factors generally highlight a temperature axis and a humidity axis (fig. 11), making it possible to construct estimators that are essential for paleoclimatology studies (see below).

It is impossible to cite here all the references of these applications, which are very numerous: in palynology, the work of J. Birks (BIRKS, GORDON 1985); in limnology, those of F. Gasse on diatoms (GASSE, TEKAIA 1979; GASSE 1986); those of J. Chaline on rodents (CHALINE, LAURIN 1984); and those of J.-J. Puisségur on molluscs (LAURIN, ROUSSEAU 1985). The applications in geoarchaeology are less

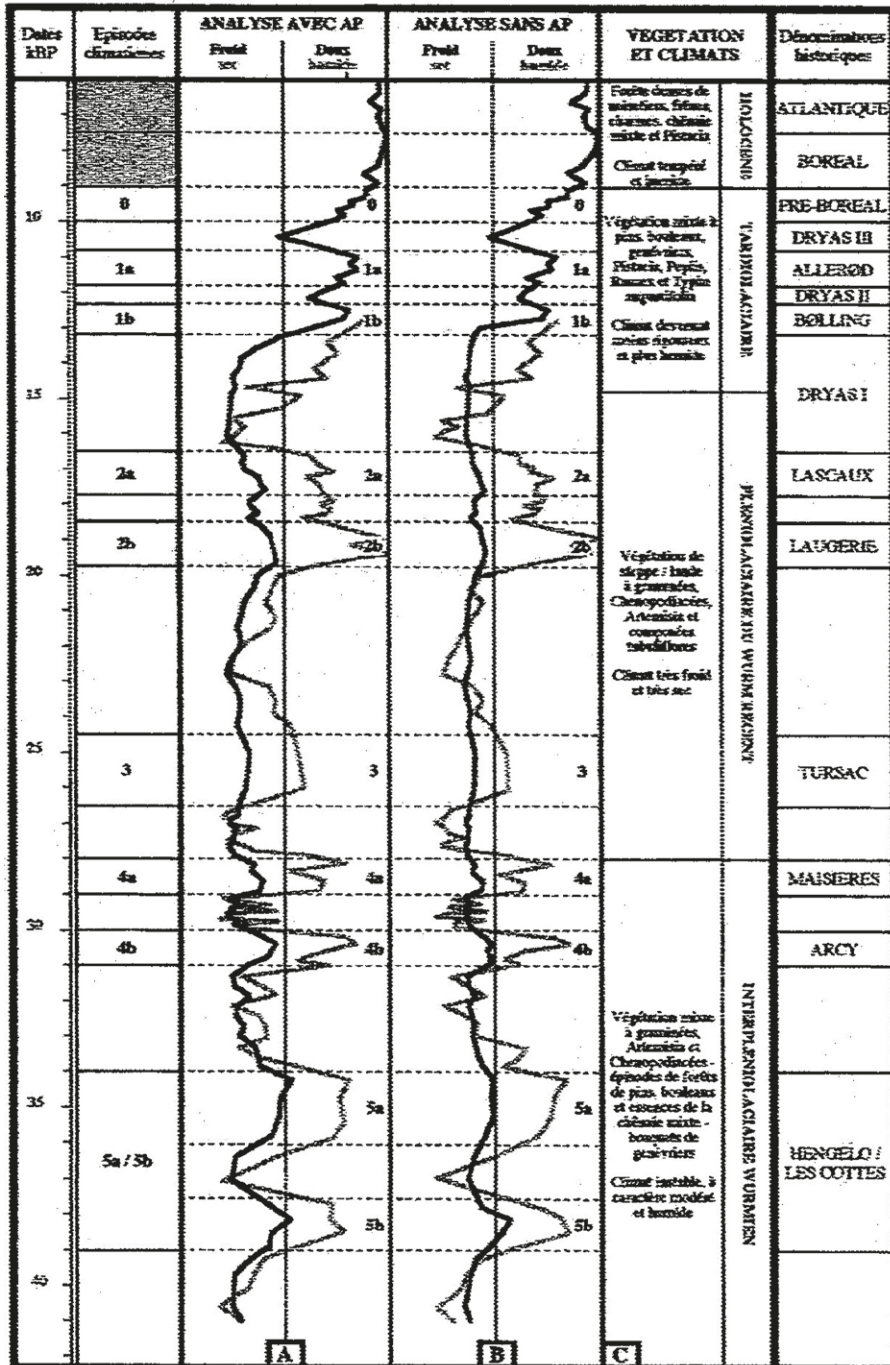


fig. 12 – Paleotemperature and paleohumidity curves for the period 50,000-10,000 BP, obtained via correspondence analysis on Tenaghi-Philippou pollen diagrams, Macedonia (BOSSELIN, DJINDJIAN 2002).

numerous. Stratigraphic sequences, with complex fills, can be critically analyzed by multiplying the samples to assess the variability of the lateral facies, as in the study of the sequence of the Riera cave in Asturias (BOSSELIN, DJINDJIAN 1999, on the data of a study by H. Laville).

The use of data analysis in paleoclimatology is linked to its ability to reveal a temperature axis that makes it possible to construct a paleo-temperature curve, as shown by an article by M. Roux, who compares the paleo-temperature curves constructed by IMBRIE and KIPP (1971) from a principal component analysis, and ROUX (1979) from a correspondence analysis on the foraminifera of an ocean core.

Many paleontological assemblages exhibit this characteristic of a high adaptation to climatic variations, such as benthic foraminifera, diatoms, molluscs, micromammals, and birds, as well as paleobotanical assemblages, such as pollens (KAY 1979) and dendrology (GUIOT 1984).

In order to deal with the correlations between the evolution of human societies and climate variations, archaeologists need paleo-temperature curves and paleo-precipitation curves, whose precision is of the same order of magnitude as the changes observed in material culture. As these curves are not always available in the work of paleoclimatologists (especially in ice cores), it is necessary to construct them from the available data, as in the case of the European Upper Palaeolithic with the pollens from the core samples of Lake Tenaghi-Philippou in Macedonia (BOSSELIN, DJINDJIAN 2002), where the analysis of the data made it possible to separate temperature and precipitation on the first two factors (*fig. 12*).

4.2 DATA ANALYSIS AS A COGNITIVE PROCESS

The application of data analysis techniques in a cognitive process in archaeology has been achieved by means of different and contradictory approaches:

- BINFORD, BINFORD (1966), in his famous case study on Mousterian assemblages, employed factor analysis, following the psychometricians, looking for the hidden factors: in this case, functional patterns (which unfortunately were not present in his typology).
- BENZÉCRI (1973) used correspondence analysis, in a totally inductive perspective, considering the algorithm able to find the pertinent structures in the data, from a supposedly exhaustive description.
- DORAN and HODSON (1975) used data analysis as a multivariate reduction method on observations only, without trying to formalize any rules on the description (variables).
- DJINDJIAN (1980a, 1985b) used data analysis as a tool in an iterative cognitive process between intrinsic data and extrinsic data, in order to find and optimize structures which can represent a system in a systemic approach (see below). In this approach, the typological method (which has always been the archaeological method of reference in prehistory since the 19th century) is only the first step in a cognitive process of archaeology that makes it possible to highlight structures (shapes, tooling, manufacturing

processes, dwelling structures, spatio-chronological ‘cultural’ facies, etc.), and to extract and enrich the relevance and semantics of the information of the initial typology.

The above considerations demonstrate the need to separate the three levels of the data analysis role in archaeological constructs:

- The technical level (capabilities and limits of data analysis techniques for each case study).
- The methodological level, which is the integration of data analysis techniques into an archaeological formalized method (seriation, culture facies study, typometry, etc.).
- The cognitive level, which is the integration of methods in an epistemological framework, producing knowledge, which can be validated or refuted.

An example of such an epistemological framework is the systemic triplet method (DJINDJIAN 1980, 2002) as a contribution to a general theory of knowledge in archaeology, schematized by the formula:

$$(I, E, S) \text{-----} (I+, E+, S+)$$

where *S* is a system of artifacts, *I* is an intrinsic description of the artifacts, and *E* is the information of the archaeological context of the artifacts. The motor of the cognitive process is data analysis on data matrices crossing *E* x *I* on artifacts of *S* (DJINDJIAN 2011, §5). The structures obtained by data analysis are optimized by an iterative process on *I* and on *E*, transforming iteratively the typological data *I* in an explicit system state variable *I+*, improving the level of pertinence of the extrinsic information *E+* and structuring the system as *S+*.

Correspondence analysis, as an *R+Q* method with distributional equivalence and the ability to process supplementary elements (individuals and variables), plays a role of systemic learning in such a cognitive framework.

4.3 TEACHING DATA ANALYSIS IN ARCHAEOLOGY

The revolution brought about by data analysis in the humanities and social sciences required specialized teaching on the practice of computer science in the 1970s, on the one hand on the computers at CNRS computing centers or universities, but also on the mastery of the use of statistical programs that only practical training could provide.

In the example of archaeology at the University of Paris 1, this training was provided as part of a DEA seminar with tutorials at the CIRCE computing center in Orsay, between 1980 and 1990. Each year, a volume of the projects produced by the students was published (e.g., DJINDJIAN 1985a).

The training of a new generation of European archaeologists was further advanced by an intensive European course in 1983 under the aegis of the Council of Europe. Thirty students, selected on the basis of their applications, were trained with lectures in the morning and practical work in the afternoon, by thirty European specialists

(DJINDJIAN, DUCASSE 1987). These students became professionals and, in turn, relayed the use of data analysis.

In the 1990s, a multi-year course, organized by the Chair of Archaeology Methods and Theory of the University of Paris 1 Panthéon Sorbonne, was established to set up a curriculum beginning in the first year of university and leading to a master's degree in specialized training in computational archaeology, including a module dedicated to data analysis. These courses, which have been highly popular among the students, have given them the essential tools for their doctoral theses.

4.4 THE POST-MODERN REACTION IN ARCHAEOLOGY

The data analysis revolution is a component of the formalist and quantitative movement in the humanities and social sciences that emerged after the Second World War, particularly in relation to structuralist and systemic approaches. In archaeology, this was associated with New Archaeology or processual archaeology (even though the latter favored the use of statistical tests, which will obviously be much criticized later, and rightly so, rather than for data analysis).

Faced with the excesses of a structuralism used in all sorts of ways, sometimes misguided, various movements came into being in reaction that can be grouped under the generic term of postmodernism. This trend is well-illustrated by the well-known term deconstruction, which came from the ideas of French Theory, which in fact brought together a diverse group of French philosophers (such as Foucault, Derrida, Baudrillard, Deleuze, etc.), and which were to find a multiplying echo from 1975 onwards in the Anglo-Saxon academic circles of the Humanities.

The movement appeared in archaeology in the 1980s under the name of post-processual archaeology or symbolist archaeology, with personalities such as I. Hodder and Ch. Tilley, who were influenced by postmodernism and Marxist anthropology, but they broke through in the 1990s and took root particularly in Anglo-Saxon countries by developing themes such as feminism, gender, indigenism, and post-colonialism, themes that can be found in today's 'wokeism'.

Post-modern archaeology, which is a hermeneutic of archaeology, abandons rigorous and refutable demonstration in favor of multiple interpretations, and consequently no longer quantifies the phenomena it studies. In this new world without tables of data to process, data analysis was also neglected, as shown by the studies of the *chaîne opératoire*, which, at the same time, sought to represent a technological alternative to morphotechnical and morphological typometry without any quantification (however possible, see above), or the studies of cultural facies after the deconstruction of typological lists.

Post-modern archaeology thus appears as a scientific regression in archaeology, the outcome of which must be assessed thirty years after its emergence. This regression could have been avoided if, instead of stumbling over the first difficulties in quantitative archaeology, archaeologists had tried to overcome them by continuing the effort on the

semiology of descriptions (ontologies), the representativeness and reliability of the data from new archaeological excavations, and the mastery of statistical processing, which their insufficient training in mathematics blocked, thus contributing to correcting the processual analysis of its numerous dross to propose a neoprocessual archaeology (DJINDJIAN 2010).

This is all the more regrettable since data analysis is still of unavoidable use in paleoanthropology as well as in paleontology, paleogenetics, paleoenvironment and paleoclimatology, disciplines that have moved away from the human sciences in France, while in Anglo-Saxon countries we speak of 'Humanities', which means that they have already dropped the term 'Sciences', which will not surprise the reader.

Paola Moscati

5. DATA ANALYSIS AND 'CLASSICAL' ARCHAEOLOGY

C'est l'ordinateur qui apparaît ici comme un outil privilégié, parce qu'il permet de traiter des quantités énormes d'information à une vitesse considérable et selon des opérations extrêmement compliquées, celles de la statistique multidimensionnelle, de l'"analyse des données" ... Or, si la théorie de la stratigraphie est enseignée aux archéologues classiques dans un grand nombre d'Universités, dans combien peuvent-ils trouver une initiation, même élémentaire, à la logique mathématique, à la statistique? (GINOUVÈS 1988, 120).

5.1 THE QUANTITATIVE APPROACH AND THE DATA ENCODING STRATEGY

Since the 1960s, the use of computer applications by the humanities community has been instrumental not only in advancing scholarship but also in reshaping theoretical and methodological approaches to data analysis. The integration of technical and scientific tools has undoubtedly triggered lively reflections concerning the appropriate terminology, the methods, and the techniques to be employed for knowledge acquisition, analysis, and representation.

In archaeology, the broad scope of laboratory and fieldwork research areas as well as methodological approaches has given rise to two distinct schools of thought advocating for the renewal of traditional investigation methods. These approaches differ in both theoretical orientation and practical application, covering periods from prehistory to modern times. One school focused on developing computer science-based technologies designed to replicate traditional inventory and cataloguing practices (CHENHALL 1971; BORILLO, GARDIN 1974). The other emphasized the application of mathematical and statistical techniques, enhancing the heuristic foundations of quantitative data analysis for typological artifact classification (SPAULDING 1953) and the study of spatial distribution patterns (HODDER, ORTON 1976).

This divarication has had significant theoretical and methodological implications, particularly in the processing of archaeological information. The establishment of information systems and database management have prompted a thorough examination of issues related to language standardization and the use of 'metalanguages', which consist of three main components: a lexicon for indexing purposes, a semantic organization of terms, and a grammar outlining their syntactic relations (GARDIN 1981, especially Chap. 3). Among the earliest and most tangible outcomes was the development of archaeological descriptive 'codes' (GARDIN 1958) and the creation of thesauri or controlled vocabularies

with structured classifications of terms (GINOUVÈS 1971; CHENHALL 1978; GINOUVÈS, GUIMIER-SORBETS 1978; for a synthesis, see FROMAGEOT-LANIEPCE 2019). Since the 1970s, archaeological databases have remained a central focus of computational archaeology, representing one of the most long-lasting areas of experimentation (CARVALE 2022). Databases were the keystone not only for incorporating computer science into the disciplines of classical archaeology and art history but also for the introduction of information systems in the inventory and management of cultural heritage.

According to scholarly tradition, mathematical and statistical methods have been more extensively applied in prehistoric archaeology to identify, to characterize, and to study the evolution of material cultures (HODSON 1971). In the statistical classification of archaeological artifacts, two core steps need to be taken to construct the analytical framework: data modeling and encoding procedures. The rigid constraints of numeric data analysis have posed significant challenges for archaeologists trained in historical, literary, and art-historical scholarship, who complement their methodological panoply with contributions from written and iconographic sources (MOSCATI 1995b).

The words of Robert WHALLON (1972) eloquently reflect the quantitative school of thought of the 1970s. In comparing the relationship between computers and archaeological thinking, he emphasized the primary role of prehistoric archaeology in advancing the use of computers:

«A sharp distinction must be made here between prehistoric archaeology and classical archaeology, for computer analyses are at present being used essentially only in prehistoric archaeological research, and where the term “archaeology” is used, it will be intended to refer only to that area of research. Classical archaeology is reluctant to adopt computers or computer analysis, preferring still to rely on intuition, “feel”, etc., in analysis (Burton, Bonin, Lourie, and Spiselman, 1970) [BURTON *et al.* 1970]. The computer finds its place in this field primarily as a large data storage, sorting, and retrieval device. The problems in computer application with which classical archaeologists are usually concerned are those of uniform coding and recording for this purpose» (WHALLON 1972, 30).

Apart from the objections raised by GARDIN (1971), this predominant perspective – although never explicitly articulated – remained largely unchanged for a long time. When tracing the history of the quantitative approach, the earliest applications typically pertain to the field of prehistoric studies. Djindjian himself delineates the main developments as follows:

«À partir de 1965, la classification automatique et l'analyse des données multidimensionnelles font leur apparition (Hodson, Sneath & Doran, 1966: une classification automatique des fibules de Münsingen; Doran & Hodson, 1966: une *multidimensional scaling* sur des assemblages du paléolithique supérieur; Binford & Binford, 1966: une *factor analysis* sur des assemblages moustériens). En 1970, le congrès de Mamaia en Roumanie (Hodson, Kendall, Tautu, 1971) qui réunit archéologues et mathématiciens du monde entier, illustre spectacu-

lairement la dynamique de cette révolution méthodologique (Doran & Hodson, 1975; Djindjian, 2009)» (DJINDJIAN 2016, 132).

Exceptions did exist within this panorama, although they were limited in number. One example from Italy is the 'Informatica e Archeologia Classica' conference convened in Lecce in 1986 by Francesco D'ANDRIA (1987), which was structured distinctively around this specific chronological framework, although only two of the papers presented employed a statistical approach (MOSCATI 1987b; MASIERO 1987). A similar absence can be noted in another conference, held in Ravello five years later on the initiative of John BOARDMAN and Donna KURTZ (1993), also devoted to classical archaeology.

Since the 1980s, the advent of dedicated software and technological advances in desktop computers have facilitated the application of statistical techniques to the social sciences and humanities. Consequently, theoretical and methodological discussions have increasingly centered on integrating various approaches – from descriptive statistics to multivariate analysis – within a more comprehensive research framework. In archaeology, for example, the concept of a 'statistical cycle' has emerged as the combination of several dynamic models, where each subsequent stage serves as an endpoint for validation processes culminating in cultural and historical interpretation (ORTON 1980).

The illustrative examples we will present here reflect this evolution as they examine not only the reception of data analysis techniques but also look at them within a wider analytical context that is both comprehensive and interdisciplinary in nature. Within this framework, the integration of scientific analyses, technological innovation, and humanistic understanding created a uniquely powerful combination of expertise for addressing complex problems. For instance, while the traditional classification of archaeological artifacts relies on comparative procedures, the quantitative approach employs statistical association and correlation mechanisms (MOSCATI 1997). This shift in theoretical perspective can influence all stages of investigation, reshaping how the modeling of archaeological phenomena is theorized and analyzed. An initial descriptive level of analysis, when combined with the application of multivariate statistical techniques, fosters an epistemological process designed to extract new information and to identify a logical structure that fulfills specific criteria, such as those necessary for the classification and seriation of archaeological artifacts.

5.2 QUANTITATIVE ARCHAEOLOGY, DATA ANALYSIS, AND ANCIENT POTTERY

To outline an itinerary with historical roots that extend into the present, it is essential to examine specific issues that have long been central to quantitative archaeology. Just as lithic industries have played a crucial role in the study of material cultures in prehistoric times, ceramics serve, more than any other category of evidence, as a key indicator for the application of statistical analyses in classical antiquity and as a bridge to current AI-driven data analysis in archaeology and the Data Science movement (see Chap. 7).

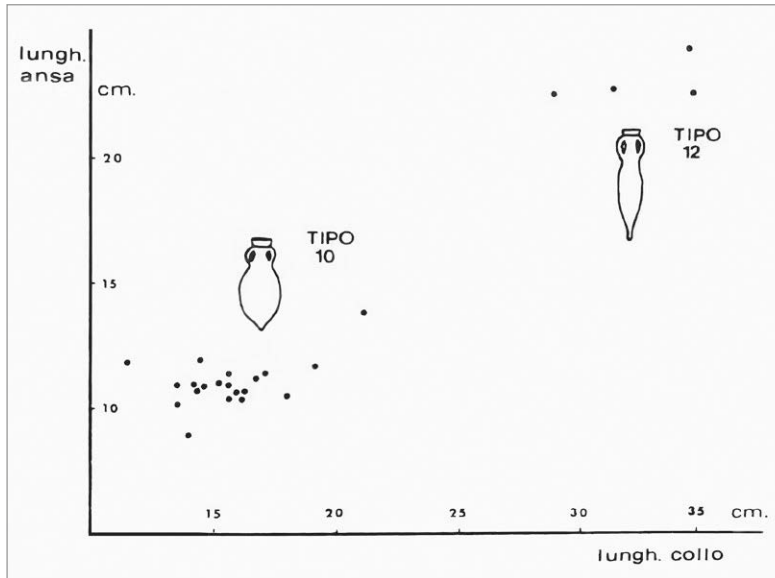


fig. 1 – Scatter diagram showing the distribution of Roman amphorae (types Dressel 10 and 12), based on metrical variables (IHM 1978, adapted in MOSCATI 1987a, *fig. 2*).

Due to their abundance and enlightening informational potential, ceramics have been studied from multiple perspectives (PEACOCK 1977; MILLETT 1979; ORTON 1982; ORTON, TYERS 1990): as ‘types’, focusing on form and fabric for classification and seriation purposes; as fragmented data, i.e., sherds that help quantify the amount of pottery in an assemblage; as an index of trade and connectivity, providing insights into distribution, exchange mechanisms, and trademarks; as evidence of consumption, deposition, and social or funerary practices in intra- and inter-site spatial analyses; and, finally, as a medium of cultural interaction that facilitated the dissemination of Greek myths and classical iconography across the Mediterranean and beyond. All these aims, along with the general rationale for practical approaches, can be summarized in Clive Orton’s ‘big three’: dating evidence, distributional evidence, and evidence for function and/or status (ORTON, TYERS, VINCE 1993, 23).

Research in this field is extensive, complex, and multifaceted. For this reason, I will focus on two ceramic classes that have attracted considerable scholarly interest: transport amphorae and Attic figured pottery. Both types of production are characterized by well-established taxonomies and long-standing research traditions, and their computer-based study provides an effective lens through which to trace major theoretical, methodological, and technical developments in automated classification procedures (SEMERARO 2016). From a data analysis perspective, amphorae research illustrates the evolution from the early application of descriptive statistical techniques to the adoption of advanced learning algorithms, which open up new possibilities for visualizing and processing large-scale datasets and integrate data from multiple sources (*fig. 1*). As for Attic figured pottery, we will explore a less well-known aspect, complex iconography studies, which will introduce the content of Section 3.

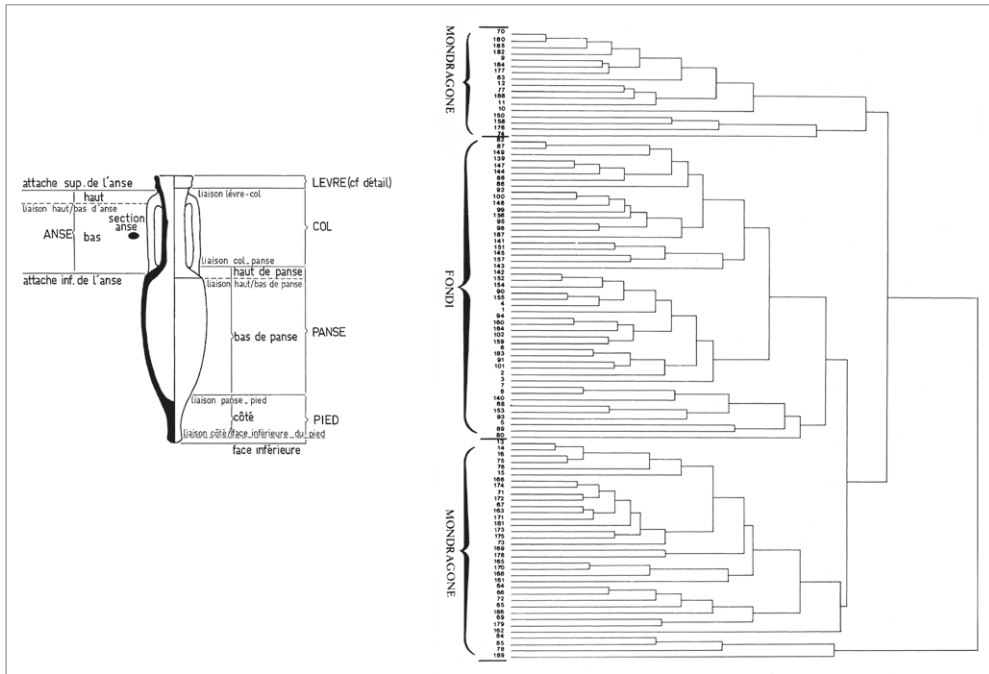


fig. 2 – Left: morphological description of an amphora through the segmentation of the curve of its profile (HAMOND, HESNARD 1977). Right: typological classification (hierarchical cluster analysis) based on compositional analysis of Roman amphorae from the Fondi and Mondragone areas (South Latium-North Campania) (HESNARD, LEMOINE 1981).

The application of computational methods for data recording began relatively early and reflected the need for a structured and formalized investigative approach to achieve greater analytical precision in quantification and classification processes. Pioneering contributions include the *Projet de Code pour l'analyse des formes de poteries* (CHRISTOPHE, GARDIN 1956, later republished by GARDIN 1976) and the studies collected in the proceedings of the conference 'Méthodes classiques et méthodes formelles dans l'étude des amphores' held at the École française de Rome in 1974 (MÉTHODES CLASSIQUES 1977; see in particular HAMON, HESNARD 1977) (fig. 2). J.-C. Gardin's conclusions to the conference state:

«M. Lamboglia redoutait que le dialogue entre archéologues et méthodologues ne restât un dialogue de sourds. Je voudrais essayer de dissiper ces craintes, en montrant que les deux parties peuvent au contraire «s'entendre», et que même elles se sont déjà entendues, si l'on en juge par certaines communications à cet égard tout-à-fait lumineuses. Un mot d'abord sur ces deux parties: archéologues et méthodologues, disais-je, ou bien plutôt, pour reprendre le titre du Colloque, adeptes des «méthodes classiques» dans un cas, et des «méthodes formelles» dans l'autre. Il serait faux de nier la réalité de cette opposition» (GARDIN 1977, 313).

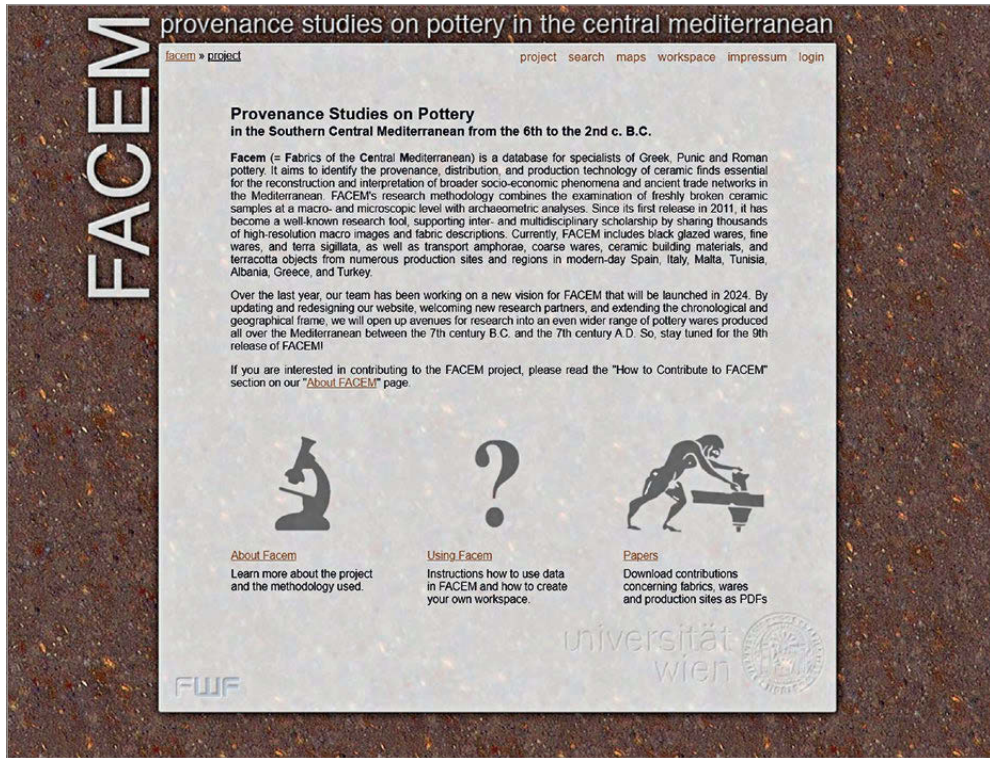


fig. 3 – The homepage of the FACEM database, started in 2011 and recently under renovation, an online collection of fabrics of Greek, Punic, and Roman pottery manufactured in the Mediterranean between the 6th and the 2nd centuries BCE.

While Gardin clearly fostered an interdisciplinary approach, he sharply criticized the statistical approach, describing it as the naive application of linguistic and mathematical operations to ‘solve’ the classification problems faced by amphorologists:

«J’ai déjà souligné l’inanité épistémologique de cette stratégie, où l’on attend du statisticien - car c’est de lui qu’il s’agit alors le plus souvent - qu’il fasse jaillir seul, du chaos de nos perceptions immédiates, l’ordre «réel» des choses, porteur de tous les messages historiques. Cette conception primaire de la collaboration entre archéologues et mathématiciens a encore ses adeptes, dans chacun des deux camps; je me félicite pour ma part qu’elle n’ait pas occupé la scène au cours de ce Colloque» (GARDIN 1977, 314).

Despite these criticisms, the quantitative approach to the study of amphorae has persisted and expanded, particularly in the analysis of the shapes and physical characteristics of the clay, to achieve a typological classification and provide information about production sites. The ‘Roman Amphorae: A Digital Resource website’ (UNIVERSITY OF SOUTHAMPTON 2014) illustrates the interesting history of the development of

quantitative studies, while the FACEM (Fabrics of the Central Mediterranean) database (*fig. 3*), designed in 2011 for interdisciplinary provenance studies, offers a web-based information system on ceramic fabric analysis (BORGERS *et al.* 2024; for other online digital initiatives on amphorae, see CARVALE 2022, 36-42).

The solid emphasis on this category in archaeological statistics stems from the extensive knowledge of well-documented productions, the abundance of ceramic finds from excavations and field surveys, and the tendency toward standardized production. The nature of the material has also facilitated the integration with archaeometric and diagnostic analyses, particularly those focusing on the physico-chemical and mineralogical properties of clay for provenance studies and manufacturing processes (PEACOCK 1971, 1977; AITCHISON 1986). In these studies, multivariate data analysis techniques, with a specific focus on principal component analysis and discriminant analysis, serve as the most appropriate tools for analyzing data and representing results (LEMOINE 1982; PICON 1984; BAXTER, HEYWORTH 1989).

In the 1980s, some experiments with innovative procedures introduced by expert systems and simulation techniques proved to be particularly forward-looking, especially in the case of Roman amphorae (GUENOCHÉ, HESNARD 1983; HESNARD 1987; see also VITALI, LAGRANGE 1988; VITALI 1989). These early approaches anticipated some solutions developed in more recent decades (see, e.g., BRUGHMANS, PECCI 2020; KOMAR *et al.* 2025), aimed at gaining insights into the exchange of goods at the local, regional, and inter-regional levels, as well as into the underlying factors behind economic efficiency or decline.

As for the history of statistical applications, the distinction between descriptive and exploratory multivariate methods has become increasingly challenging to define over time. Since the conference in Rome, descriptive techniques have been used recurrently for typometric and morphometric analyses, whereas exploratory methods, though less common, are generally employed as an extension of the results obtained from the former. From both methodological and technological perspectives, this field is well represented in a wide range of conferences related to the subject (for the 1990s, see, e.g., POLLARD 1992; OLCESE 1994; SANTORO BIANCHI, FABBRI 1997), as well as in numerous contributions from Italian scholarship, where the research tradition based on the exploration and visualization of high-dimensional data acquired with analytical techniques is long-established (see, e.g., CUOMO DI CAPRIO *et al.* 1976; CUOMO DI CAPRIO 1985).

Over the last decade, data analysis techniques applied to amphorae studies – as in other standardized ceramic classes – have undergone significant transformation, as exemplified in the recent EMAC 23 (European Meeting on Ancient Ceramics) conference, which opened with a special session on digital archaeology and pottery studies. This session was entirely dedicated to data visualization and portable analytical tools for archaeometry. Influenced by digital cultural heritage policies, the formalized description of artifacts has increasingly incorporated scientific visualization and 3D computer graphics, while the distribution phenomena have been analyzed through spatial analysis techniques in a GIS environment (*fig. 4*).

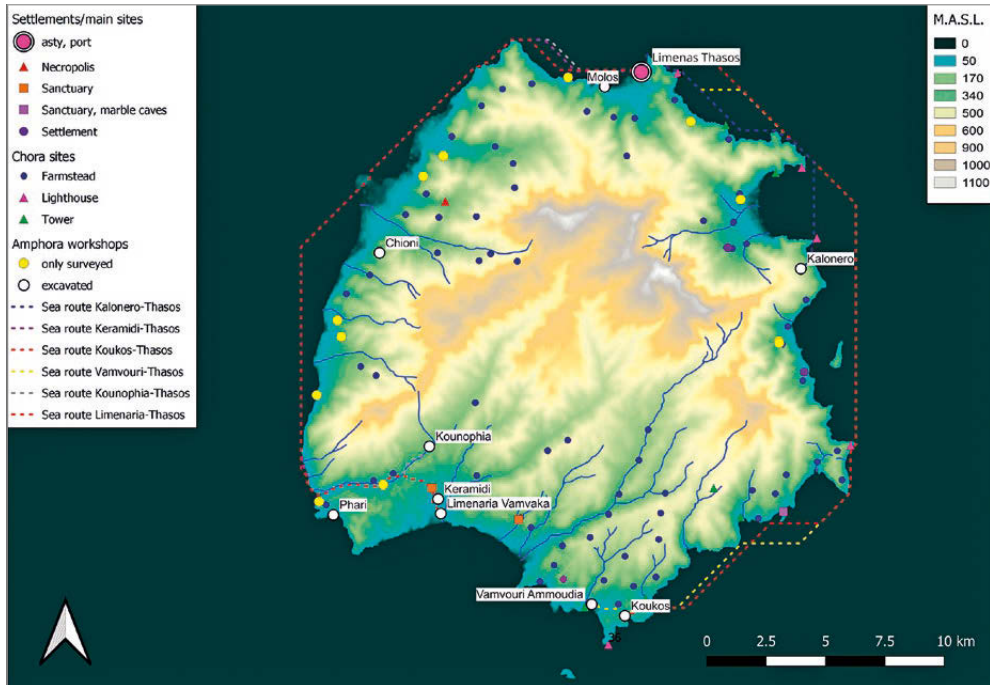


fig. 4 – A GIS application of Least-Cost Path Analysis to reconstruct potential transport routes connecting amphorae production sites on the island of Thasos, revealing strategic workshop placement and connectivity between countryside and port (TOMEI 2024).

Once again, the journal «Archeologia e Calcolatori» offers valuable insights into this transformation, which emphasizes the development of automated shape analysis techniques and the semantic enrichment of digital representations of heritage objects in an open science environment (fig. 5). Early studies introduced automated shape analysis for typological comparisons (MODRZEWSKA, TARONI, PIANETTI 2010, based on a method already described in MODRZEWSKA *et al.* 1993). This approach evolved into a method combining photogrammetry with ontological definitions of amphorae to generate accurate 2D and 3D reconstructions in GIS format (DRAP 2012). New computational approaches then enabled point cloud comparison for pixel-level 3D geometry characterization, leveraging Machine Learning techniques for efficiency and scalability (VO-PHAMHI, LEIDWANGER 2020). Optical scanning micro-profilometry has permitted in situ high-resolution surface analysis and 3D printing (MAZZOCATO *et al.* 2021). Legacy data have proven valuable for reconstructing a virtual 3D model of the *Grado I* Roman shipwreck, integrating excavation measurements, archival images, and digitized scale models (COSTA, BELTRAME 2021).

An intriguing precursor to more recent projects focusing on the automatic recognition of pottery fragments is the VASARI project (STANKE 1993), which was presented at the ‘Data and Image Processing in Classical Archaeology’ conference held in Ravello

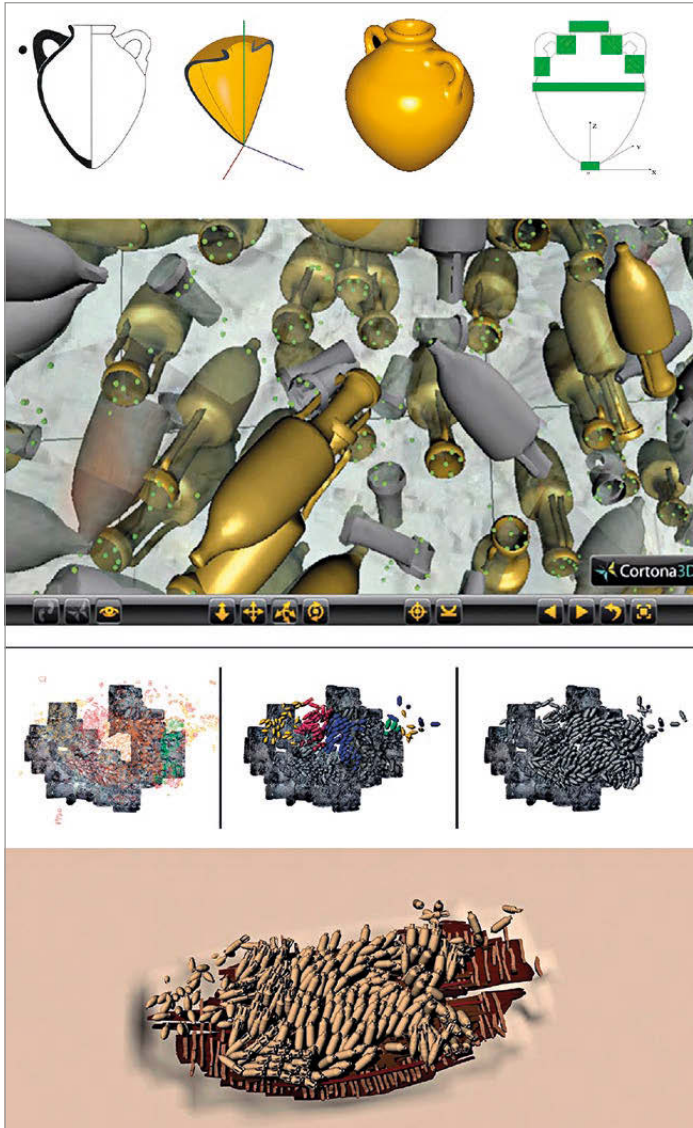


fig. 5 – Photogrammetric model optimization and 3D modeling to generate accurate representations of ancient amphorae and plausible reconstructions of their original positions in shipwrecks (top: DRAP 2012; bottom: COSTA, BELTRAME 2021).

in 1992, the proceedings of which were published in a special section of «Archeologia e Calcolatori». Although limited by the technological capabilities of its time, the project aimed to address a significant challenge – particularly relevant in museum storehouses – namely, the identification of archaeological ceramic fragments and potsherds through image processing.

Today, pioneering projects such as Arch-I-Scan (VAN HELDEN *et al.* 2022) or ArchAIDE (Archaeological Automatic Interpretation and Documentation of Ceramics) exemplify how the development of AI applications using Deep Learning algorithms

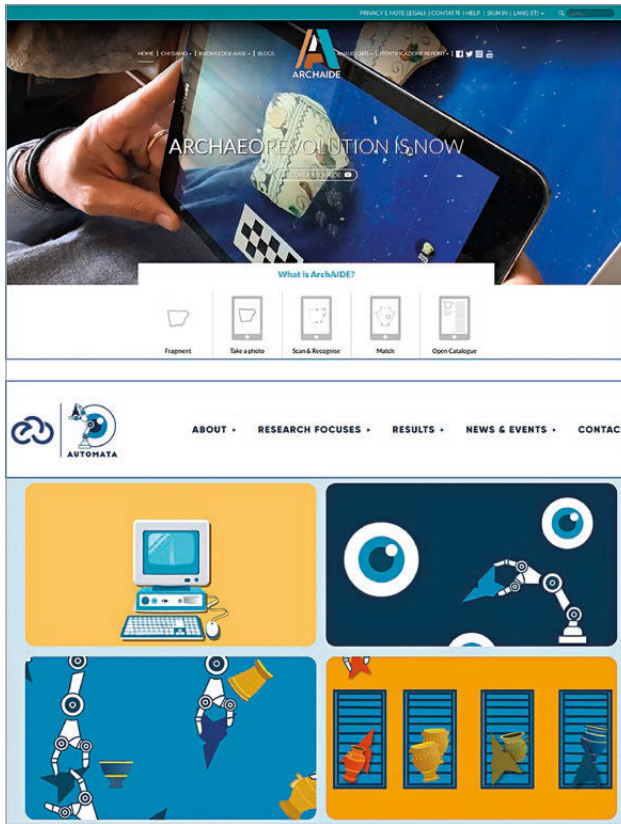


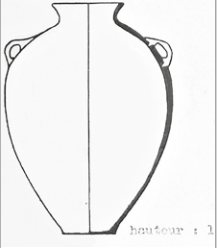
fig. 6 – Two European projects, ArchAIIDE and AUTOMATA, where Artificial Intelligence and robotics play a crucial role in the automated acquisition and typological recognition of standardized types of ancient ceramics.

can effectively tackle the same issue, namely, the automatic recognition of pottery fragments, including various types of amphorae (GATTIGLIA 2022; see also Chap. 7). But this is not the final frontier in the study of amphorae and ceramics in general, at least according to the goals of the promoters of the AUTOMATA (AUTOMated enriched digitisation of Archaeological liThics and cerAmics) project, recently funded by the European Union (*fig. 6*). AUTOMATA, in fact, aims to facilitate archaeological documentation through the development of an advanced digitization system that integrates archaeometric sensing, robotic automation, and Artificial Intelligence. In the operational phase, the system is expected to use a robotic arm to pick up ceramic or stone fragments, pass them to sensors for chemical and physical analysis, and then create a 3D model complete with associated data and information. And this is nothing compared to the METATOOL project, also funded by the European Innovation Council, which studies how robots can develop cognitive capabilities to invent tools, as ancient humans did some three million years ago.


Current approaches are certainly far removed from the theoretical basis of cognitive archaeology, so clearly illustrated by GARDIN (1996) and based on the contribution of formal methods (logicism included) to the process of acquiring archaeological knowledge.

Références :
Mexsin 207, fig. 127, n° 12

TYPES	A	⊗	p q r	r1 r2		
PANSE	B	haut:	x	o	liaison:	n/i
		bas :	x	l		r
BASE	C	⊙	profil	inclin.	extrémité	dimens.:
		e		i u o	'	'
	D	y	face ext.:	d	face int.:	liaison:
COL	E	⊖	profil:	liaison:	dimens.:	⊖ => ⊕ +
			d, v i	C		détails:
LEVRE	F	inflexion:	z	faces:	i ⊙ o	extrém.:
		inclinaison:	<			"
A types	G	⊗	at a' ef et e'	nombre:	z	< = >
N place	H		point inf.:	4 / 4	point sup.	w
S forme	I		profil:			
E section	J	p r	...			
		détails	K	z		
BEC types	L	⊗	...			
ou						
VER- forme	M		côté sup.			
SOIR sect.	N		p q r			
DETAILS	P		B C E			
TECHNO-LOGIE	Q	dimens.:				
		R	technique			
DECOR	S	décor:	⊖			
TYPES PARTIC.	Z	⊖	1 2			



hauteur : 1



53

Dans les scènes que nous avons analysées, Hélène est conduite par un hoplite qui la tient souvent par le poignet ou essaie de la tirer par l'himation (parfois ce personnage la poursuit en la menaçant d'une épée). Hélène porte un peplos et a la tête recouverte par son manteau.

THEMES - E I enlèvement, m/g d'une femme par un homme
PERSONNAGES TYPES PURS - 9 l'hoplite ou - 9' (G) selon l'existence des éléments obligatoires
PERSONNAGES TYPES PURS - 6 femme en peplos (G)
VETEMENTS 1. Corps - c sur la tête (1/G)

La spécification du vêtement d'Hélène est utile pour séparer ce thème d'autres scènes d'enlèvement d'une femme par un hoplite, comme par exemple Antiope (en costume phrygien) enlevée sur un char par Thésée.

fig. 7 – An amphora record in the *Projet de Code pour l'analyse des formes de poteries and the analytical description of the scene of Helen's abduction in the Code pour l'analyse des représentations figurées sur les vases grecs*, both kept in the Gardin Fonds of the Archives of the Maison des Sciences de l'Homme Mondes in Nanterre.

In discussing the computerized analysis of Attic figured pottery, I would like to focus on a less fully explored aspect. As we have already highlighted, Attic pottery has long been a privileged field of research in the history of archaeological computing. This has led to the development of dedicated electronic archives, some of which, such as the Beazley Archive Computer Project, were established as early as the 1970s (KURTZ 2009) and to the quantitative analysis of the distribution of Attic pottery in the Mediterranean (in Italy, see e.g., ARIAS 1978; ARIAS, DI BARI, ORSOLINI RONZITTI 1985; SEMERARO 1990; GIUDICE 1993; GIUDICE *et al.* 1999).

In addition, this class of pottery has generated particular interest in both iconography and iconology. The origins of this interest can be traced back in time, with Gardin and his team again playing a crucial role. Building on the *Code pour l'analyse des représentations figurées sur les vases grecs* (SALOMÉ 1965, later republished by SALOMÉ 1980) (fig. 7), this line of inquiry was further advanced at LISH (Laboratoire d'Informatique pour les Sciences Humaines) through experiments involving the SNARK expert system

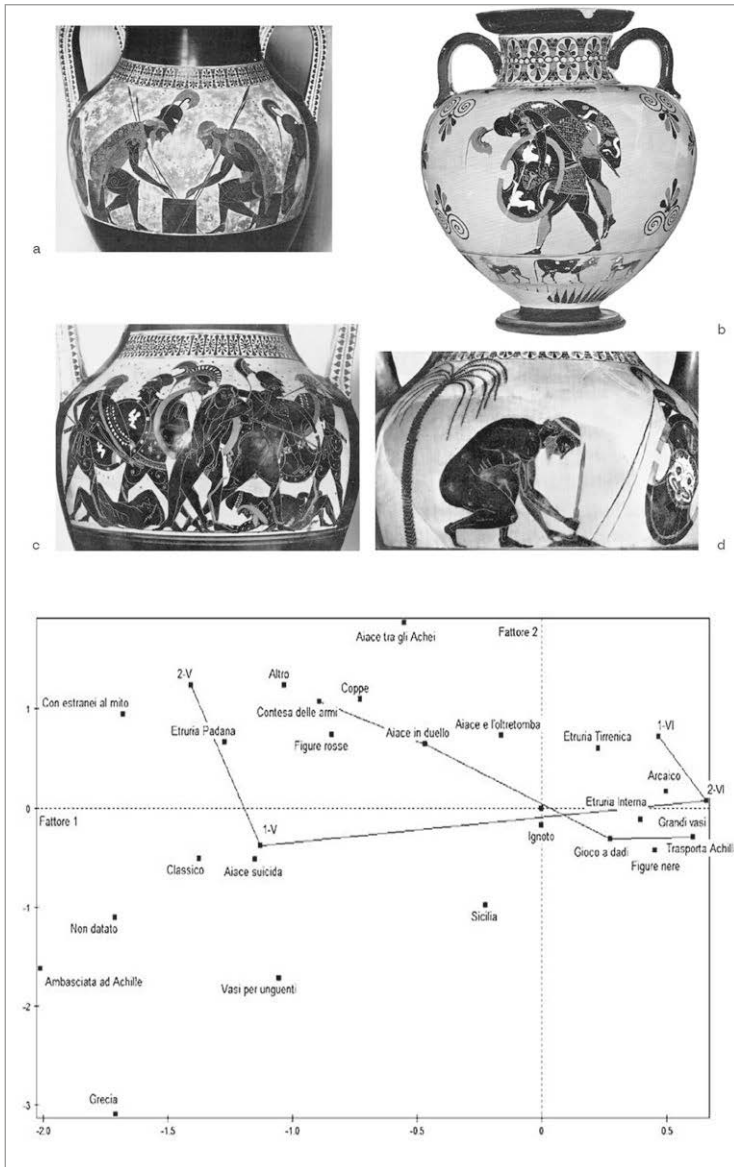


fig. 8 – Examples of Greek vases extracted from the dataset, with images of the Ajax myth (top). Multiple Correspondence Analysis: distribution on the factorial map of the ‘date’ and ‘iconography’ variables relevant to Attic figured pottery (bottom).

(LAGRANGE, RENAUD 1983a, 1983b, 1985; LAGRANGE, MONTEIRO RODRIGUES 1994). Less well known are some studies conducted since the late 1980s that employed exploratory data analysis techniques, such as correspondence analysis and multidimensional scaling, to examine the distribution of Attic pottery across the Mediterranean, particularly in relation to the spread of cults and rites from classical antiquity. For example, in the database of Attic pottery from Spina, statistical analyses enabled archaeologists to identify significant relationships between pottery shapes and

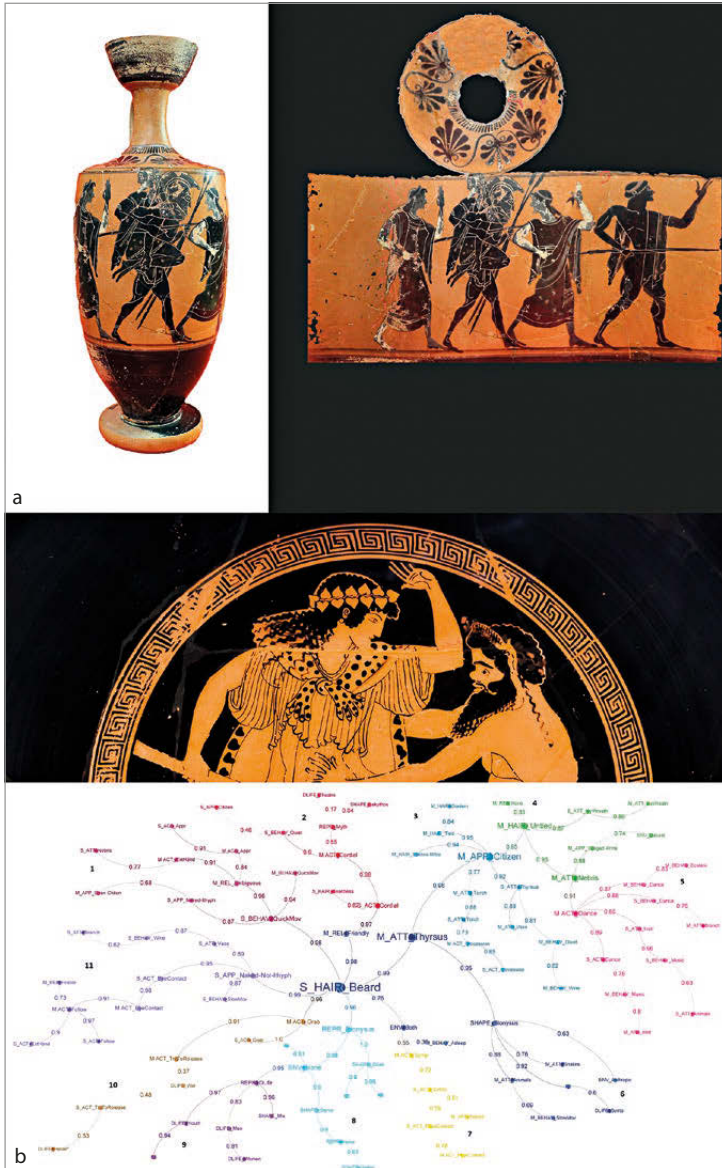


fig. 9 – Two innovative methods for the digital documentation and analysis of the iconography of Attic figured pottery: image acquisition by ‘unwrapping’ the decoration onto a flat surface, and the application of Artificial Neural Networks (ANN) for the analysis of satyr and maenad scenes (a: PACE, BURSICH 2021; b: WAYENBERG, CAPRIOTTI 2014).

subjects over time (ROSATI, QUARTILI, GUERMANDI 1989; GUERMANDI, MIGNANI, MONTANARI 1996).

To the best of my knowledge, the only research project that has applied data analysis techniques – specifically multiple correspondence analysis and discriminant analysis – to investigate iconographic aspects of the classical world is the work by Sergio CAMIZ and Eleonora FERRAZZA (2006), focused on the myth of Ajax, son of Telamon, within a collection of artifacts, primarily pottery, from Greece and pre-Roman Italy (fig. 8). This

study examines the iconography of the classical myth alongside various other variables, such as object type, production methods, painting techniques, place of discovery, and chronological attribution. Multiple correspondence analysis revealed significant relationships between the artifacts and these variables, and the interpretation of the resulting factorial maps helped illustrate how different meanings and functions of this Greek myth were represented throughout classical antiquity.

In line with recent scholarly developments, even in the field of iconographic studies of Attic pottery there has been a notable evolution towards incorporating methods employed in computer graphics. This advancement improves both the acquisition of images that ‘unwrap’ convex surfaces, facilitating the reading of their two-dimensional representations, and the restitution of 3D models that enhance the handling of these artifacts (BURSICH, PACE 2017; PACE, BURSICH 2021; and, more recently, TRINKL *et al.* 2024) (*fig.* 9a). Furthermore, a separate study has explored the use of Artificial Neural Networks (ANN) in investigating Greek painted iconology (WAYENBERG, CAPRIOTTI 2014) (*fig.* 9b). This research employed descriptive statistics and chi-square tests for the iconographic analysis of a selected collection of artifacts to redefine the identity of maenads and satyrs based on their appearance and the nature of their multiple interactions. An ANN-based analysis was then conducted, which confirmed and expanded the correlation between the satyrs’ appearance and the maenads’ reactions, revealing the presence of two distinct iconographies of maenads and highlighting divergent representations of feminine identity in Athenian red-figure pottery.

Finally, building on recent advances in Open Science and interoperability research, projects such as Kerameikos.org (<https://kerameikos.iath.virginia.edu/>) aim to define the intellectual concepts underpinning pottery studies through the principles of Linked Open Data and the development of ontologies for representing and sharing ceramic data across heterogeneous data systems.

5.3 ‘AUTOMATIZATION OF ETRUSCAN CORPORA’: A PIONEERING APPLICATION OF DATA ANALYSIS IN THE ARCHAEOLOGY OF PRE-ROMAN ITALY

In the early 1980s, a research project on the computerized analysis of Etruscan engraved bronze mirrors was initiated at the Accademia Nazionale dei Lincei, in the Centro Linceo Interdisciplinare di Scienze Matematiche e loro Applicazioni. This initiative was part of an interdisciplinary program exploring the application of computational methods in scientific research. The project gained early recognition as a distinct line of research under the name ‘Automatization of Etruscan Corpora’ at the Centro di Studi per l’Archeologia Etrusco-Italica of the Italian CNR, directed by Mauro Cristofani. Its primary objective was to test a computer-based methodology for analyzing and classifying homogeneous groups of artifacts from the Etruscan civilization (MOSCATI 1984, 1986).

The project aimed to contribute to the analysis of large archaeological corpora, particularly those compiled since the 19th century, whose systematic organization had

The figure displays two screenshots of the Heidelberg historic literature - digitized project website. The top screenshot shows the page for 'Etruskische Spiegel' by Gerhard, Eduard (Editor), Berlin, 1843-1897. The page includes a title, author information, keywords, bibliographic information, and citation links. A circular illustration of an Etruscan mirror is shown on the right. The bottom screenshot shows the page for 'I rilievi delle urne etrusche' by Brunn, Heinrich von (Editor); Körte, Gustav (Editor), Rom, 1870. This page also includes title, author information, keywords, bibliographic information, and citation links. A rectangular illustration of an Etruscan urn relief is shown on the right. Both pages feature a navigation bar at the top with the project name and university logo, and a footer with 'You might also be interested in...'.

fig. 10 – Homepage of the two websites of the ‘Heidelberg historic literature - digitized project’ promoted by the Digital Library of Heidelberg University, which reproduces the 19th-century corpora of Etruscan mirrors (GERHARD 1840-1867) and urns (BRUNN, KÖRTE 1870-1916).

proven suitable for integration into a computerized recording system (fig. 10). However, our objective was neither to create a database of Gerhard’s 19th-century monumental corpus of Etruscan mirrors (GERHARD 1840-1867), nor to examine aspects such as state of preservation, morphometry, or bronze composition, which were systematically incorporated only later in the *Corpus Speculorum Etruscorum* initiative, launched in the 1970s and published from the early 1980s onward. Instead, we explored the feasibility of integrating art-historical and statistical approaches with a view to analyzing specific qualitative attributes that shed light on artistic, social, and religious patterns.



fig. 11 – Some Etruscan mirrors depicting scenes from Greek mythology with the participation of Etruscan characters: a) the birth of Athena (*Menrva*) from the head of Zeus (*Tinia*); b) scene with Aphrodite (*Turan*), Helen (*Elina*), and Menelaus (*Menle*); c) the Judgment of Paris.

Using remote terminals connected to a Sperry UNIVAC 1100 mainframe computer at the Sapienza University of Rome, the research employed SPSS (Statistical Package for the Social Sciences) and SPAD (Système Portable pour l'Analyse des Données) software packages. In the first phase, the analysis began with the recording of approximately 1,000 Etruscan mirrors. All the selected variables were qualitative in nature, including the mirrors' provenance; the type and motifs of border decoration; the number, arrangement, and posture of the figures engraved on the reverse, or non-reflecting side of the mirror; specific characters and scenes represented within the border; the presence of inscriptions.

The dataset was structured as a data matrix, where rows corresponded to an individual mirror and columns represented a selected variable, resulting in a total of 898 records. Preliminary elementary statistical analyses provided an initial quantitative assessment of the sample. These analyses included frequency distributions and contingency tables, used in conjunction with the chi-square test to evaluate statistical significance. This approach established an initial interpretative framework, as a basis for more detailed analysis.

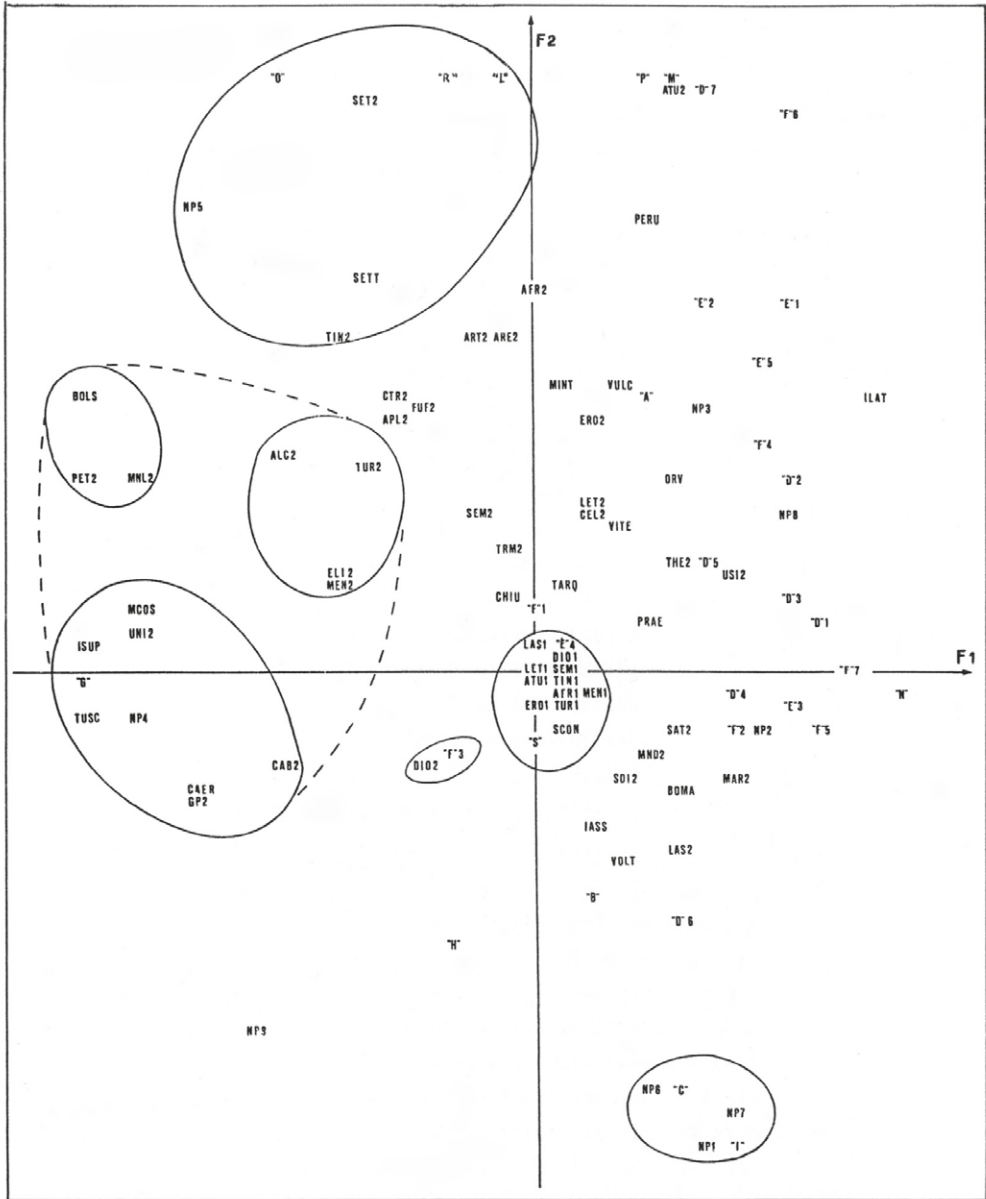


fig. 12 – Multiple Correspondence Analysis: projection on the factorial map of the qualitative attributes representing the mirrors’ provenance, decoration, iconography, and style.

Significant findings pertained to the frequency and distribution of mythological figures, such as the triad of deities *Tinia* (Zeus), *Uni* (Hera), and, most prominently, *Menrva* (Athena), as well as iconographic themes like the Judgment of Paris, the story of Helen of Troy, and abduction scenes, together with depictions of scenes from

```

<URNA>
<INFO>
<NUMERO_Catalogo>26</NUMERO_Catalogo>.
<OGGETTO>Uma</OGGETTO> in <MATERIALE>tufo</MATERIALE>;
<LUOGO_CONSERVAZIONE>Volterra, Museo Guarnacci</LUOGO_CONSERVAZIONE>;
<NUMERO_Inventario> inv. 538</NUMERO_Inventario>.
<DIMENSIONI>
<COPERCHIO>Cop.:
<ALT>0,36</ALT>;<LUNG>0,57</LUNG>x<LARG>0,23</LARG>.
</COPERCHIO>
<CASSA>Cassa:
<ALT>0,45</ALT>;<LUNG>0,57</LUNG>x<LARG>0,23</LARG>.
</CASSA>
</DIMENSIONI>
</INFO>
<DESCRIZIONE>
<COPERCHIO classe="antropomorfo" tipo="femminile">Sul coperchio recumbente femminile vestita di <abito>chitone
manicato e mantello che copre le gambe, il dorso, la nuca, il braccio e la spalla s., formando una grossa piega sotto l'addome
</abito>. <capo>Sulla testa <ornamenti>un diadema decorato nei bordi da una doppia fila di perline</ornamenti></capo>; <collo>al
collo <ornamenti>un torques</ornamenti></collo>. <abito>Il chitone presenta sulla spalla s. <ornamenti>un fermaglio a tre dischi
</ornamenti> ed è cinto in vita da <ornamenti>una cintura decorata da placche rettangolari</ornamenti></abito>. <capo>
<acconciatura>I capelli, con scriminatura centrale, si dispongono attorno al viso in ciocche ondulate</acconciatura></capo>.
<bracciodx>Il braccio d. è disteso lungo il corpo e <manodx>la mano, che poggia sul ginocchio d. piegato, regge <oggetto_mano>un
flabello con la punta rivolta verso l'alto e lungo manico. Il flabello è decorato internamente da un motivo perlinato a forma di foglia e da
piume incise</oggetto_mano></manodx></bracciodx>. <bracciosx>Il braccio s. poggia su </bracciosx><cuscino>due cuscini col
bordo ricamato da linee ondulate incise e nappine agli angoli</cuscino>, <bracciosx><manosx>la mano con le dita disposte nel
<gesto>gesto apotropaico</gesto>, tiene <oggetto_mano>un oggetto non identificabile</oggetto_mano></manosx></bracciosx>.
</COPERCHIO>
.....
.....
</DESCRIZIONE>
<CONSERVAZIONE>Figura sul cop. mutila della prima falange dell'indice della mano s.; manca l'attributo nella mano s. Il manico
del flabello ha una lacuna nella parte mediana</CONSERVAZIONE>.
<BIBLIOGRAFIA>Bibl.: B.-K. III, p. 225, tav. CLII, 1; Ducati, tav. 275, fig. 666; Giglioli, tav. CCCCVIII, 2; Gambetti, p. 21, n. 1, p.
71; Nielsen, pp. 305-7, 309, 314; Fiumi 1976, fig. 38</BIBLIOGRAFIA>.
<RINVENIMENTO>Acquisita dal Museo tra il 1875 e il 1877</RINVENIMENTO>.
</URNA>
    
```


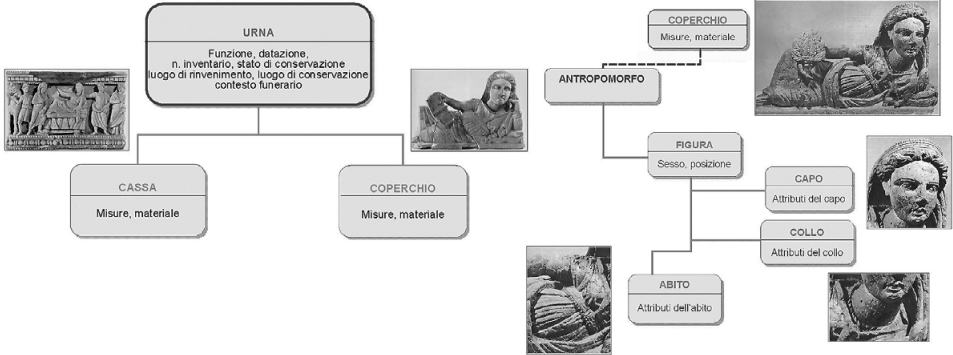



fig. 13 – Formalized classification of Volaterran urn lids as part of the project ‘Semantic and ontological research applied to cultural heritage: the classification of archaeological materials’ (SIGNORE, MISSIKOFF, MOSCATI 2005).

daily life associated with the *mundus muliebris*. The findings highlighted a significant relationship between figures from both Greek and Etruscan traditions, showing the integration of distinctly Etruscan mythological characters in scenes of Greek origin and the adaptation of Greek deities into local Etruscan narratives (fig. 11). Additionally, Discriminant Analysis was applied to a subset of mirrors of known provenance. This

method identified key iconographic and stylistic elements associated with different geographical production and consumption areas, providing insights into regional artistic preferences.

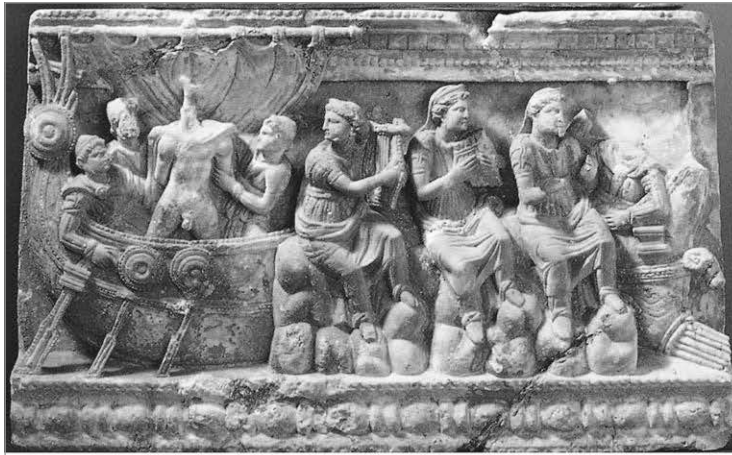
Following this initial phase, further multivariate analytical techniques were applied using SPAD, a statistical package distributed by CISIA (Centre International de Statistique et d'Informatique Appliquées, Paris). A dimensionality reduction method was integrated with clustering algorithms, specifically Multiple Correspondence Analysis, followed by Hierarchical Ascendant Classification. In Italy, this approach, still relatively uncommon in archaeology, had been primarily employed in social sciences, particularly in psychology (MUSSINO 1984), due to its effectiveness in analyzing large datasets characterized by categorical variables and in identifying the principal associative components of the phenomenon under investigation.

This approach facilitated the synthesis of information while preserving its intrinsic complexity, and allowed for the identification of homogeneous groups of mirrors without imposing a priori classifications. By clustering the mirrors into coherent groups, new insights into the spatial and chronological distribution were also proposed with a view to gaining a better understanding of Etruscan religion and daily life, the engravers' expertise, and their workshop organization, as well as the progressive shift toward simplification and standardization in decorative schemes, likely driven by evolving artistic conventions and client demands (*fig. 12*).

In subsequent years, the research program on the 'Automatization of Etruscan Corpora' was extended to include the corpus of approximately 1,200 stone cinerary urns produced in Volterra during the Hellenistic period (the *Volaterrae* project). The study initially collected data from published catalogues: the *Corpus delle Urne Etrusche di età ellenistica* (CUE), museum collections (such as the Guarnacci Museum in Volterra and the Archaeological Museum in Florence), and private collections. The study was divided into two research areas, one relevant to the lids and the other to the chests (*fig. 13*). This approach responded to the difficulty of correctly associating lids and chests in the absence of reliable excavation data and of managing such a large number of variables.

The corpus of scenes carved on the front of the chests was thus analyzed, focusing on the interrelationship between morphology and iconography, as well as on the chronological distribution of mythological or funerary themes (MOSCATI 1990, 1994b, 1995a, 2004). The formal description of architectural moldings – considered in the literature as distinctive markers of individual workshops – facilitated the identification of manufacturing techniques based on the stone-carving practices of local artisans (*fig. 14a*). The quantitative analysis revealed some characteristic elements of standard production, in contrast to high-quality pieces adorned with framing decorations, which reflect the deceased's desire to celebrate their social and civic engagement.

The study then focused on 351 chests featuring Etruscan funerary iconography (MOSCATI 1997): *dextrarum iunctio*; the deceased summoning his wife to the underworld; the journey to the underworld (on foot, on horseback, *in carpentum*, or in a four-horse chariot); procession of a magistrate to the underworld (*fig. 14b*). Descriptive and mul-

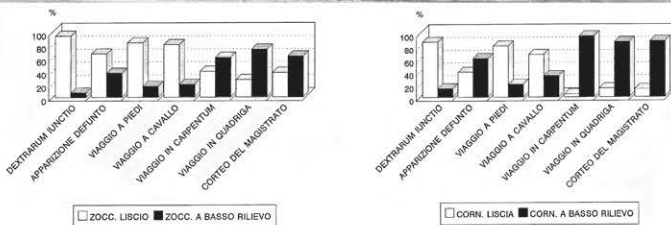


Crosstabulation: CG
By ZF

ZF →	Count Row Pct Col Pct	CG		Row Total
		0,0	1,00	
CG	0,0	508 93,0 92,5	38 7,0 53,5	546 88,1
	1,00	41 55,4 7,5	33 44,6 46,5	74 11,9
Column Total		549 88,5	71 11,5	620 100,0

Chi-Square	D.F.	Significance	Min E.F.	Cells with E.F. < 5
87,35263	1	,0000	8,474	None
91,02624	1	,0000	(Before Yates Correction)	

a



b

fig. 14 – Descriptive statistics applied to the analysis of Etruscan urn chests. a) Crosstabulation of the variables describing the upper and lower frames in an urn chest representing the myth of Odysseus and the Sirens. b) Statistical distribution of the chests featuring Etruscan funerary iconography.

tivariate statistical methods were utilized to examine the distribution of variables and their interrelationships. The *dextrarum iunctio* group, comprising 106 chests, exhibited remarkable uniformity in the use of tufa, indicating a standardized production. In contrast, the 68 chests depicting the deceased summoning his wife showed greater diversity in materials and styles, suggesting distinct workshop practices. The journey to the underworld presented notable typological richness, while chests representing this journey *in carpentum* tended to be more uniform. Meanwhile, the chests featuring the four-horse chariot and magistrate's procession revealed intricate craftsmanship, hinting at their high-status origins.

This quantitative study validated and refined existing knowledge, confirming that technical and typological features are primary classification criteria, while iconographic variables serve to track workshop choices and chronological trends. The study also explored framing elements as potential workshop markers. Finally, the methodological approach emphasized that preliminary iconographic classification must be tested against a comprehensive dataset to ensure reliability. The study proved that quantitative analysis is a valuable tool in archaeological classification, helping to reveal production trends and workshop organization in Hellenistic Volaterran urns.

Alessandro Di Ludovico

6. TEXTUAL ANALYSIS IN ARCHAEOLOGY: PROPOSALS, METHODS, AND APPLICATIONS

To recover the uniqueness of a phenomenon, through its several versions – which may all be correct, or at least equivalent as to the degree of approximation – one would have to put them through a sort of rectifier which would polarize observations around a few “points,” chosen according to a series of reasoned conventions. To a certain extent, such procedure recalls that adopted for the analysis of archeological data (GARDIN 1958, 355).

6.1 TEXTUAL ANALYSIS: ASSUMPTIONS AND NEEDS OF ARCHAEOLOGICAL RESEARCH

It is perhaps significant that a substantial part of the experiments in textual analysis of ancient artifacts has been conducted in the context of studies on pre-classical Western Asia. The cultures referable to antiquity in this vast region have for centuries remained wholly or partially unknown and forgotten, only gradually to re-emerge in the knowledge of scholars and the general public over the past two centuries. This propitiated the emergence of new disciplines, with their corresponding specialists, but also progressively confronted scholars with the hermeneutic difficulties that research on largely forgotten cultures poses. While the biblical tradition and the tradition of classical studies may have given some scholars the impression that they had firm historical and cultural reference points, the scientific community realized relatively late that it had to deal with centuries-old cultural systems, some of which had little in common with European cultures (see, e.g., TRIGGER 1996, 41-43).

The creativity of a few pioneering researchers led them to undertake complex and systematic investigations that resulted in considerable progress in epigraphy and philology. Important discoveries regarding the functioning mechanism of Mesopotamian cuneiform writing, as well as the first repertories of signs, were the outcome of investigations marked by intensive analytical research, a remarkable formalization of graphic data and a careful examination of the relationships and recurrences of the minimal elements identified (FRIEDRICH 1954). In other words, the fundamental contributions to the understanding of ancient Mesopotamian writing systems derive from systematic analyses of data that are structurally comparable to textual analyses, only explicitly considered from the 1950s onwards. Moreover, the protagonists of such pioneering studies include several researchers trained in the mathematical and natural sciences

(FALKENSTEIN 1964, 1-5; EDZARD 1976-1980, 544; MARSAL 2014, 2015; STRECK 2006-2008, 277; ANDRÉ-SALVINI 2003, 117-118; 2013, 419-420).

Epigraphists and philologists particularly felt the need to arrive at solid points of reference when it came to identifying the operating principles of writing systems and the different languages they expressed. It was essential for them to recognize the structure and function of the graphic sign system used, on the one hand, and the nature – and possibly also the family to which it belonged – of the language that was expressed in the specific epigraphic specimens, on the other. These were indispensable preliminary operations for the initiation of the reconstruction of syntax, morphology and vocabulary allowing for the decipherment of the texts. In this way, the brilliant scholars mentioned above were able to recognize and begin investigating languages such as Sumerian, Elamite, Old Persian, Akkadian, and Hittite. On the other hand, researchers who engaged in the study of the artifacts of preclassical West Asian cultures and their figurative languages were under the illusion that they could approach them with greater decisiveness and less hermeneutic and methodological caution, partly by virtue of ethnocentric prejudices that have long contaminated research (see, e.g.: MOOREY 1991; MATTHEWS 2003, 1-32; LIVERANI 2005; POLLOCK, BERNBECK 2005, 6-8).

With the progress of field investigations and related research, between the second half of the 19th and the first half of the 20th centuries, the scientific world acquired an increasingly less biased and more critical and calibrated view of the artistic, craft and architectural productions of the cultures in question. The roots of the various art-historical and archaeological currents were, however, still in the classical and European tradition, without diverging from it significantly. It cannot be said that this approach has been revolutionized in more recent years, but certainly the approach of much of the scientific world has been refined considerably, equipping itself with complex critical and theoretical tools that are often the result of highly productive interdisciplinary reflections.

6.2 TEXTUAL ANALYSIS AND ARCHAEOLOGY: PREMISES, BEGINNINGS AND DEVELOPMENTS

The 1950s marked a moment of transition in which some scholars decided to adopt an approach that would prove both mature and experimental (if not *avant-garde*) towards the visual languages of ancient cultures, starting with those of West Asia. The challenge of understanding such languages and their underlying structures proved to be particularly stimulating, since the researchers in question were now aware that they had to deal with diverse and highly complex communication systems. Particularly in the case of ancient Near Eastern cultures, once they had embarked on the road of the in-depth formal investigation of iconographies, they also had to be ready to reconsider the historical periodization and cultural reference horizons that were habitually accepted and shared in the scientific world.

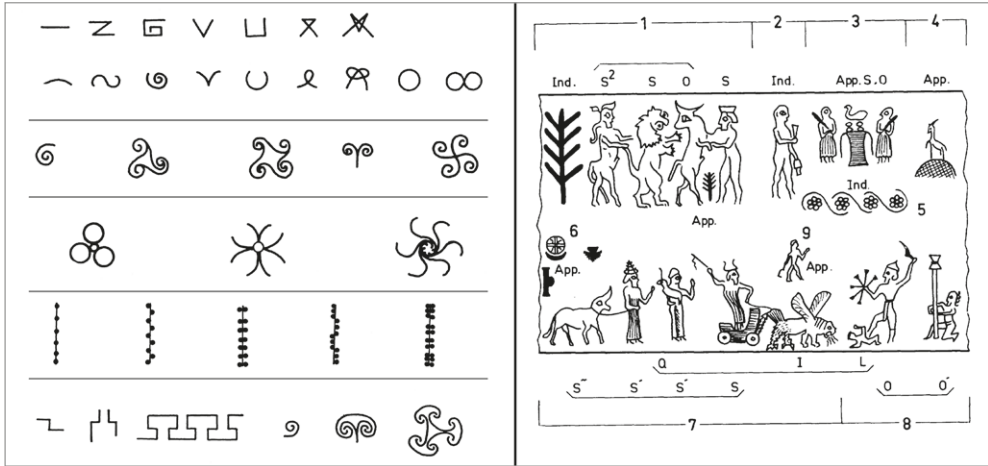


fig. 1 – Left: Formalizations of signs from a corpus of decorations, coded according to an arbitrary system which, by the agglomeration of predetermined syllables, makes it possible to record their number, arrangement, and combination. From the first row (above): 1) elementary signs of decoration; 2) examples of combinations of the ‘spiral’ element; from left to right: ‘spi’, ‘spiti’, ‘spika’, ‘spim’, ‘spiro’; 3) various types of rosettes with branches of different number and type; 4) types of symmetrical repetition of elementary signs within a sequence; from left to right: ‘ma-’, ‘mi-’, ‘mo-’, ‘mu-’, ‘mou-’; 5) examples of combining elementary characters; from left to right, ‘z’, ‘zim’, ‘zimuli’, ‘spi’, ‘spimimb’, ‘spimimbuti’ (GARDIN 1958, 342-344). Right: example of a segmentation scheme of West Asian iconographies (GARDIN 1967, 23, fig. 19).

As mentioned above, the systematic codification and formalization of basic archaeological data constituted the first fundamental issues to be addressed. These are complex operations that are usually planned and calibrated according to the needs of the specific investigation and the tools and procedures it makes use of, whereas their declination in a universal key is a very ambitious project, even if it remains within very general parameters (which is inevitable). In addition to weighing procedural choices according to the objectives, the coding and formalization of the data impose on the scholar an intensely critical and self-critical reflection on how to read and interpret the data. Indeed, he or she will have to establish the modalities of punctuation and segmentation of the figurative representations and the underlying conceptual and logical phenomena. This is already a research activity in which the scholar is called upon to properly ponder and analyze his or her view of ancient cultures and their products, possibly identifying preconceptions and fundamental flaws in certain interpretations that are widely shared in the scientific world. Against the backdrop of these themes and principles, it seems significant and emblematic that the first innovative uses of quantitative approaches in archaeology were developed within the Institut Français d’Archéologie in Beirut at a time when Jean-Claude Gardin and his colleagues were developing important methodological reflections.

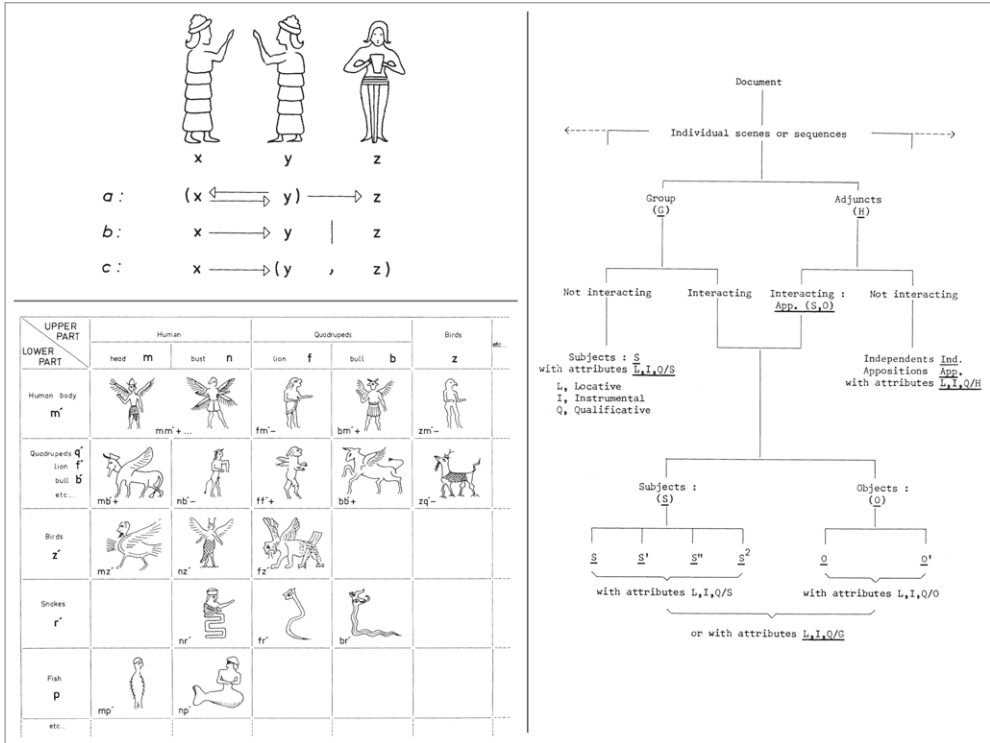


fig. 2 – Left: examples of different interpretations of the same composition (top) and example of a differentiation scheme of West Asian iconographies of ‘Hybrids’ (bottom). Right: explanatory diagram of the analysis of a pictorial representation (GARDIN 1967, 16, 22, 26, figs. 6, 20, tab. 2).

Towards the end of the 1950s, Gardin, who had founded the Centre d’Analyse Documentaire pour l’Archéologie in 1957, outlined the basic principles of a system for the quantitative and automated study of artifacts belonging to different classes (GARDIN 1958, 335-336; 1967, 13; for the general principles, see GARDIN 1955). The ideas he expounded with such clarity stemmed from his many years of reflection on the methods and problems of classifying and interpreting archaeological data, as well as the proposals and ideas that were emerging internationally at the time (fig. 1). Since then, one of the main themes of his research has been the development – with the aim of promoting widespread adoption – of a system of universal codes that could adequately represent and describe artifacts of archaeological and art-historical interest in sufficient depth to enable scholars to publish, compare and classify materials with flexibility. This approach was based on the assumptions of the French logicist school (or *analyse logiciste*) and aimed at constructing a universal language that, as mentioned, would be able to describe any archaeological artifact with both accuracy and information richness.

Inspired mainly by linguistics, Gardin wanted to base his system on a code with a high level of productivity, which could be used systematically to describe, with a small

number of elements, all the information needed for any kind of finding (GARDIN 1966; 1967, 18-26). At first, therefore, he developed a series of distinct codes, each of which could be used to describe a macro-category of artifacts (glyptics, ceramics, tools, etc.) in a suitably rigorous and systematic manner. These codes, structurally similar to one another but differing in their ontologies and applications, constituted an initial milestone, the first important stage of a quest whose declared objective was to pursue an evolution leading to a formal system that could have universal use (*fig. 2*).

The concrete starting point was the assumption that the coding of objects should be based on three fundamental concepts: orientation, segmentation and differentiation. The individual components of the object could then be isolated and related to each other via a system of alphanumeric labels univocally associated with them (GARDIN 1967, 13-18). The data collected in this way were then recorded on perforated cards, which made them computer-organizable and usable. The scholars of the Centre National de la Recherche Scientifique (CNRS), who belonged, on a theoretical-methodological level, to the school of 'logistic archaeology', organized and actively engaged in the years to come in the design and realization of a considerable number of systematic and reasoned experiments and catalogues aimed at collecting rigorous descriptions of the evidence of antiquity.

The logicist school distinguished itself in the international context not so much as an alternative to theories developed in the Anglo-American sphere, but rather as a different point of view, original and potentially integrative of them. It is, moreover, a school of thought that, while remaining recognizable in its general approach, has expressed itself, in theory and practice, in very different ways, depending on the personal attitudes of the scholars and the contexts or scientific questions they were facing. One characteristic of the logicist approach is undoubtedly the constant focus on the attitude and perspective of the scholar, and thus an uninterrupted tendency towards reflexivity and critical consideration of the researcher's point of view. The consequence of this is a focus on archaeological discourse, as well as on its terminologies and the ways in which data is formalized (for a general view on the logicist school, see, among others: GALLAY 1989, 2007; GALLAY, GARDIN 2009).

In the intellectual context just described, for example, a series of interesting and groundbreaking proposals for universal codification of classes of materials, such as those for ceramic forms, decorations, coins, Greek vase paintings, etc., developed (see, e.g., LE RIDER 1975; GARDIN 1976, 1978; SALOMÉ 1980). Also within the intellectual milieu of the CNRS in Paris, various investigations and experiments based on lexicographic and semiotic concepts sought to develop suitable coding systems and datasets for archaeological research (see, e.g., GINOUVÈS, GUIMIER-SORBETS 1978).

Following the same logic, and in the same period, Gardin and his collaborators implemented the first proposal for a computational treatment of ancient West Asian visual languages. The iconographies of some cylinder seals were the subject of one of the most impressive applications of this first extensive experiment, translated into a formalization based on already carefully outlined criteria (GARDIN 1967, 21-26). The expansion of

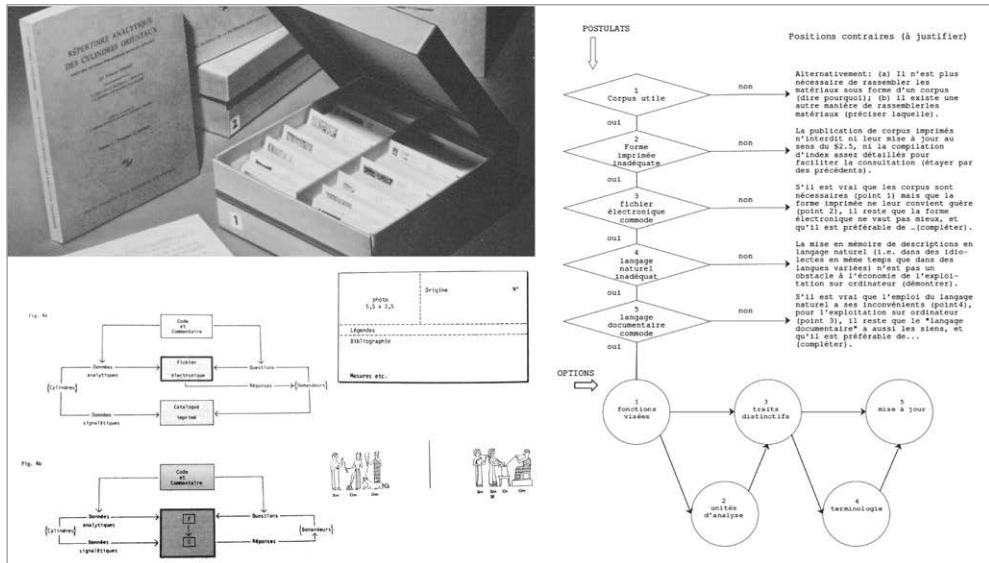


fig. 3 – The *Répertoire analytique des cylindres orientaux* (DIGARD 1975), top left, and excerpts from the description of its content and structure (graphical elaborations).

the project and the use of the tested system on a larger scale led to the consolidation of collaboration between the scholars involved. The result was a body of work in which numerous French scholars participated, mostly working in separate locations, in the retrieval and adequate recording of data on cylinder seals whose primary publications were not museum catalogues but rather excavation reports or heterogeneous articles scattered among different international journals (thus having been studied and analyzed at different times and by different authors).

The result of this effort was an admirable and voluminous catalogue, the *Répertoire analytique des cylindres orientaux* (DIGARD 1975), in which the essential information for each seal was collected rigorously and in uniform format cards (fig. 3). The *Répertoire* was a particularly suitable experiment to illustrate concretely the ideas and intentions of the research group of the Institut Français d'Archéologie in Beirut. Not only did it clearly express the methodological and theoretical elaborations of the logicist school but it was also a useful tool for research, ready to be used by other scholars. It was a collection of cylinder seal encodings that could be expanded at any time (e.g. following the discovery of new specimens), within which automated searches and comparisons could be carried out. It was a structured, orderly and simultaneously flexible archive that could support research of various kinds.

The codings devised and developed for the *Répertoire* and the other works mentioned constituted archives of rigorously formalized expressions, on occasion very concise, composed and condensed in a manner comparable to the syntax of a natural language. These characteristics allowed for an agile archival procedure, secure data storage and

the possibility of retrieving information easily and without dispersion. The *Répertoire* and similar archives remained as basic research tools with the potential for secondary interventions, for example work that employed stored data to carry out quantitative research. The latter did not find ready-made approaches in the archives themselves, as they were rightly expected to occasionally be developed in accordance with the scientific requirements of scholars and their research tools.

As an instrument, the *Répertoire* was warmly welcomed by the scientific community, but it was never expanded as it should have been, so it slightly outgrew the embryonic form of the experiment that it had been. This is (paradoxically) due to its nature of free and open source archive *ante litteram*: as Gardin himself imagined, very few if any scholars would have wanted to share their primary materials, and perhaps even fewer would have given up their own time to contribute to the expansion of such a shared resource. However, the contribution of the *Répertoire* to the development of the scientific community and the use of new digital technologies in the research into ancient cultures has been of great importance. On one hand, considering the methodological aspects, linguistics was identified as a stable and reliable point of reference and the need for a flexible and dynamic investigation tool for the future quantitative and qualitative glyptic (and comparable) studies was affirmed. On the other hand, there was an explicit stress on the importance of a conscious reflection on the interpretation and the recognition of typologies in archaeological reasoning: this has been stated before and would be further developed by Gardin in the years that followed (GARDIN 1958, 350-355; 1967, 28-29; 1997; 2002; MOSCATI 2013, 7).

In the 1970s a rigorous textual formalization system for ancient Western Asiatic glyptic was also developed and experimented with in California, by researchers at UCLA, led by Marilyn Kelly-Buccellati (KELLY-BUCCELLATI, ELSTER 1973; KELLY-BUCCELLATI 1977, 1979-1980). Initially, this experiment was part of a project called 'Computer Aided Analysis of Mesopotamian Material', which sought to cover a broad spectrum of classes of materials (KELLY-BUCCELLATI 1977, 45, n. 2). However, only the section related to glyptic was fully developed. Although pursuing different goals and organized according to different logics, this experiment started from the codifications of the Centre d'Analyse Documentaire pour l'Archéologie of the CNRS (KELLY-BUCCELLATI, ELSTER 1973, 200). Thus, the textual codings of the iconographies of cylinder seals were arranged according to procedures that allowed for rigorous descriptions of the specimens and facilitated quantitative studies. In this case, however, they rather followed a binary system. Then, such codings enabled statistical analyses of themes of cylinder seals and comparisons or assessments of proximity between the subjects represented (whole compositions or elements of the compositions that can be considered comparable – *fig. 4*).

According to the plans, the first concrete and meaningful product of this project was intended to be a large digital catalogue that included all possible logical combinations among the features recorded in the specimens. This means that, unlike the CNRS catalogue, this was a closed catalogue that collected a very large sequence of retrievals already carried out and made available to the user. The project developed at UCLA

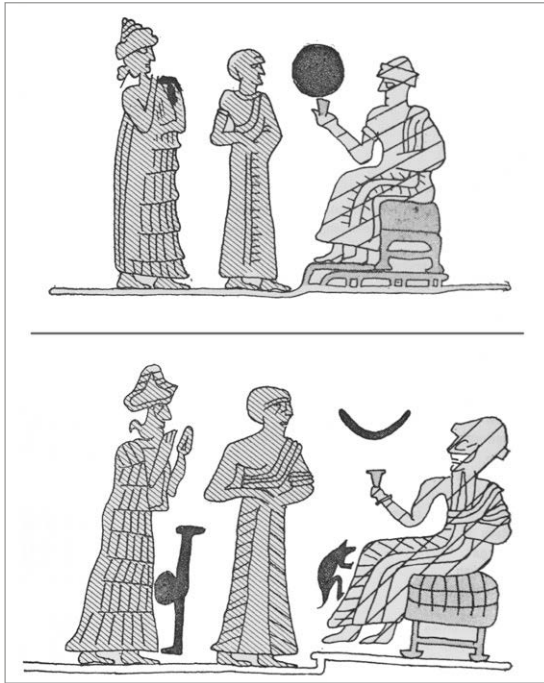


fig. 4 – Example of an iconographic concordance between a seal from Ugarit, top, and an Old Babylonian specimen, bottom (KELLY-BUCCELLATI 1979-1980, 277).

seems never to have been enlarged as originally intended, and to have ended after the first large experiments with Old Babylonian glyptic production.

However, in both the open and dynamic catalogue developed at CNRS and the closed UCLA catalogue, largely composed of pre-processed information, textual analysis was relegated to the preliminary formation phase of the archive itself, that is to the construction of the consistently formalized dataset, as well as the possible phases of data retrieval.

A largely novel approach was proposed some years later by Paola Moscati in a research on Etruscan bronze mirrors (MOSCATI 1984, 1986). This work was methodologically aligned with the tradition of the French logicist school (see the considerations in MOSCATI 1986, 11-31), but unlike the projects mentioned above it was conceived from its inception as a quantitative research on a specific corpus of ancient artifacts (see Chap. 4). In a preliminary stage of this research, Moscati carried out a supervised classification of the artifacts, while the follow-up focused on more complex and productive data mining approaches by employing multivariate statistical methods. Moscati's research was carried out and published in the same period in which, in the field of computer applications in archaeological studies, this transition from data management to data processing («passer de la gestion des données aux traitements des données», DUCASSE 1986, 292) was recognized as both necessary and timely, as the moment seemed ripe for it (see also DJINDJIAN 1986b).

These research experiences by Paola Moscati did not involve the use of textual coding for the artifacts, but they introduced statistical methods particularly well-suited for

investigating iconographies of relatively homogeneous classes of materials. For these reasons, her work played a role in the first developments of Textual Correspondence Analysis (TCA), carried out in the 1990s (see below).

6.3 TEXTUAL ANALYSIS AND THE FRENCH SCHOOL OF DATA ANALYSIS: THE INVESTIGATION OF VISUAL LANGUAGES

From a theoretical and methodological perspective, a significant number of applications of textual data analysis – or text mining – in archaeology owe much to the research in computational linguistics conducted and developed by Jean-Paul Benzécri and his collaborators and colleagues in France, starting at least in the 1960s. Benzécri identified statistics, and particularly correspondence analysis, as the most appropriate methodological framework for advancing his research, thanks to its versatility and capacity to ascertain proximity and associations between individuals and variables documented in a data set and its capability to project the outcomes of each analysis onto a map and obtain parameters useful for evaluating the qualitative level of the projection itself. To facilitate the interpretation of results and to complete the analysis, he also proposed the use of clustering algorithms (see, e.g., BENZÉCRI 1992; BEAUDOUIN 2016).

Correspondence analysis enables the association of groups of elements, through the provision of geometric representations for these associations. Benzécri systematically oriented his work toward developing methods that could not only have a broad range of practical applications but also remain as manageable as possible to those not particularly familiar with statistics. The significant advancements in the field of electronic computer design during the latter third of the 20th century played a crucial role in the realization of these objectives. Notably, many researchers who collaborated with Benzécri or were trained in his school went on to develop free or open-source computer applications, with the objective of promoting constant implementation, and portability of their algorithms. In general, the algorithms and methods in question have been especially successful in Mediterranean countries, but they have met with less favor in the Anglo-American world.

Ludovic Lebart was among the earliest collaborators of Benzécri and has served as a member of the scientific committee of the Journées Internationales d'Analyse des Données Textuelles (JADT), a conference series founded in 1990, since its inception. During his extended tenure at Crédoc (Centre de recherche pour l'observation des conditions de vie), Lebart engaged in extensive research in the domain of textual statistics. A notable achievement of this period was his collaboration with Alain Morineau, which resulted in the development of the SPAD system (Système Portable pour l'Analyse des Données; LEBART, MORINEAU 1982).

In the early 1990s, the debate regarding the coding and quantitative investigation of archaeological materials became increasingly intense and productive (see, e.g., DJINDJIAN 1991 and the fifth issue of the journal «Archeologia e Calcolatori», published in 1994 and dedicated to *Choice, Representation and Structuring of Archaeological Infor-*

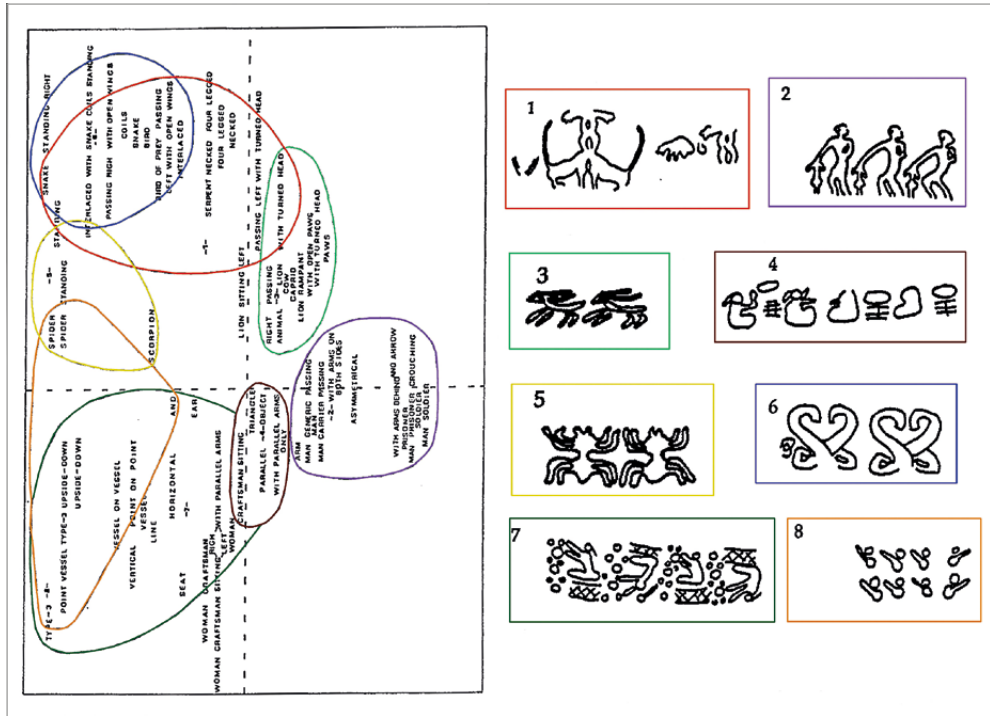


fig. 5 – Correspondence analysis of protohistoric seals: most characteristic seals of the eight clusters and distribution of their centroids on the first two factors (graphic reworking of CAMIZ, ROVA 1991, figs. 1, 3).

tion (MOSCATI 1994a). In this context, explicit references emerged to the possibility of enriching the investigation of archaeological artifacts through the use of statistical methods applied to textual data (see, e.g., GUERMANDI, MIGNANI, MONTANARI 1994, 129-130), after pioneering research adopting TCA had already been carried out by the mathematician and statistician Sergio Camiz and the archaeologist Elena Rova (CAMIZ, ROVA 1991).

The common main reference, methodologically speaking, was the work by Lebart and Morineau, which has since been expanded, refined and updated until recent years, also through the intervention of other scholars (LEBART, SALEM 1988, 1994; LEBART, SALEM, BERRY 1998; LEBART, PINCEMIN, POU DAT 2019).

Thus, about fifteen years after the first published proposals for the archiving of Western Asiatic cylinder seals based on highly formalized textual coding, a new wave of studies on the same class of materials was conceived and developed at Rome University ‘La Sapienza’. Camiz and Rova experimented with different types of codings in order to perform exploratory analyses on a huge and seemingly heterogeneous corpus of 963 iconographies recorded on protohistoric cylinder seals or seal impressions. In their first experiment, they envisioned the use of three types of coding, but developed only the

first two, postponing the application of the third, concerning the syntactic structure of composition, to a future date (*fig. 5*).

The first coding recorded the presence or absence of specific elements in the composition and formed a dataset which was investigated using multiple correspondence analysis, to detect proximities and differences among subjects and places of origin or hypothetical function of the seals, and cluster analysis, to group the specimens into classes (the tool used was SPAD-N: LEBART, MORINEAU, LAMBERT 1987). The second coding was textual, which means that it recorded iconographies in the form of highly formalized and homogeneously structured texts, and this proved particularly useful for the description and investigation of the sets of elements, or sub-patterns, of the compositions, as well as their single elements. Both the textual coding and the first experiments on it were very innovative since for the first time they not only allowed for the recording of the presence of elements, sub-patterns and patterns singled out in the compositions, but also the spatial, and possibly logical, relations between them (CAMIZ, ROVA 1991, 131-135).

The patterns and sub-patterns could be detected and studied thanks to tools that helped in locating the ‘repeated segments’ of the dataset (SPAD.T: LEBART, MORINEAU, BÉCUE 1988). Inevitably the textual coding required a series of attempts and corrections (following the problems observed on occasion), which were carried out in

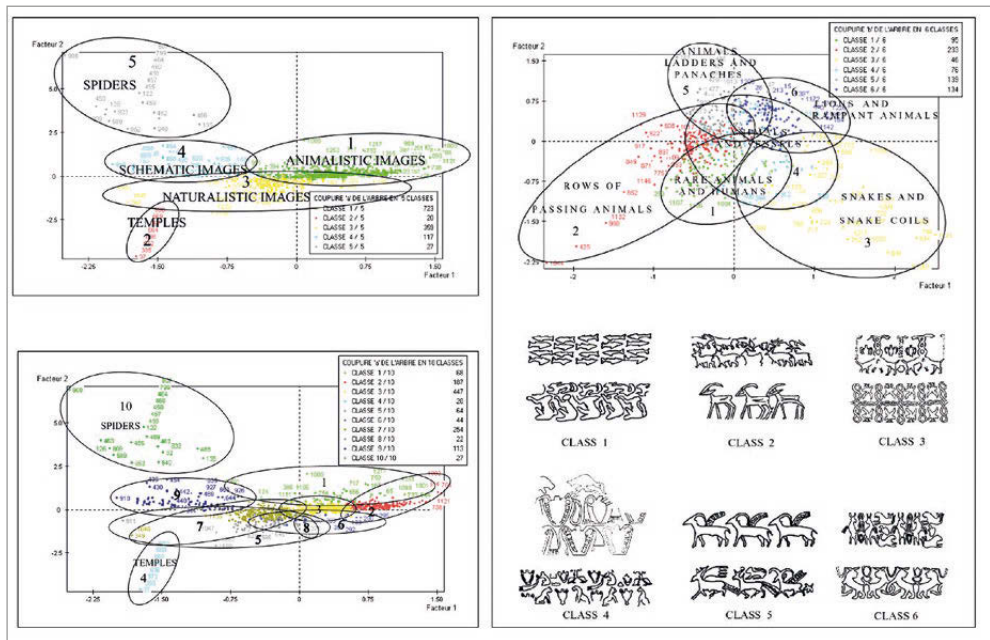


fig. 6 – Correspondence analysis: projection of classes on the first two factorial axes. Left: results of the analysis on the whole corpus (1247 seals iconographies) with the projection of five (top) and ten (bottom) classes. Right: results of the analysis on the corpus of animalistic iconographies (723 specimens) with the projection of six classes (excerpts from CAMIZ, ROVA 2003b, figs. 4-5, 8).



fig. 7 – The multimedia cultural itinerary of the Virtual Museum of Archaeological Computing dedicated to ‘Quantitative approach to ancient Near Eastern glyptic’ (CAMIZ, DI LUDOVICO, ROVA 2017).

a number of subsequent projects. While preparing a general synthesis of the first series of experiments (ROVA 1994), an expansion of the dataset enabled the scholars to obtain results that were both more satisfying and more readable (CAMIZ, ROVA 1992, 1993a, 1993b, 1996, 2001, 2003a, 2003b; ROVA 2005 – *fig. 6*). These also stimulated new theoretical-methodological reflections, both regarding the coding of archaeological material in general, and more specifically about the corpus and the optimal ways to encode it and represent its internal structures (CAMIZ 1994, 2004; ROVA 1995, 1996, 2000). The various refinements in the theory and preparation of textual datasets also made it possible, at last, to conduct experiments on the syntactical structure of the compositions investigated (CAMIZ, ROVA, TULLI 1998, 2003, 2010).

Quantitative instruments were, in these cases, embedded in investigations of visual languages that sought to share the logic of these languages, while the chosen model interacted dynamically with the corpus. The work developed by Rova and Camiz served as a precedent from which to draw inspiration, primarily from perspective of logic, the underlying scientific questions, and the type of quantitative investigations to be carried out (see CAMIZ, DI LUDOVICO, ROVA 2017 – *fig. 7*).

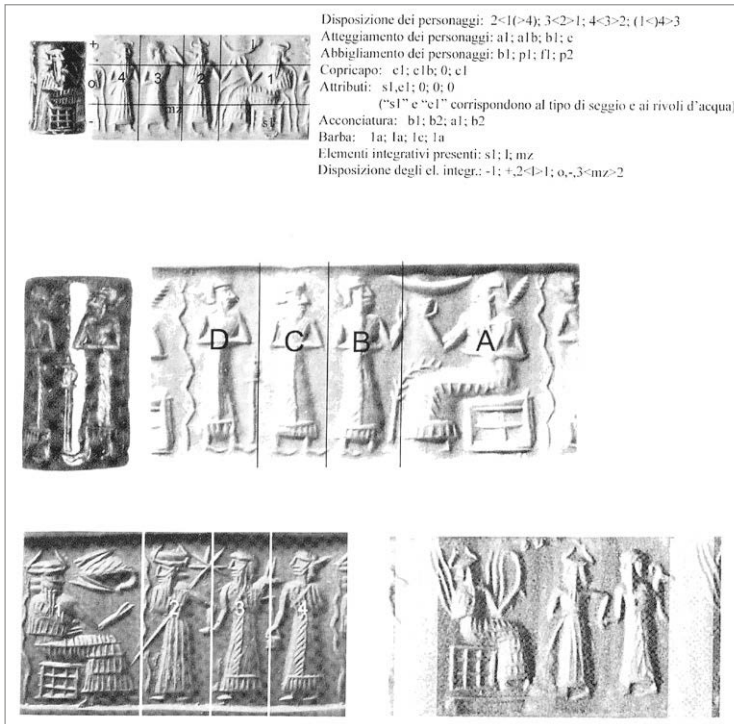


fig. 8 – Encoding criteria and experiments for third millennium presentation scenes (excerpts from DI LUDOVICO 2005, 90-92).

Similar principles were the basis for a research project on Mesopotamian glyptics at the end of the third millennium, which was initiated by Alessandro Di Ludovico in the early 2000s (DI LUDOVICO 2005). Initially, this work developed as a qualitative and quantitative study on the evolution of the *presentation* theme in the glyptic of Lower Mesopotamia from the Akkadian Age to the Third Dynasty of Ur (ca 2350-2000 BC). From the outset the basic approaches were essentially based on structural linguistics and led to an initial analogical classification of specimens, and a reconstruction of the origins and diachronic transformations of a significant number of the iconographic traits recorded (*fig. 8*).

A major and important innovation in the coding of this and the following datasets is the fact that, for the first time, each element of the composition is recorded in its relative position, with respect to the other elements, and absolutely, with respect to the cylindrical surface of the seal, and not in relation to its artificial projection on the plane. Alphanumeric and fuzzy logic-based encodings were later adapted to a binary system (presence/absence) for investigations using Neural Network-type algorithms (DI LUDOVICO, RAMAZZOTTI 2008; DI LUDOVICO 2011). The procedure was rather complex and once again involved a critical rethinking of the chosen categories, and the internal segmentation of iconographies. Further adaptations were thus necessary for subsequent works in which, on the one hand, the chronological horizon of the corpus was narrowed down to the Third Dynasty period only, and, on the other, a path was sought for the automated reproduction of the internal structure of the presentation

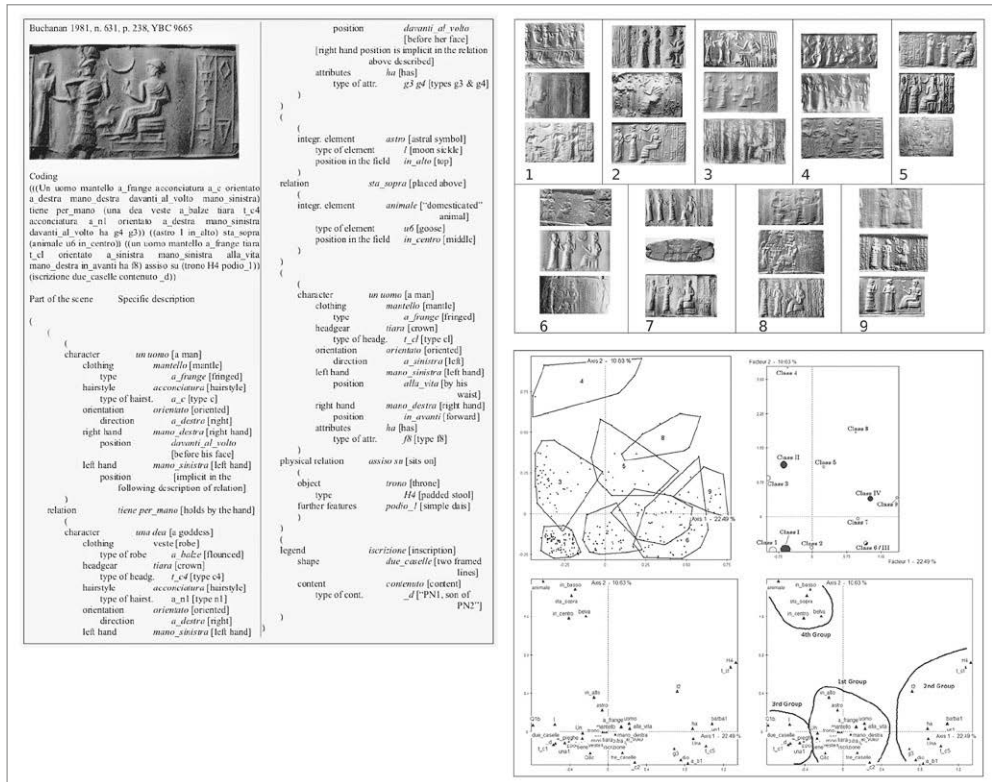


fig. 9 – Textual encoding and classification of seals’ presentation scenes (DI LUDOVICO, CAMIZ 2014, figs. 2, 6, 8).

scenes (DI LUDOVICO 2011; DI LUDOVICO, PIERI 2011a, 2011b) while the basic logic and general vision of the coding remained almost unchanged.

Some way in, this path of research made a breakthrough, enriched by the expansion to the field of statistics, starting with an initial experiment comparing the performance of Neural Networks and statistical applications with respect to the cultural and art-historical issues being investigated (DI LUDOVICO, CAMIZ, PIERI 2013). The elaborations carried out using statistical methods involved Sergio Camiz, and the algorithms are of the same type as those previously used by Camiz and Roa, thus mainly referring to the TCA (DI LUDOVICO, CAMIZ 2014, 2015; DI LUDOVICO 2018). The dataset was originally made of 354 textual codings describing the same number of scenes.

The TCA was first carried out on the forms (the minimal elements making up the descriptions), whose distribution and participation in the most significant factors gave clear indications about the main features distinguishing the specimens of presentation scenes. These results also permitted some of the features typical of the geographic areas of origin of the scenes to be identified. The classifications of the compositions (obtained using a Hierarchical Ascendant Classification tool) into nine classes were much

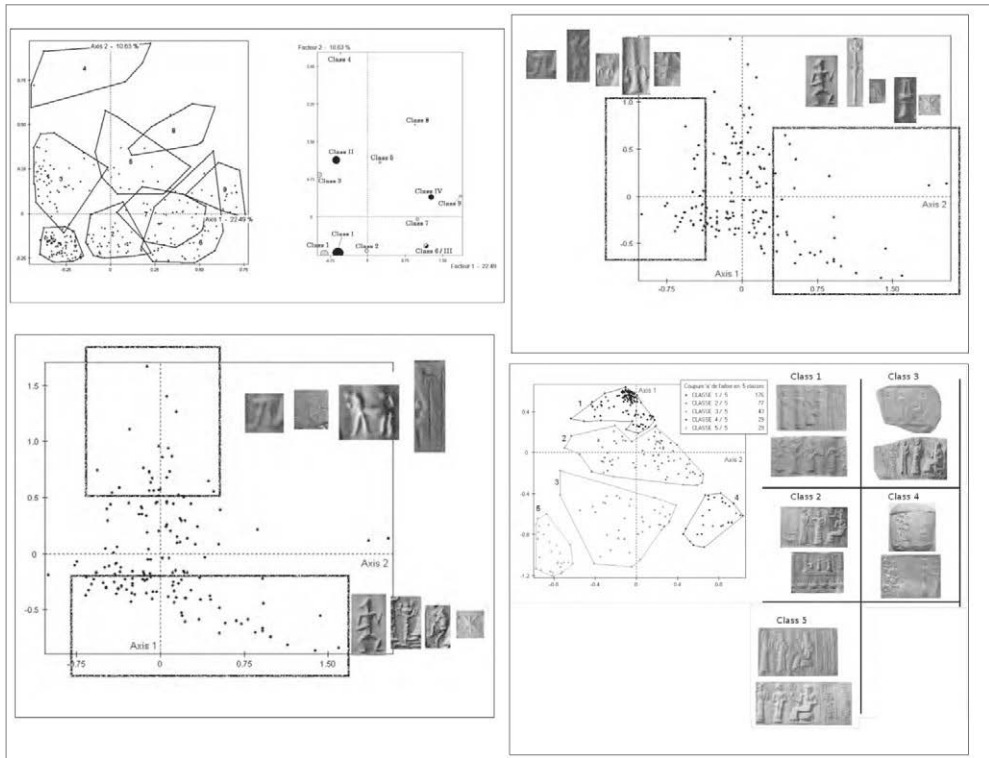


fig. 10 – Automated recognition, correspondence analysis and classification of the repeated segments of *presentation scenes* (from DI LUDOVICO, CAMIZ 2015, figs. 3, 4, 5, 7): classification of the scenes through the forms (top-left); correspondence between legends and integrating motifs (bottom-left and top-right); classification of the scenes through the repeated segments (bottom-right).

clearer (DI LUDOVICO, CAMIZ 2014 – fig. 9). The second part of the investigation of the same dataset was the automated recognition and analysis of the repeated segments (DI LUDOVICO, CAMIZ 2015). The overall results were not much different from those of the forms, but there were many more possibilities for the reconstruction of well-structured patterns of the visual languages of *presentation scenes* (fig. 10). For a third investigation of these artifacts, the dataset was enlarged (425 textually coded specimens) in order to improve the recognition of similarities and differences among the places of origin of the scenes (DI LUDOVICO 2018). The results were more than encouraging, prompting the authors of this research to investigate it further, as far as was allowed by the primary data available. An attempt to thin the dataset by means of Procrustes analysis was decided on, trying to reduce the internal noise and obtain more explicit results, as long as there was no possibility of considerably integrating the dataset (DI LUDOVICO, CAMIZ 2020, 2022).

Compared to the dataset of protohistoric seals investigated by Camiz and Rova, the one describing the *presentation scenes* was chronologically more compact and typologi-

cally much more homogeneous, providing considerable advantages for the use of the above-mentioned analysis and classification methods. However, considering current developments in research and applications which can be used for quantitative research, it can be assumed that textual analysis practices such as those used for the iconographic studies described above are likely to experience interesting developments in the near future.

Simona Marchesini proposed a particular use of correspondence analysis in the early 2000s to investigate the diachronic evolution of the form of signs in some groups of Italic inscriptions, using the Bonn Archaeological Software Package (BASP), which has been developed in Germany since the 1970s. The approach and the algorithms were innovative, but to our knowledge the project never progressed beyond the first phase, devoted to seriation (MARCHESINI 2004, 2010).

A similar approach is found in a more recent, yet more in-depth research with particularly ambitious goals, likely requiring further investigation in the near future. In his PhD dissertation, Andrea Santamaria used correspondence analysis to examine the morphological evolution of the Cretan hieroglyphic system and its geographical and chronological relationship to Linear A (SANTAMARIA 2023). Utilizing dedicated packages in the free software R, Santamaria investigated the relationship between hieroglyphs and the shape and material of the stamp seals on which they appear.

The research experiences of both Marchesini and Santamaria evoke the pioneering works referenced at the outset of this paper. Utilizing an archaeological and quantitative approach, they address inquiries primarily within the domain of epigraphy and philology.

6.4 TEXTUAL ANALYSIS FOR METHODOLOGICAL AND META-ARCHAEOLOGICAL RESEARCH

As previously mentioned, Benzécri's school of *Analyse des Données* played a predominant role in inspiring statistical applications for investigations into concrete textual corpora in archaeological studies. However, a salient point that merits attention is the persistent influence of the logicist approach on the theoretical underpinnings of quantitative archaeological research. At present, some of the initial research proposals and experiences stemming from reflections within the logicist school, and concerning the domain of coding and textual analysis in archaeology, are demonstrating their vitality and long-term solidity.

One of the areas in which the aims of the logicist school have recently found new life remains that of ancient West Asian glyptics. More than forty years after the birth and the rapid end of the CNRS *Répertoire* project, Elisa Roßberger and Anna Kurmangaliev developed the project 'Concept for the Digitization and Labeling of Ancient Near Eastern Seals and Sealings' (DigANES). In terms of content and technical concept, this project can be considered a new version of the *Répertoire*, although it was built from scratch and is totally independent from it. From a methodological and technological point of view, DigANES is in fact an updating to the linked world of the 2010s-2020s

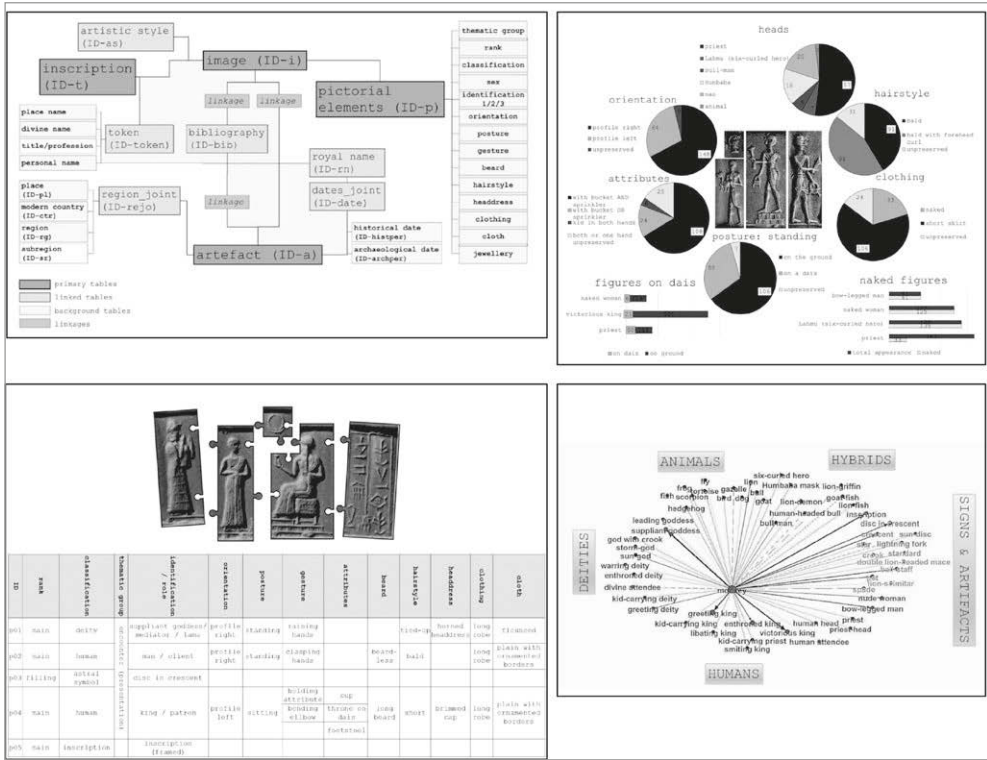


fig. 11 – Structure and tests of the project DigANES/ACAWAI-CS (from ROßBERGER, KURMANGALIEV 2023, figs. 4.1, 4.2, 4.5, 4.6).

(and the Linked Open Data principles) of the main ideas on which the *Répertoire* was based. After 2017, DigANES was reformulated and enriched in a new guise, giving rise to the project ‘Annotated Corpus of Ancient West Asian Imagery: Cylinder Seals’ (ACAWAI-CS), which is ongoing and in an expansion phase (ROßBERGER, KURMANGALIEV, OTTO 2018; ROßBERGER, KURMANGALIEV 2023 – fig. 11). This open, flexible and dynamic database is, amongst other things, particularly suitable for possible future textual analyses on diverse corpora of cylinder seals¹.

The other area in which, over time, the experience of the logicist school has been able to prudently sprout interesting shoots has been the textual analysis of archaeological writings. It is a type of meta-archaeological research that covers topics as diverse as the schematic representation and comparison of the content of archaeological reports, history of studies, methodology and the possibility of integrating the results of past research experiences into new projects. As early as the 1970s, Jean-Claude Gardin and Marie-Salomé Lagrange proposed developing, in line with the principles of the

¹ <https://www.acawai-cs.gwi.uni-muenchen.de/>.

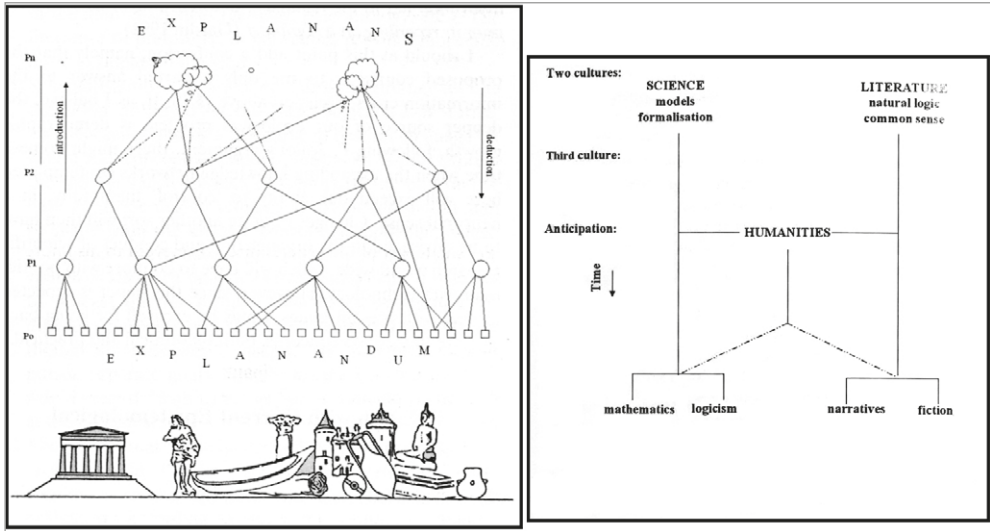


fig. 12 – The logicist view of the archaeological discourse. Left: the standard schematisation of arguments in archaeological constructs. Right: diagram on possible future evolutions of archaeological discourse (GARDIN 2003, figs. 1-2).

logicist school, the analysis and translation of the content of entire excavation reports (and similar texts belonging to the category of *construction interprétative*) into a highly formalized language (GARDIN, LAGRANGE 1975; GARDIN 1980 – fig. 12).

The importance of logic-discursive modelling would subsequently be stressed and updated by a number of scholars educated in the same cultural and academic environment, especially by Gardin and Lagrange themselves (GARDIN 1997, 2002; LAGRANGE, BONNET 1978; LAGRANGE, MONTEIRO RODRIGUES 1994; see also the synthesis by MOSCATI 1996b, 583-586). Insightful observations and a very clear and explanatory experiment were produced, for example, by Marie-Salomé Lagrange and Maria da Conceição Monteiro Rodrigues in 1994, who used the declarative programming language SNARK (LAURIÈRE 1986) to analyze the extract of a monograph by Marija Gimbutas. The starting principle is that the rigorous and correct evaluation of the logical parts of an archaeological report can produce a valid representation of its strengths and weaknesses, while the basic scheme and organization are comparable to those of an expert system. This allows for the recognition of inconsistencies and considerably improves the reliability of the proposed interpretations. Archaeological reasoning can be represented, according to the logicist view, as a chain of inferences which leads from observation in the field to the final interpretation of data.

In more recent times some archaeologists who, from the point of view of theory and methodology, are separate from the logicist tradition have also raised the issue of formalizing the content of archaeological documentation texts. Sveta Matskevitch and Ilan Sharon (MATSKEVITCH, SHARON 2018) were faced with the problem of managing

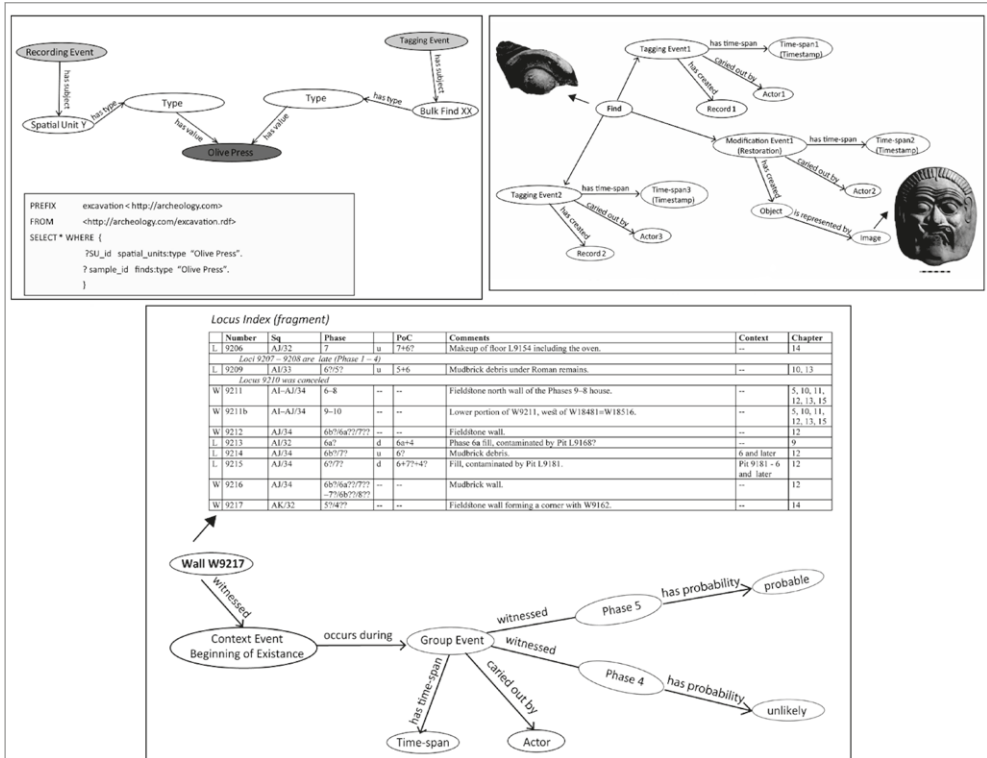


fig. 13 – Modeling for the interoperability of archaeological archives. Modeling: ambiguity, top-left; multiple interpretation of the same object, top-right; uncertainty of the stratigraphic definition of a feature, bottom (MATSKEVITCH, SHARON 2018, 50, 53-54, figs. 1.6-1.8).

and coordinating the information contained in archival records of various scholars and excavation expeditions (including explorations in underwater areas) who came to the site of Dor with different interests over the course of almost a century. Their proposal of creating meta-databases and the interoperability of very heterogeneous texts is mainly rooted in the Anglo-American tradition, as are the tools used. Matskevitch and Sharon adopted the model Simple Knowledge Organization System (SKOS)² for taxonomies and vocabularies, and the CIDOC Conceptual Reference Model (CIDOC-CRM)³ to map the data. The system they adopted allowed them rigorously to address the organization and mapping of ambiguous information, information originating from different interpretations of the same data or multiple interpretations resulting from uncertainties or doubts, from a perspective of Linked Open Data (fig. 13).

Of particular interest and originality for its meta-archaeological approach and reflections is the recent experiment by Gertjan Plets, Pim Huijnen, and David van Oe-

² <https://www.w3.org/2004/02/skos/>.

³ <https://cidoc-crm.org/>.

veren (PLETS, HUIJNEN, VAN OEVEREN 2021). In this paper, the authors aim to use textual analysis to investigate how the way of writing archaeological texts in Dutch-speaking Belgium - and the underlying heuristic practices - has changed over time. The corpus of data they examine and investigate with text mining is extensive and covers a significant time span (from 1945 to 2017). The results obtained allowed the authors to reconstruct diachronic variations in theoretical trends within the aforementioned production (which only concerns archaeological investigations conducted in Belgium).

It also permitted the identification of a qualitative decline in the production of archaeological texts, and recognition of the changes induced in the interpretation of the past by the influence of nationalism. The use of topic modelling, which identifies semantic co-occurrences between jointly used words, and, more precisely, of the Mallet package⁴ made the tracing of paths followed by the different schools of thought possible. Subsequently, the authors purposefully constructed a thesaurus of key theoretical terms in Dutch, since there were no ready-made ones, and conducted a diachronic analysis of the developments of the different schools using Voyant Tools⁵. The third stage of investigation concerned measuring the conformity of texts to 'boiler templates' provided by the bureaucracy, or the use of 'tailored methods' to draft them. In this case, the authors chose to use the 'scikit-learn'⁶ machine learning tool, which was also useful for measuring the rate of plagiarism and its diachronic dynamics. The more than satisfactory results suggest that this work may have paved the way for a very stimulating line of investigation.

6.5 CLOSING REMARKS

The history of textual analysis in archaeology is long and complex. The technological innovations and trends of recent years seem to propitiate a further renewal of the theories, methods and techniques of textual analysis available to archaeologists. On the other hand, if the requirements prevailing in recent times remain those which have emerged from the first pioneering studies, that is to systematize and harmonize the available information and explore large corpora of artifacts (especially their iconographies), we are given a glimpse of an ever-increasing popularity for meta-archaeological investigations.

The enormous diffusion of Artificial Intelligence (AI), and the ever-increasing access and ease of use of such applications, seems to be inspiring both old and new types of analysis of the scientific texts produced by archaeologists. The proposals and challenges may not always seem satisfying, but the overall situation appears promising, especially in view of the rightly cautious and disenchanted attitude that archaeologists who have experience in the use of quantitative methods show towards AI (TRUST ISSUES 2023; GATTIGLIA 2025, especially 229-231).

⁴ <https://mimno.github.io/Mallet/>.

⁵ <https://voyant-tools.org/>.

⁶ <https://scikit-learn.org/stable/>.

Lorenzo Cardarelli

7. THE LEGACY: THE DATA SCIENCE MOVEMENT

The convergence of computing and communication has produced a society that feeds on information. Yet most of the information is in its raw form: data. If data is characterized as recorded facts, then information is the set of patterns, or expectations, that underlie the data. There is a huge amount of information locked up in databases—information that is potentially important but has not yet been discovered or articulated. Our mission is to bring it forth (WITTEN *et al.* 2011, xxi).

7.1 THE LEADING ACTORS: DATA SCIENCE, MACHINE LEARNING AND DEEP LEARNING

It is now fifteen years since the contribution of Michael Baxter and Hilary Cool in issue 21 of «Archeologia e Calcolatori» (BAXTER, COOL 2010), which was characterized as innovative due to its presentation as a true computer guide. In addition to a theoretical introduction, the text included the commented R code strings necessary to reproduce the analyses and to replicate them elsewhere. This educational approach has since become a staple feature in numerous contemporary coding and data analysis textbooks (VANDERPLAS 2016; WICKHAM, GROLEMUND 2016; RASCHKA 2022).

Since 2010, the computing and computer science landscape has undergone significant transformations, driven by a series of innovations that have elevated AI to a central role. The current period has been characterized by the proliferation of large language models (LLMs), a development that has accelerated in the aftermath of the COVID-19 pandemic. Anticipating the typical dynamics of the modern approach to data analysis, Baxter and Cool use the R programming language and specifically the MASS package (VENABLES, RIPLEY 2002). The employment of this statistical programming language, which has attained widespread recognition within the computational archaeology community (CARLSON 2017), facilitates the circumvention of numerous limitations, including those pertaining to licensing (as R is a fully open-source software), as well as the ‘didactic’ and academic constraints associated with a graphical interface (GUI) program. Indeed, the ‘prompt’ approach, characterized by the sequential insertion of textual strings, finds optimal application in scientific publications and replicability. Conversely, R continues to be a preferred tool among computational archaeologists, despite the growing prominence of Python, the *lingua franca* of machine learning (ML). As a drawback, the absence of a graphic interface may deter new users, who are

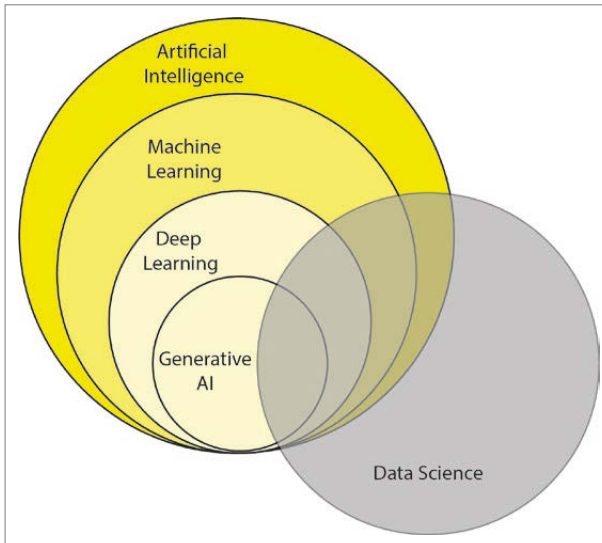


fig. 1 – Relationship between Data Science and Artificial Intelligence.

less versed in the domain of computers. This issue will be revisited in the conclusion of this contribution.

The concept and use of ‘Data Science’ can be traced back to the mid-20th century. However, it was not until the mid-2010s that this discipline began to gain widespread popularity, coinciding with the development of Big Data and the Internet of Things (CAO 2017; DONOHO 2017). The multifaceted and multidisciplinary nature of Data Science (EMMERT-STREIB, DEHEMER 2019; MIKE, HAZZAN 2023) is probably the reason why it is difficult to give a definition of it. Several authors have defined it as a point of contact and sharing between mathematics and statistics, computer science and application domain. Within the overarching framework of technical knowledge, such as statistics and computer science, there is also a requirement for specific knowledge of the application domain. Consequently, the data scientist is required to function as a transdisciplinary figure, necessitating a comprehensive understanding and appreciation of the intricacies inherent in the discipline with which they are engaged. The data scientist’s knowledge encompasses a synthesis of programming, data visualization, command-line tools, databases, statistics, and ML (DONOHO 2017).

To provide a more detailed discussion of the skills required of the new computational archaeologist, it is necessary to consider the Data Science within the surrounding and relative disciplines. While maintaining its own distinct identity, Data Science shares many tools and approaches with the world of AI. In fact, AI is a very broad field that encompasses several branches, including ML, which is currently the dominant approach in the field. The latter includes DL, which represents the most advanced frontier of ML techniques (*fig. 1*). The subject has a long history (see MCCULLOCH, PITTS 1943; ROSENBLATT 1958) and has also undergone periods of significant difficulty (the so-called *AI winter*, *fig. 2* - SCHUCHMANN 2019; TOOSI *et al.* 2021).

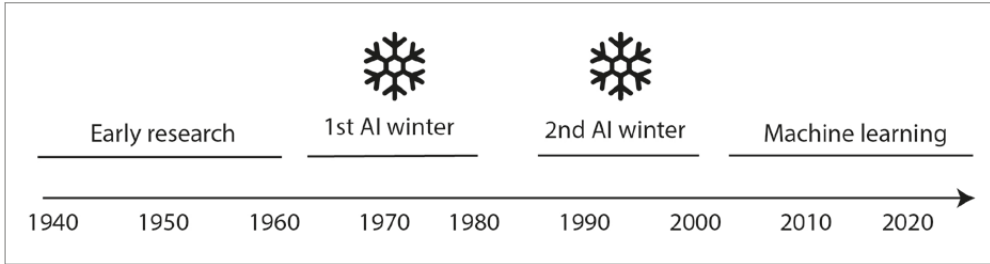


fig. 2 – AI winter chronology (based on https://www.researchgate.net/publication/372616731_A_Non-technical_Introduction_to_Machine_Learning).

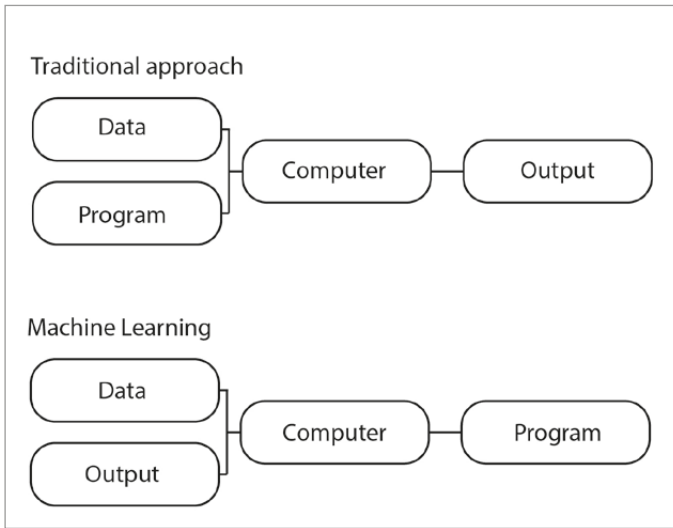


fig. 3 – Classical Programming and ML approach.

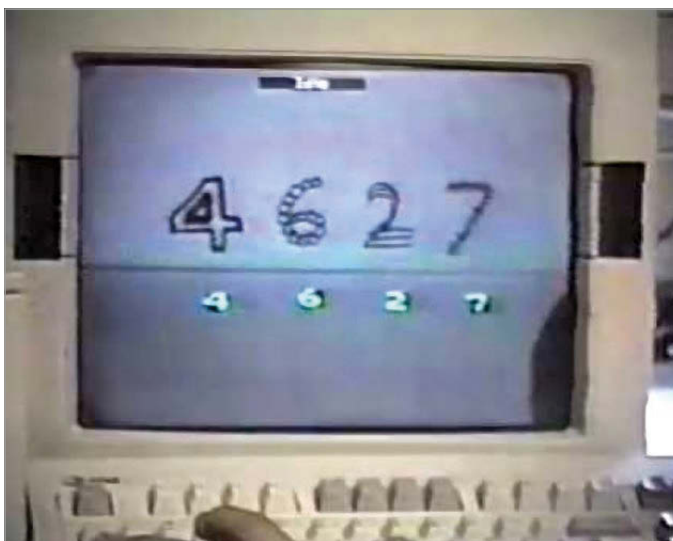


fig. 4 – Early demonstration of a convolutional network (full video: https://www.youtube.com/watch?v=FwFduRA_L6Q).

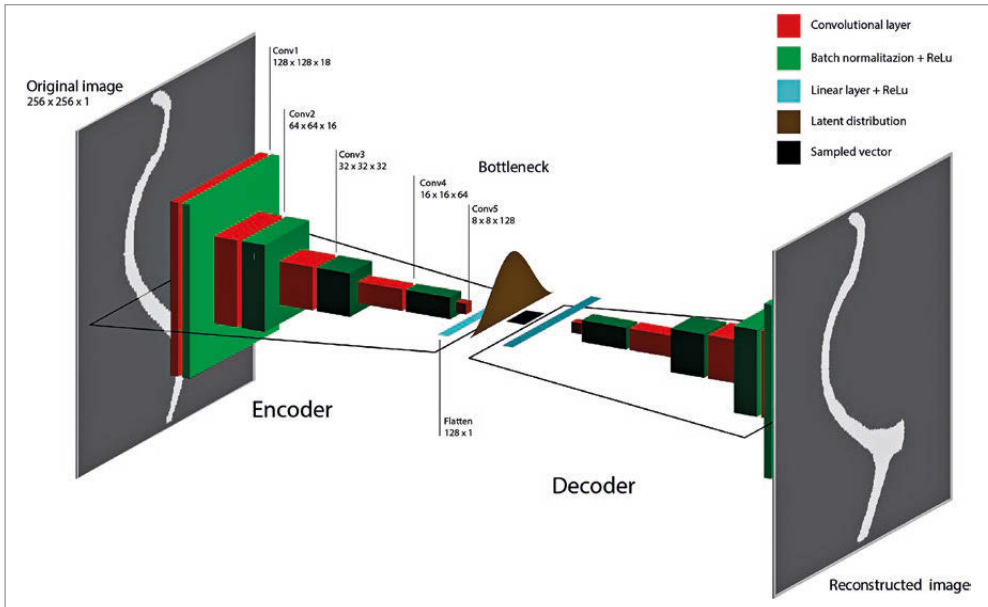


fig. 5 – An autoencoder architecture used in CARDARELLI 2022.

Since the mid-2010s, however, several positive factors have emerged that have enabled the rapid development we observe today. ML has been shown to overturn the classical programming perspective, replacing the definition of rules to obtain a result with the use of the result to obtain rules (MOHRI *et al.* 2012) (fig. 3). This inductive approach necessitates substantial data and computing capabilities to facilitate the modelling of these relationships. The increased availability of information, coupled with the advanced capabilities of contemporary personal computers, has fostered a conducive environment for the advancement of this field. A sub-branch of ML known as DL has been the most popular approach since the mid-2010s. In 2015, Yann LeCun and colleagues (LECUN *et al.* 2015) published an important paper defining the DL concept. Here we have a fragment from the abstract: «Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction».

Put simply, DL is characterized by deep neural networks, algorithms that attempt to mimic the behavior of neurons in the human brain. Convolutional networks, a type of DL algorithm, were first developed in 1989 (LECUN *et al.* 1989 - fig. 4), however, it was not until 2010 that the rapid and substantial development of DL and its applications began to take shape. In 2009, the first version of ImageNet (<https://www.image-net.org/>) was published, followed by highly advanced architectures such as AlexNet (KRIZHEVSKY *et al.* 2012), VGG (SIMONYAN, ZISSERMAN 2015) and ResNet (HE *et al.* 2015). Concurrently, the notion of generative AI is being formalized: in 2013, the Variational Autoencoder (VAE) architecture was proposed (KINGMA, WELLING

2013) (*fig. 5*); in 2014, the Generative Adversarial Network (GAN) architecture was introduced (GOODFELLOW *et al.* 2014); and a year later, diffusion models were proposed (SOHL-DICKSTEIN *et al.* 2015; DHARIWAL, NICHOL 2021). The transformer architecture was introduced in 2017 (VASWANI *et al.* 2023) and had a significant impact on text generation.

The advent of LLMs marked a significant turning point in the realm of AI and especially their knowledge and utilization by the public. The release of ChatGPT by OpenAI in late 2020 catalyzed the development that has come to fruition in the present era. The evolution of recent DL development was characterized by a primary focus on image classification until 2015. The period spanning from 2015 to 2020 can be characterized as a transitional phase, during which the spotlight shifted towards generative AI, albeit with a concurrent emphasis on image generation. The period from 2020 to the present is characterized by the sustained predominance of generative AI, accompanied by a shift towards text generation, as exemplified by LLMs, and, to a lesser extent, image generation. In the field of archaeological applications, archaeologists have endeavored to leverage these sophisticated and demanding techniques from the earliest stages (BICKLER 2021; CACCIARI, POCOBELLI 2022; CARDARELLI 2024a; GATTIGLIA 2025).

The applications are extensive, ranging from artifact recognition and classification (e.g., ANICHINI *et al.* 2021), to spatial and remote sensing applications (e.g., SAKAI *et al.* 2024). This encompasses the reconstruction of fragmented data (e.g., ALTAWHEEL *et al.* 2024; CARDARELLI 2024b), and extends to data analysis in general (e.g., CARDARELLI 2022; PARISOTTO *et al.* 2022). Significant initiatives are beginning to emerge in Italy, such as the ArchAIDE (ARCHaeological Automatic Interpretation and Documentation of cERamics) project at the University of Pisa (<https://archaide-desktop.inera.it/>). This project is funded by the European Union's Horizon 2020 programme, and a similar project, known as AUTOMATA (AUTOMated enriched digitisation of Archaeological liThics and cerAmics), is also funded by the European Union and is currently being conducted by the University of Pisa (<https://automata-eccch.eu/>). In addition, the archaeological literature is beginning to feature the first systematic reviews and studies on the subject (BELLAT *et al.* 2025).

7.2 NOT ONLY DEEP LEARNING: THE ROLE OF MANIFOLD LEARNING AND DIMENSIONALITY REDUCTION

While DL represents a powerful frontier in computational archaeology, it is crucial to recognize that not all archaeological problems require such approaches. In fact, many archaeological questions can be better addressed through other sophisticated methods. This is particularly true when dealing with multivariate archaeological data (BAXTER 2015), where understanding the relationships between variables and identifying patterns is often more important than automated classification or prediction. While DL excels at learning complex patterns from large datasets, dimensionality reduction techniques offer complementary strengths: they help archaeologists visualize and interpret their

data, often providing insights that are more directly applicable to archaeological interpretation. Dimensionality reduction can be referred as the transformation of data characterized by high dimensionality into more manageable representations of smaller dimensions (VAN DER MAATEN *et al.* 2009).

To illustrate this, let us consider a hypothetical scenario. Imagine a laboratory provides an analyst with a table containing the chemical values of certain elements of any archaeological data, specifically ceramics. Imagine that we have 100 ceramic samples analyzed for 15 different chemical elements (Fe, Cu, Pb, Zn, etc.): each sample is represented by a point in a 15-dimensional space, which is impossible to visualize or interpret directly. The question that arises is how to manage, interpret or understand this dataset. Three primary multivariate techniques have been utilized in most archaeological applications in this domain: Correspondence Analysis (CA) for qualitative data, Principal Component Analysis (PCA) (PEARSON 1901; HOTELLING 1933) for unsupervised quantitative data, and Linear Discriminant Analysis (LDA) (FISHER 1936) for supervised quantitative data. Although applicable to different types of data, the functioning of these techniques is relatively similar: the algorithm proceeds to reduce the size of the dataset through a linear re-projection of the original variables within the reduced space. In other words, this means that the main components of PCA, for example, are linear combinations of the original variables and that the algorithm ‘fails’ when it comes to describing non-linear phenomena within a dataset. However, it is important to note that ‘non-linear’ relationships are frequently present in real-world datasets.

To address this limitation, various non-linear techniques have been developed, primarily based on the concept of Manifold Learning (MEILĂ, ZHANG 2023). The underlying principle of Manifold Learning is that, despite the data residing in a high-dimensional space, the inherent structure can be generalized to a smaller dimension (*fig. 6*). To illustrate this, consider the example of facial images, which, although represented by thousands of pixels (thus in a very high-dimensional space), exhibit variations in dimensions such as the angle of head rotation, or facial expression, which can be described by a much smaller number of variables.

These few ‘real’ dimensions form what is known as the manifold, the small space in which our data actually resides. Several different manifold learning techniques have been published over time. Two of them will be discussed in this paper due to their close similarity and use: t-Distributed Stochastic Neighbour Embedding (t-SNE) (VAN DER MAATEN, HINTON 2008) and Uniform Manifold Approximation and Projection (UMAP) (MCINNES *et al.* 2017). We start with t-SNE, the introduction of which will allow us to better understand UMAP. T-SNE converts similarities between data points into joint probabilities and minimizes the divergence between them in different spaces. This translates in the ability to preserve local relationships between points, while keeping the natural clusters of the data separate. Such an approach makes t-SNE particularly effective for exploratory data analysis. UMAP can be seen as an evolution of t-SNE as it clearly improves on many aspects of it. To simplify, UMAP constructs a graph in high-dimensional space, which it then attempts to simplify in low-dimensional space,

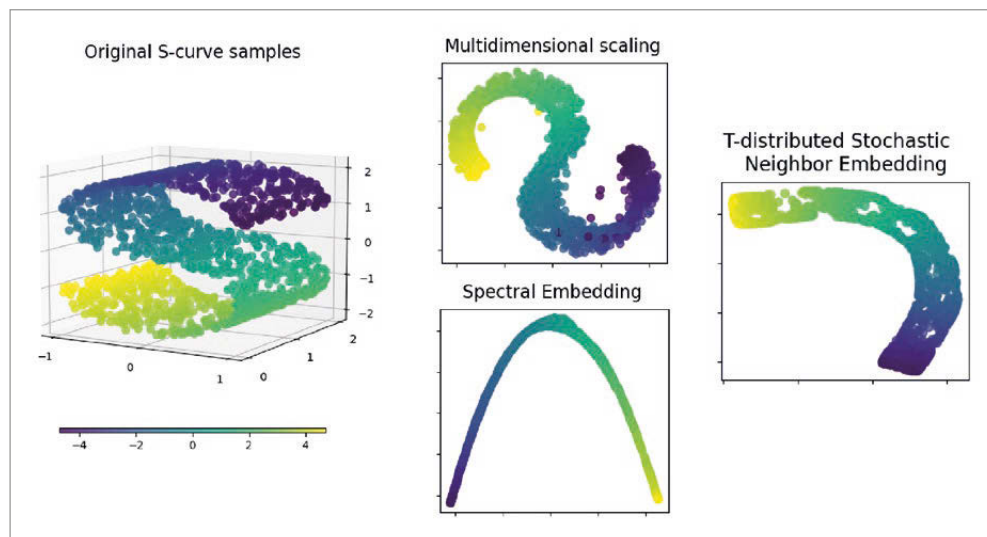


fig. 6 – A manifold learning example: the original three-dimensional shape can be approximated by 2D representations (https://scikit-learn.org/stable/auto_examples/manifold/plot_compare_methods.html).

trying to preserve as much as possible the connections established in the high-dimensional graph (McINNES *et al.* 2017).

This type of approach, together with more efficient optimization techniques, makes it possible to reduce computational complexity and better preserve both the local structure of the data (clusters) and the global structure (distances between clusters). For example, in a dataset of cell gene expression, UMAP will not only keep cells of the same type clustered together (local structure) but will also preserve the correct distances between different cell types, showing biologically related cell types closer together and less related cell types further apart (global structure). This means that the distances we see between clusters within a UMAP graph tend to be more reliable than those we see in a t-SNE graph. UMAP is used in various scientific fields, such as biology, medicine (DORRITY *et al.* 2020; ARMSTRONG *et al.* 2021; DO, CANZAR 2021) and its use in archaeology is steadily increasing (NAVARRO *et al.* 2021; SUN *et al.* 2023; CARDARELLI 2024a; CARLETTI *et al.* 2024).

UMAP is a state-of-the-art method that is also widely used in industry. For example, it is the method used by Spotify to suggest new songs to a user based on what he or she has listened to (<https://engineering.atspotify.com/2023/12/recursive-embedding-and-clustering/>), and numerous algorithms pursue this goal by improving certain characteristics (CARDARELLI 2024a, 74-75). But to return to the original example, how should the use of UMAP improve the analysis of our multivariate chemical element dataset? The difference becomes clear when we compare the results with those obtained by PCA: whereas the latter, being a linear transformation, tends to show our

Linear methods (PCA)	
Pro	Cons
Results immediately interpretable	Limited to linear relationships between variables
High computing speed	Cannot capture complex non-linear patterns
Determinism (always gives the same outcome)	May 'flatten' important structures in the data
Allows you to quantify the contribution of each component to the total variance	Reduced performance on high dimensional data
Easy to implement and replicate	Sensitive to outliers
Nonlinear methods (UMAP)	
Pro	Cons
Capture complex non-linear relationships	Less interpretability of results
Better preservation of local data structure	Non-deterministic (results can vary between different runs)
Excellent ability to detect clusters	Sensitivity to input parameters (perplexity, n_neighbours, etc.)
Greater flexibility to adapt to different data structures	Takes more computational resources than PCA
Better computational performance compared to other non-linear methods (e.g. t-SNE)	It can be difficult to choose optimal parameters

tab. 1

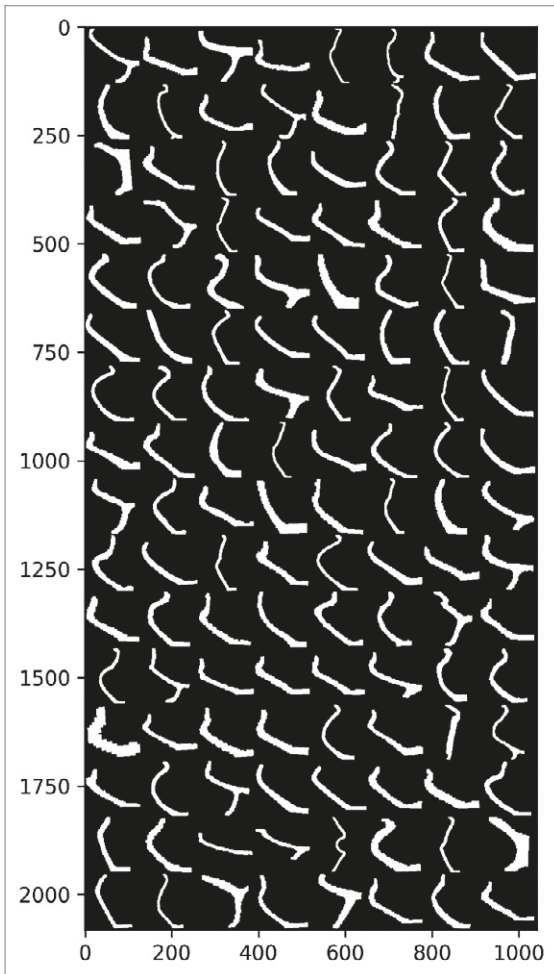


fig. 7 - A batch of binary images.

archaeometric data as a single diffuse point cloud, as non-linear relationships between chemical elements are ‘flattened’. UMAP can reveal more complex structures. In our imaginary case, well-defined clusters of artifacts with similar chemical compositions can emerge, suggesting different production ‘recipes’ or different sources of raw materials. These clusters, which may appear overlapping or indistinguishable in PCA, become clearly separated in two-dimensional space in UMAP, allowing us to identify patterns that may have significant archaeological implications. In the next paragraph we will give tangible proof of this, but before moving on to the practical example, a table is proposed (*tab. 1*) in which the pros and cons of linear methods (in this case defined by PCA) and non-linear methods (represented by UMAP) are summarized (for more on this subject, see CARDARELLI, LAPADULA 2022).

7.3 A PRACTICAL EXAMPLE: UMAP AND ARCHAEOLOGICAL CERAMICS

One of the main features of UMAP is its ability to deal effectively with both quantitative (such as isotopic and dimensional measurements) and qualitative variables. An illustrative case study is the one developed in my doctoral thesis (CARDARELLI 2024a), where the object of analysis were black and white images representing profiles of ar-



fig. 8 – Detail of the book cover dedicated to the publication of the excavations conducted at the necropolis of Osteria dell’Osa (BIETTI SESTIERI 1992).

chaeological pottery. From a technical point of view, these binary images are essentially matrices where the value 0 corresponds to the ‘absence’ of the pot and the value 1 to the ‘presence’ of the pot (*fig. 7*). The procedure requires the ‘flattening’ of these matrices for each record, generating over 65,000 variables for each element.

The analysis of the Osteria dell’Osa necropolis (BIETTI SESTIERI 1992 - *fig. 8*) provides an instructive application of non-linear dimensionality reduction on binary variables and comparison with PCA. Applying both methods to the same data set, with the markers colored according to the functional class of each record (cup, amphora, etc.), significant differences emerge:

– PCA results (*fig. 9*): Although a macro division between functional classes is visible, the overall structure appears to be overlapping and not well-defined. Without the support of color coding, identification of the groupings would be problematic.

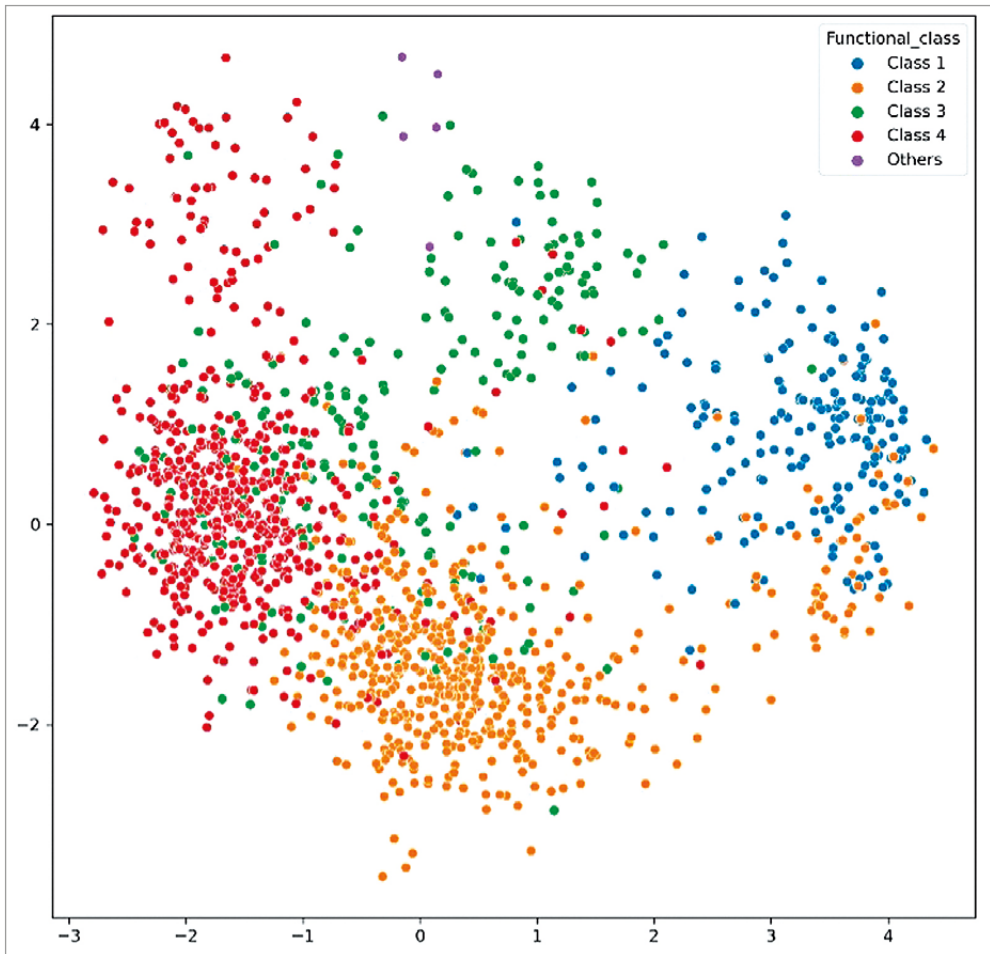


fig. 9 – PCA dimensionality reduction applied to Osteria dell’Osa necropolis.

– UMAP results (*fig. 10*): The method produces a true ‘map’ of the ceramic assemblage in which the functional classes form clearly defined clusters, while retaining some natural overlap. This clear structuring not only facilitates visual interpretation but also lends itself to subsequent clustering analyses.

Despite the high dimensionality of the original data, UMAP is able to reveal patterns and structures that stimulate archaeological interpretation, providing a more coherent and morphologically meaningful representation than traditional linear methods.

In the Data Science landscape, R and Python are emerging as the two dominant programming languages, each with its own history and strengths. R was created specifically for statistical analysis and data visualization, by statisticians for statisticians. Python, on the other hand, is a general-purpose programming language that has gained enormous popularity in Data Science due to its versatility and rich ecosystem of specialized libraries (*fig. 11*).

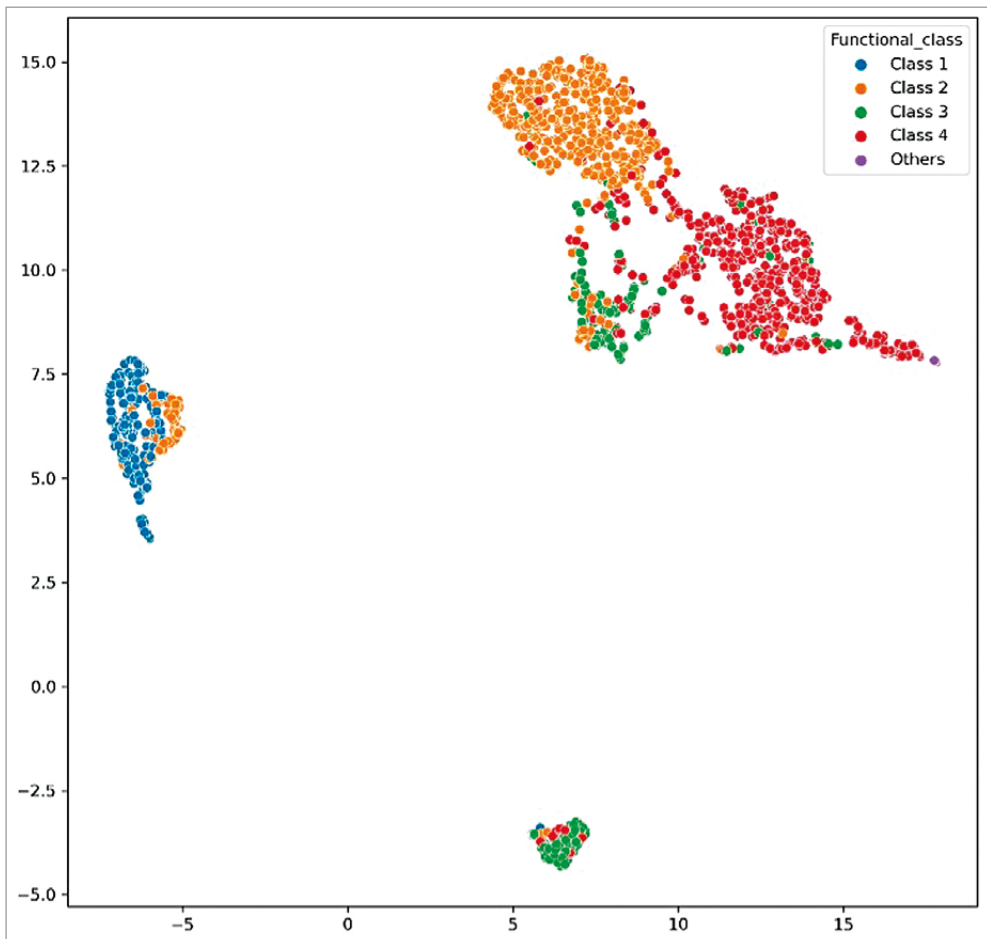


fig. 10 – UMAP dimensionality reduction applied to Osteria dell’Osa necropolis.

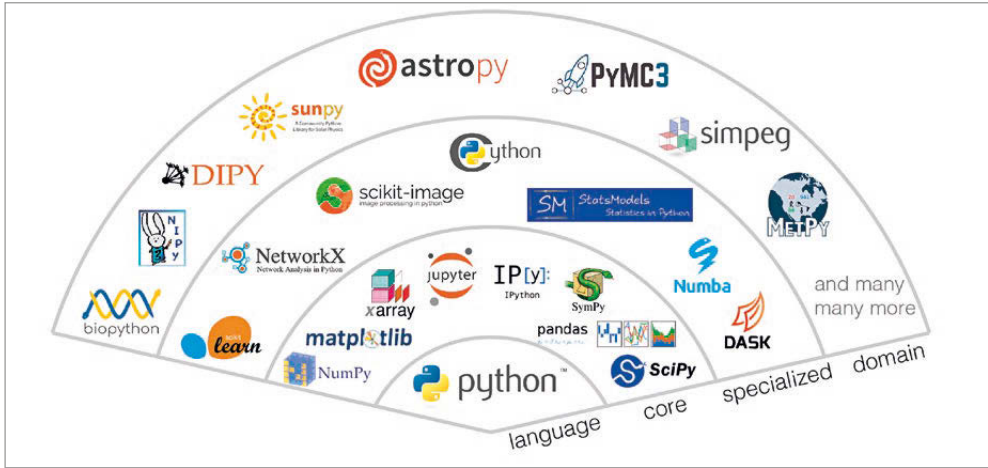


fig. 11 – Python ecosystem (<https://jupyterearth.org/jupyter-resources/introduction/ecosystem.html>).

The choice of language to use in archaeology must be based on specific considerations. Although both languages have the necessary tools to read, clean and visualize data, R retains a privileged position in the archaeological field (CARLSON 2017; <https://benmarwick.github.io/aswr/>) due to the availability of numerous domain-specific packages. For a newcomer to computational archaeology interested in basic statistical analysis and visualization of results, R is often the most accessible choice: it offers a more structured bibliography and numerous published papers that serve as practical reference examples.

The scenario changes radically when it comes to implementing ML or DL. Here, Python dominates unchallenged thanks to libraries such as TensorFlow (<https://www.tensorflow.org/>) and PyTorch (<https://pytorch.org/>), specifically designed for the creation of complex neural networks and optimized to take advantage of the hardware acceleration of GPUs. It should be noted, however, that there is currently a lack of organic resources, such as textbooks or manuals, specifically dedicated to learning Python in an archaeological context.

It is important to stress that these two languages, both open-source, do not compete with each other, but rather complement each other. An important example of this is Quarto (<https://quarto.org/>), a framework for publishing scientific material, which, although born for R, also fully supports the Python environment. The same applies to tools such as Jupyter Notebook (<https://jupyter.org/>), which allows both R and Python to be used in the same working environment, facilitating a hybrid approach. Both benefit from extremely active and vibrant communities, and it would be ideal for a computational archaeologist to become familiar with both development environments.

In conclusion, while recognizing the importance of both languages, Python is currently recommended primarily for those interested in ML, while R remains the more accessible choice for those new to computational archaeology due to its well-established integration with the archaeological field.

7.4 CONCLUSIONS AND PERSPECTIVES

The landscape of computer science and computational archaeology has evolved significantly since the pioneering work of Baxter and Cool in 2010. Today's computational archaeologist operates at the intersection of multiple disciplines, leveraging both traditional statistical methods and cutting-edge AI technologies. This paper has demonstrated how different approaches – from DL to non-linear dimensionality reduction methods like UMAP – can complement each other in archaeological analysis, each offering unique perspectives on archaeological data.

The integration of these tools presents both opportunities and challenges. While DL methods excel at pattern recognition and automated analysis, techniques like UMAP offer powerful ways to explore and visualize complex archaeological datasets, revealing structures that traditional methods might miss. The choice between Python and R programming languages reflects this duality: R maintains its strength in statistical analysis and archaeological applications, while Python dominates in ML implementations.

However, the growing sophistication of these tools raises important questions about archaeological education and training. The scarcity of computational archaeology courses in Italian universities creates a significant gap between available tools and the ability to use them effectively. While Large Language Models offer an appealing shortcut to code generation, they should be viewed as assistive tools rather than substitutes for fundamental understanding. The future of computational archaeology depends not just on adopting new technologies, but on developing comprehensive educational frameworks that combine:

- 1) Strong theoretical foundations in both archaeology and Data Science;
- 2) Practical programming skills in both R and Python;
- 3) Critical understanding of different analytical approaches and their appropriate applications;
- 4) Awareness of both the potential and limitations of AI tools in archaeological research.

Looking forward, the field of computational archaeology stands at a crucial juncture. The challenge lies not just in adopting new technologies, but in integrating them thoughtfully into archaeological practice while maintaining rigorous methodological standards. This requires developing new educational paradigms, fostering collaboration between archaeologists and computer scientists, and establishing best practices for the application of these tools in archaeological research.

The future success of computational archaeology will depend on our ability to bridge the gap between traditional archaeological expertise and modern computational methods. This means not only teaching technical skills but also cultivating critical thinking about how these tools can best serve archaeological inquiry.

8. REFERENCES

- ABI-BOUTROS B., BELLIER L. 1977, *Contribution à la taxinomie des micromammifères. Application au genre Crocidura*, «Les Cahiers de l'analyse des données», 2, 4, 435-450 (https://www.numdam.org/item/CAD_1977__2_4_435_0/).
- AITCHISON J. 1986, *The Statistical Analysis of Compositional Data*, London-New York, Chapman and Hall (<https://doi.org/10.1002/bimj.4710300705>).
- ALBERTI G. 2013, *An R Script to Facilitate Correspondence Analysis. A Guide to the Use and the Interpretation of Results from an Archaeological Perspective*, «Archeologia e Calcolatori», 24, 25-53 (<https://www.archcalc.cnr.it/journal/articles/661>).
- ALTAWHEEL M., KHELIFI A., ZAFAR M.H. 2024, *Using Generative AI for Reconstructing Cultural Artifacts: Examples Using Roman Coins*, «Journal of Computer Applications in Archaeology», 7, 1, 301-315 (<https://doi.org/10.5334/jcaa.146>).
- AMMERMAN A.J., CAVALLI-SFORZA L.L. 1984, *The Neolithic Transition and the Genetics of Populations in Europe*, Princeton, Princeton University Press.
- ANDRÉ J. 2004, *Préhistoire de l'informatique à l'Université de Rennes. Des origines au Général de Gaulle*, in J. ANDRÉ, P. MOUNIER-KUHN (eds.), *Actes du 7^e Colloque sur l'histoire de l'informatique et des transmissions (Cesson-Sévigné 2004)*, Rennes, INRIA-Université de Rennes, 228-236.
- ANDRÉ-SALVINI B. 2003, *Oppert, Jules (Julius)*, in *Reallexikon der Assyriologie*, 10, 1-2, Berlin-New York, De Gruyter, 117-118.
- ANDRÉ-SALVINI B. 2013, *Talbot, William Henry Fox*, in *Reallexikon der Assyriologie*, 13, Berlin-New York, De Gruyter, 419-420.
- ANICHINI F., DERSHOWITZ N., DUBBINI N., GATTIGLIA G., ITKIN B., WOLF L. 2021, *The Automatic Recognition of Ceramics from Only One Photo: The ArchAIDE App*, «Journal of Archaeological Science: Reports», 36, 102788, 301-315 (<https://doi.org/10.1016/j.jasrep.2020.102788>).
- APELLANIZ J.M., GOMEZ F.C. 1999, *La forma del arte paleolítico*, Bilbao, Universidad de Deusto.
- ARIAS P.E. 1978, *Problemi passati e attuali della ceramica greca. L'elaborazione elettronica dei dati emergenti dall'analisi dei vasi greci figurati*, Quaderni della Ricerca Scientifica 2, Roma, CNR, 201-232.
- ARIAS P.E., DI BARI V.C., ORSOLINI RONZITTI G. 1985, *La ceramica attica a figure nere e rosse del Corpus Vasorum Antiquorum. L'analisi computerizzata dei dati*, Contributi del Centro Linceo Interdisciplinare di Scienze Matematiche e loro Applicazioni 70, Roma, Accademia Nazionale dei Lincei.
- ARMSTRONG G., MARTINO C., RAHMAN G., GONZALEZ A., VÁZQUEZ-BAEZA Y., MISHNE G., KNIGHT R. 2021, *Uniform Manifold Approximation and Projection (UMAP) Reveals Composite Patterns and Resolves Visualization Artifacts in Microbiome Data*, «mSystems», 6, 1-6 (<https://doi.org/10.1128/mSystems.00691-21>).
- AUDOUZE F. 1984, *Échanges d'objets et évolution stylistique. Un exemple: les épingles à tête vasiforme de Savoie*, in *Éléments de pré- et protohistoire européenne. Hommages à J.P. Millot*, Paris, Les Belles-Lettres, 37-44.
- AZOURY I., HODSON F.R. 1973, *Comparing Palaeolithic Assemblages: Ksar-Akil, a Case Study*, «World Archaeology», 4, 3, 292-306 (<https://doi.org/10.1080/00438243.1973.9979540>).
- BAXTER M.J. 1989, *Multivariate Analysis of Data on Glass Compositions: A Methodological Note*, «Archaeometry», 31, 45-53 (<https://doi.org/10.1111/j.1475-4754.1989.tb01055.x>).
- BAXTER M.J. 1994, *Exploratory Multivariate Analysis in Archaeology*, Edinburgh, Edinburgh University Press.
- BAXTER M.J. 2015, *Exploratory Multivariate Analysis in Archaeology*, New York, Percheron Press.
- BAXTER M.J., COOL H.E.M. 2010, *Correspondence Analysis in R for Archaeologists: An Educational Account*, «Archeologia e Calcolatori», 21, 211-228 (<https://www.archcalc.cnr.it/journal/articles/573>).

- BAXTER M.J., HEYWORTH M.P. 1989, *Principal Components Analysis of Compositional Data in Archaeology*, in RAHTZ, RICHARDS 1989, 226-240.
- BEAUDOUIN V. 2016, *Statistical Analysis of Textual Data: Benzécri and the French School of Data Analysis*, «Glottometrics», 33, 56-72.
- BEH E.J., LOMBARDO R. 2012, *A Genealogy of Correspondence Analysis*, «Australian & New Zealand Journal of Statistics», 54, 2, 137-168 (<https://onlinelibrary.wiley.com/doi/10.1111/j.1467-842X.2012.00676.x>).
- BELLAT M., FIGUEROA J.D.O., REEVES J.S., TAGHIZADEH-MEHRJARDI R., TENNIE C., SCHOLTEN T. 2025, *Machine Learning Applications in Archaeological Practices: A Review*, 1-43 (<https://doi.org/10.48550/arXiv.2501.03840>).
- BENZÉCRI J.P. 1966, *Linguistique et mathématique*, «Revue Philosophique de la France et de l'Étranger», 156, 309-374.
- BENZÉCRI J.-P. 1969, *Statistical Analysis as a Tool to Make Patterns Emerge from Data*, in S. WATANABE (ed.), *Methodologies of Pattern Recognition*, New York, Academic Press, 35-74.
- BENZÉCRI J.P. 1973, *L'analyse des données. Tome 1: La taxonomie; Tome 2: L'analyse des correspondances*, Paris, Dunod.
- BENZÉCRI J.P. 1976a, *Histoire et préhistoire de l'analyse des données. Partie I La préhistoire*, «Les Cahiers de l'analyse des données», 1, 1, 9-32 (https://www.numdam.org/item/CAD_1976__1_1_9_0/).
- BENZÉCRI J.P. 1976b, *Histoire et préhistoire de l'analyse des données. Partie II La biométrie*, «Les Cahiers de l'analyse des données», 1, 2, 101-120 (https://www.numdam.org/item/CAD_1976__1_2_101_0/).
- BENZÉCRI J.P. 1976c, *Histoire et préhistoire de l'analyse des données. Partie III Era piscatoria*, «Les Cahiers de l'analyse des données», 1, 3, 221-241 (https://www.numdam.org/item/CAD_1976__1_3_221_0/).
- BENZÉCRI J.P. 1976d, *Histoire et préhistoire de l'analyse. Partie IV La psychométrie*, «Les Cahiers de l'analyse des données», 1, 4, 343-366 (https://www.numdam.org/item/CAD_1976__1_4_343_0/).
- BENZÉCRI J.P. 1977a, *Histoire et préhistoire de l'analyse. Partie V L'analyse des correspondances*, «Les Cahiers de l'analyse des données», 2, 1, 9-40 (https://www.numdam.org/item/CAD_1977__2_1_9_0/).
- BENZÉCRI J.P. 1977b, *Actualité: l'analyse des données en archéologie préhistorique*, «Les Cahiers de l'analyse des données», 2, 2, 193-214 (https://www.numdam.org/item/CAD_1977__2_2_193_0/).
- BENZÉCRI J.P. 1977c, *Typologie de l'outillage préhistorique en pierre taillée. Application à la définition du type burin de Noailles*, «Les Cahiers de l'analyse des données», 2, 2, 215-238 (https://www.numdam.org/item/CAD_1977__2_2_215_0/).
- BENZÉCRI J.P. 1982, *Histoire et préhistoire de l'analyse des données*, Paris, Dunod.
- BENZÉCRI J.P. 1992, *Correspondence Analysis Handbook*, Statistics, Textbooks and Monographs 125, New-York-Basel-Hong Kong, Dekker.
- BENZÉCRI J.P. et al. 1981, *Pratique de l'analyse des données. Linguistique et lexicologie*, Paris, Dunod.
- BERGMANN J. 1970, *Die ältere Bronzezeit Nordwestdeutschlands*, Marburg, Elwert.
- BERGOUGNAN D., MOHEN J.P. 1980, *L'histoire du camp néolithique de Semussac*, «Les dossiers de l'archéologie», 42, 75-79.
- BERTHOUD T. 1979, *Étude par l'analyse des traces et la modélisation de la filiation entre minerai de cuivre et objets archéologiques du Proche-Orient*, Thèse de doctorat, Physique, Université de Paris VI.
- BERTIER P., BOUROCHE J.M. 1981, *Analyse des données multidimensionnelles*, Paris, PUF.
- BICKLER S.H. 2021, *Machine Learning Arrives in Archaeology*, «Advances in Archaeological Practice», 9, 186-191 (<https://doi.org/10.1017/aap.2021.6>).
- BIETTI A. 1982, *Tecniche matematiche nell'analisi dei dati archeologici*, Contributi del Centro Linceo Interdisciplinare di Scienze Matematiche e loro Applicazioni 61, Roma, Accademia Nazionale dei Lincei.
- BIETTI A. 1993, *Nuove prospettive nelle analisi di correlazioni spaziali in preistoria*, «Archeologia e Calcolatori», 4, 39-60.
- BIETTI A., BURANI A., ZANELLO L. 1985, *Interactive Pattern Recognitions on Prehistoric Archaeology. Some Applications*, in VOORRIPS, LOVING 1985, 191-203.
- BIETTI A., RAMBALDI A., ZANELLO L. 1978, *DENDR. Un programma Fortran di "cluster analysis" per applicazioni archeologiche, paleontologiche e paleoecologiche*, «Quaternaria», 20, 49-85.
- BIETTI SESTIERI A.M. (ed.) 1992, *La necropoli laziale di Osteria dell'Osa*, Roma, Quasar.

- BINFORD L.R., BINFORD S.R. 1966, *A Preliminary Analysis of Functional Variability in the Mousterian of Levallois Facies*, «American Anthropologist», 68, 238-295 (<https://doi.org/10.1525/aa.1966.68.2.02a001030>).
- BIRKS H.J.B., GORDON A.D. 1985, *Numerical Methods in Quaternary Pollen Analysis*, New York, Academic Press.
- BOARDMAN J., KURTZ D. (eds.) 1993, *International Conference on Data and Image Processing in Classical Archaeology (Ravello 1992)*, «Archeologia e Calcolatori», 4 (<https://www.archcalc.cnr.it/journal/volumes/4>).
- BØLVIKEN E.E., HELSKOG K., HOLM-OLSEN I., SOLHEIM L., BERTELSEN R. 1982, *Correspondence Analysis: An Alternative to Principal Components*, «World Archaeology», 14, 41-60 (<https://doi.org/10.1080/00438243.1982.9979848>).
- BOOKSTEIN F.L. 1997, *Two Shape Metrics for Biomedical Outline Data: Bending Energy, Procrustes Distance, and the Biometrical Modeling of Shape Phenomena*, in *Proceedings of 1997 International Conference on Shape Modeling and Applications*, IEEE Computer Society Press, 110-120 (<https://doi.org/10.1109/SMA.1997.634888>).
- BORGERS B., GASSNER V., BECHTOLD B., HASENZAGL C. 2024, *The FACEM Online Database: State of Research (2011-2020) and New Perspectives*, «Archeologica Data», 4, 2, 13-24 (<https://doi.org/10.13131/UNIPI/PN6W-5A77>).
- BORILLO M. (ed.) 1978, *Archéologie et calcul*, Paris, Editions 10/18, UGE.
- BORILLO M., GARDIN J.-C. (eds.) 1974, *Banques de données archéologiques. Actes du Colloque (Marseille 1972)*, Paris, Éditions du CNRS.
- BOSSELIN B., DJINDJIAN F. 1988, *Un essai de structuration du Magdalénien français à partir de l'outillage lithique*, «Bulletin de la Société préhistorique française», 85, 10-12, 304-331 (<https://doi.org/10.3406/bspf.1988.9862>).
- BOSSELIN B., DJINDJIAN F. 1997, *Une révision du Solutréen de Laugerie-Haute et le problème des transitions Gravettien-Solutréen et Solutréen-Badegoulien en Aquitaine*, «Bulletin de la Société préhistorique française», 94, 4, 443-454 (<https://doi.org/10.3406/bspf.1997.10719>).
- BOSSELIN B., DJINDJIAN F. 1999, *Une révision de la séquence de la Riera (Asturies) et la question du Badegoulien cantabrique*, «Bulletin de la Société préhistorique française», 96, 2, 153-176 (<https://doi.org/10.3406/bspf.1997.10719>).
- BOSSELIN B., DJINDJIAN F. 2002, *Un essai de reconstitution du climat entre 40 000 BP et 10 000 BP à partir des séquences polliniques de tourbières et de carottes océaniques et glaciaires à haute résolution*, in DJINDJIAN, MOSCATI 2002, 275-300 (<https://www.archcalc.cnr.it/journal/articles/351>).
- BOUCHET J.P. 1986, *Les gisements du Paléolithique moyen de Biache-Saint-Vaast et Seclin: organisation spatiale des vestiges archéologiques*, Diplôme EPHE, 3^e section.
- BOURDIEU P., DE SAINT-MARTIN M. 1978, *Le patronat*, «Actes de la recherche en sciences sociales», 20-21, 3-8 (<https://doi.org/10.3406/arss.1978.2592>).
- BOURILLON R., FRITZ C., SAUVET G. 2012, *La thématique féminine au cours du Paléolithique supérieur européen: permanences et variations formelles*, «Bulletin de la Société préhistorique française», 109, 1, 85-103 (<https://doi.org/10.3406/bspf.2012.14144>).
- BOUROCHE J.M., SAPORTA G. 2006, *L'analyse des données*, Paris, PUF (9th ed.).
- BRAINERD G.W. 1951, *The Place of Chronological Ordering in Archaeological Analysis*, «American Antiquity», 15, 293-301 (https://doi.org/10.1007/978-1-4615-5911-5_24).
- BRUGHMANS T., PECCI A. 2020, *An Inconvenient Truth: Evaluating the Impact of Amphora Reuse through Computational Simulation Modelling*, in C.N. DUCKWORTH, A. WILSON, *Recycling and Reuse in the Roman Economy*, Oxford, Oxford University Press, 191-234.
- BRUNN H., KÖRTE G. 1870-1916, *I rilievi delle urne etrusche*, I-III, Roma-Berlino (<https://doi.org/10.11588/diglit.4976>).
- BUCK C.E., CAVANAGH W.G., LITTON C.D. 1996, *Bayesian Approach to Interpreting Archaeological Data*, Chichester, John Wiley & Sons.
- BUISSON D., FRITZ C., KANDEL D., PINÇON G., SAUVET G., TOSSELLO G. 1996, *Les contours découpés de têtes de chevaux et leur contribution à la connaissance du Magdalénien moyen*, «Antiquités Nationales», 6, 99-128.

- BURDUKIEWICZ J.M. 1981, *Gresswellian and Hamburgian*, in J.K. KOZŁOWSKI, S.K. KOZŁOWSKI (eds.), *Préhistoire de la grande plaine de l'Europe*, Archaeologia Interregionals 1, Cracow-Warsaw, 43-56.
- BURSICH D., PACE A. 2017, *Ripensando il "metodo Beazley". Ceramica attica e fotomodellazione 3D: il caso del Painter of Syracuse 19861*, «Archeologia e Calcolatori», 28.1, 73-91 (<https://doi.org/10.19282/AC.28.1.2017.05>).
- BURT C. 1950, *The Factorial Analysis of Qualitative Data*, «British Journal of Statistical Psychology», 3, 3, 166-185 (<https://doi.org/10.1111/j.2044-8317.1950.tb00296.x>).
- BURTON V., BONIN A., LOURIE J., SPIELMAN T. 1970, *The Computer and Archaeology*, «American Journal of Archaeology», 74, 3, 221-223 (<https://doi.org/10.2307/503095>).
- CACCIARI I., POCOBELLI G.F. 2022, *Machine Learning: A Novel Tool for Archaeology*, in S. D'AMICO, V. VENUTI (eds.), *Handbook of Cultural Heritage Analysis*, Cham, Springer International Publishing, 961-1002 (https://doi.org/10.1007/978-3-030-60016-7_33).
- CAHEN D., MARTIN P. 1972, *Classification formelle automatique et industries lithiques. Interprétation des hachereaux de la Kamoia*, Musée Royal de l'Afrique Centrale, Annales - Sciences Humaines 76, Tervuren.
- CAILLEZ F., PAGÈS J.P. 1976, *Introduction à l'analyse des données*, Paris, SMASH.
- CALLOW P., WEBB R.E. 1981, *The Application of Multivariate Statistical Techniques to Middle Palaeolithic Assemblages from Southwestern France*, «Revue d'Archéométrie», 5, 330-338.
- CAMIZ S. 1994, *Metodi di codifica ed analisi dei dati per lo studio di sigilli sumerici*, in *Atti della XXXVII Riunione Scientifica della Società Italiana di Statistica*, Roma, CISU, vol. 2, 105-112.
- CAMIZ S. 2004, *On the Coding of Archaeological Finds*, in MOSCATI 2004a, 201-218 (<https://www.archcalc.cnr.it/journal/articles/380>).
- CAMIZ S., DI LUDOVICO A., ROVA E. 2017, *Quantitative Approach to Ancient Near Eastern Glyptic*, Virtual Museum of Archaeological Computing (<https://archaeologicalcomputing.cnr.it/itineraries/projects/quantitative-approaches-to-ancient-near-eastern-glyptic-art/>).
- CAMIZ S., FERRAZZA E. 2006, *Studio sull'iconografia di Aiace Telamonia con metodi di analisi esplorative dei dati*, «Archeologia e Calcolatori», 17, 45-70 (<https://www.archcalc.cnr.it/journal/articles/419>).
- CAMIZ S., ROVA E. 1991, *Analysis of Ancient Near-Eastern Cylinder Seals (Late Fourth Millennium B.C.)*, in E. DIDAY, Y. LECHEVALIER (eds.), *Symbolic-Numeric Data Analysis and Learning*, New York, Nova Science, 131-142.
- CAMIZ S., ROVA E. 1992, *Analyse des sceaux cylindriques du Proche-Orient de la période de Uruk/Jemdet Nasr*, in M. BÉCUE, L. LEBART, N. RAJADELL (eds.), *Jornades Internacionales d'Anàlisi de Dades Textuals 1990*, Barcelona, Universitat Politècnica de Catalunya, 213-216.
- CAMIZ S., ROVA E. 1993a, *L'analyse des segments répétés pour l'étude des sceaux du Proche-Orient*, in *JADT. Secondes Journées Internationales d'Analyse des Données Textuelles (Montpellier 1993)*, Montpellier, ENST-Telecom, 437-445.
- CAMIZ S., ROVA E. 1993b, *Multivariate Analysis Techniques Applied to Ancient Near Eastern Seals of the Late Fourth Millennium B.C.*, in J. PAVÚK (ed.), *Actes du XII^e Congrès International des Sciences Préhistoriques et Protéohistoriques (Bratislava 1991)*, Bratislava, Archeologický ústav Slovenskej akadémie vied, 133-138.
- CAMIZ S., ROVA E. 1996, *Metodi di analisi per lo studio di un gruppo di sigilli cilindrici vicino-orientali e di altre immagini strutturate*, «Archeologia e Calcolatori», 7, 647-659 (<https://www.archcalc.cnr.it/journal/articles/172>).
- CAMIZ S., ROVA E. 2001, *Exploratory Analyses of Structured Images: A Test on Different Coding Procedures and Analysis Methods*, «Archeologia e Calcolatori», 12, 7-45 (<https://www.archcalc.cnr.it/journal/articles/313>).
- CAMIZ S., ROVA E. 2003a, *Quantitative Study of Images in Archaeology: I. Textual Coding*, in M. SCHADER, W. GAUL, M. VICHI (eds.), *Between Data Science and Applied Data Analysis*, Berlin, Springer, 624-632.
- CAMIZ S., ROVA E. 2003b, *Textual Data Analysis for the Study of Images in Archaeology*, in *Proceedings of the 14th Table ronde Informatique et Égyptologie (Pisa, Italy, 2002)* (CD-ROM only).
- CAMIZ S., ROVA E., TULLI V. 1998, *Exploratory Analysis of Images Engraved on Ancient Near-Eastern Seals Based on a Distance among Strings*, «Statistica», 58, 670-689 (<https://doi.org/10.6092/issn.1973-2201/1114>).

- CAMIZ S., ROVA E., TULLI V. 2003, *Quantitative Study of Images in Archaeology: II. Symbolic Coding*, in M. SCHADER, W. GAUL, M. VICHI (eds.), *Between Data Science and Applied Data Analysis*, Berlin, Springer, 633-641.
- CAMIZ S., ROVA E., TULLI V. 2010, *Modelling the Archaeologist's Thinking for the Automatic Classification of Uruk/Jamdat Nasr Seals Images*, in F. NICCOLUCCI, S. HERMON (eds.), *Beyond the Artifact. Digital Interpretations of the Past. Proceedings of CAA2004 (Prato 2004)*, Budapest, Archaeolingua, 429-432.
- CANAL E., CAVAZZONI S. 1990, *Antichi insediamenti antropici nella laguna di Venezia: analisi multivariata di tipo 'Fuzzy c-means clustering'*, «Archeologia e Calcolatori», 1, 165-177 (<https://www.archcalc.cnr.it/journal/articles/10>).
- CAO L. 2017, *Data Science: A Comprehensive Overview*, «ACM Computing Surveys», 50, 3, 1-42 (<https://doi.org/10.1145/3076253>).
- CARVALE A. (ed.) 2021, *Archaeological Computing: Selected Papers from the 2020 IMEKO TC-4 METROARCHAEO International Conference*, «Archeologia e Calcolatori», 32.2 (<https://www.archcalc.cnr.it/journal/volumes/32/2>).
- CARVALE A. 2022, *Le banche dati archeologiche. Articolazione e formalizzazione delle conoscenze*, 30 anni di informatica archeologica 2, Firenze, All'Insegna del Giglio (<https://doi.org/10.19282/FA.6.2022>).
- CARVALE A., DURAN-SILVA N., GRIMAU B., MOSCATI P., RONDELLI B. 2023, *Developing a Digital Archaeology Classification System Using Natural Language Processing and Machine Learning Techniques*, «Archeologia e Calcolatori», 34.2, 9-32 (<https://doi.org/10.19282/ac.34.2.2023.01>).
- CARDARELLI L. 2022, *A Deep Variational Convolutional Autoencoder for Unsupervised Features Extraction of Ceramic Profiles. A Case Study from Central Italy*, «Journal of Archaeological Science», 144, 105640, 1-12 (<https://doi.org/10.1016/j.jas.2022.105640>).
- CARDARELLI L. 2024a, *Morphological Variability and Standardisation of Vessel Shapes in the 2nd and First Half of the First Millennium BC in Continental Italy*, Bari, Edipuglia (<https://doi.org/10.4475/0840>).
- CARDARELLI L. 2024b, *From Fragments to Digital Wholeness: An AI Generative Approach to Reconstructing Archaeological Vessels*, «Journal of Cultural Heritage», 70, 250-258 (<https://doi.org/10.1016/j.culher.2024.09.012>).
- CARDARELLI L., LAPADULA A. 2022, *Dimensionality Reduction for Data Visualization and Exploratory Analysis of Ceramic Assemblages*, «Archeologia e Calcolatori», 33, 33-52 (<https://doi.org/10.19282/ac.33.2.2022.03>).
- CARLETTI E., CONATI BARBARO C., ALLEGRETTA I., CARDARELLI L., FORTI L., MONNO A., TERZANO R., ERAMO G. 2025, *Chert Sources and Territorial Behaviour after the Neolithization Process: An Exploratory Analysis from Grotta Battifratta (Rieti, Central Italy)*, «Quaternary International», 717, 109633, 1-22 (<https://doi.org/10.1016/j.quaint.2024.109633>).
- CARLSON D.L. 2017, *Quantitative Methods in Archaeology Using R*, Cambridge, University Press.
- CAUVIN M.C., GOURGAUD A., GRATUZE B., ARNAUD N., POUPEAU G., POIDEVIN J.L., CHATAIGNER C. 1998, *L'obsidienne au Proche et Moyen-Orient, du volcan à l'outil*, BAR International Series 738, Oxford, Archaeopress.
- CAVALLI-SFORZA L. 2000, *Genes, People and Languages*, San Francisco, North Point Press.
- CAVALLI-SFORZA L., CAVALLI-SFORZA F. 1993, *Chi siamo. La storia della diversità umana*, Milano, Mondadori.
- CHALINE J., LAURIN B. 1984, *Le rôle du climat dans l'évolution graduelle de la lignée *Mimomys occitanus-ostromosensis* (arvicolidae, rodentia) au pliocène supérieur*, «Geobios», 8, 323-331 ([https://doi.org/10.1016/S0016-6995\(84\)80187-4](https://doi.org/10.1016/S0016-6995(84)80187-4)).
- CHENHALL R.G. 1971, *The Archaeological Data Bank: A Progress Report*, «Computers and the Humanities», 5, 159-169 (<https://doi.org/10.1007/BF02402182>).
- CHENHALL R.G. 1978, *Nomenclature for Museum Cataloging: A System for Classifying Man-made Objects*, Nashville, Tenn., American Association for State and Local History.
- CHRISTOPHE J., GARDIN J.-C. 1956, *Projet de Code pour l'analyse des formes de poteries*, Paris, CADA.
- CIBOIS P. 1983, *L'analyse factorielle*, Paris, PUF.
- CIBOIS P. 1996, *Une méthode post-factorielle: Tri-deux. Application à l'évolution de la couleur des yeux de 1810 à 1940*, «Histoire & Mesure», 1, 1, 19-34.

- CIÉZAR P. 1990, *Sériation de la nécropole wisigothique de Duratón (Ségovie, Espagne)*, «Histoire & Mesure», 5, 1-2, 107-144.
- CITERNE P. 2003, *Les poissons dans l'art paléolithique. Un thème figuratif au cœur de l'iconographie magdalénienne*, Thèse de doctorat en Préhistoire, Université de Toulouse-Le Mirail.
- CLARKE D., RIVETT B.H.P. 1978, *A Structural Mapping Approach to Complex Decision-Making*, «The Journal of the Operational Research Society», 29, 2, 113-128 (<https://doi.org/10.1057/jors.1978.26>).
- COSTA E., BELTRAME C. 2021, *3D Modelling from Archive and Legacy Data: Preliminary Data Processing on the Roman Shipwreck Grado I*, in CARVALE 2021, 156-166 (<https://doi.org/10.19282/ac.32.2.2021.14>).
- COURBIN P. 1983, *Classement informatisé des skyphoi protocorinthiens*, «Bulletin de Correspondance Hellénique», 107, 1, 85-110 (<https://doi.org/10.3406/bch.1983.1875>).
- COWGILL G. 1994, *Unknown Sampling Bias Is Not a License to Ignore Statistical Theory*, in JOHNSON 1994, 7-11.
- COWGILL G. 2002, *Getting Bayesian Ideas across to a Wide Audience*, in DJINDJIAN, MOSCATI 2002, 191-196 (<https://www.archcalc.cnr.it/journal/articles/344>).
- CUOMO DI CAPRIO N. 1985, *La ceramica in archeologia. Antiche tecniche di lavorazione e moderni metodi d'indagine*, Roma, L'Erma di Bretschneider.
- CUOMO DI CAPRIO N., MAINONI I., SACCHI F., SPINOLO G.M. 1976, *Il metodo della termoluminescenza nella datazione ed autenticazione di reperti fittili archeologici*, Roma, CNR.
- D'ANDRIA F. (ed.) 1987, *Informatica e archeologia classica. Atti del Convegno Internazionale (Lecce 1986)*, Galatina, Congedo.
- DECORMEILLE A., HINOUT J. 1982, *Mise en évidence des différentes cultures mésolithiques dans le Bassin Parisien par l'analyse des données*, «Bulletin de la Société préhistorique française», 79, 3, 81-88 (<https://doi.org/10.3406/bspf.1982.5366>).
- DE LEEUW J. 1983, *On the Prehistory of Correspondence Analysis*, «Statistica Neerlandica», 37, 4, 161-164 (<https://doi.org/10.1111/j.1467-9574.1983.tb00810.x>).
- DE LEEUW J., MAIR P., GROENEN P. 2016, *Gifi Analysis of Multivariate Data*, Unpublished book draft.
- DELPORTE H., MAZIÈRE G., DJINDJIAN F. 1977, *L'Aurignacien de La Ferrassie: observations préliminaires à la suite de fouilles récentes*, «Bulletin de la Société préhistorique française», 74, 1, 343-361 (<https://doi.org/10.3406/bspf.1977.8457>).
- DENYS C. 1985, *Palaeoenvironmental and Palaeobiogeographical Significance of the Fossil Rodent Assemblages of Lactoli (Pliocene, Tanzania)*, «Palaeo», 52, 77-97 ([https://doi.org/10.1016/0031-0182\(85\)90032-X](https://doi.org/10.1016/0031-0182(85)90032-X)).
- DESACHY B. 2008, *Le STRATIFLIANT, un outil de traitement des données stratigraphiques*, in DJINDJIAN et al. 2008, 187-194 (<https://www.archcalc.cnr.it/journal/articles/487>).
- DESACHY B., DJINDJIAN F. 1990, *Sur l'aide au traitement des données stratigraphiques des sites archéologiques*, «Histoire & Mesure», 5, 1-2, 51-88 (<https://doi.org/10.3406/hism.1990.1365>).
- DESBAT A., PICON M. 1986, *Sigillée claire B et "luisante": classification et provenance*, «Figlina», 7, 15-18.
- DHARIWAL P., NICHOL A. 2021, *Diffusion Models Beat GANs on Image Synthesis*, arXiv: 2105.05233, 1-44 (<https://doi.org/10.48550/arXiv.2105.05233>).
- DIDAY E. 1980, *Optimisation en classification automatique*, Paris, INRIA.
- DIGARD F. (ed.) 1975, *Répertoire analytique des cylindres orientaux*, Paris, Éditions du CNRS.
- DI LUDOVICO A. 2005, *Scene-in-frammenti: una proposta di analisi delle 'scene di presentazione' dei sigilli a cilindro mesopotamici orientata all'elaborazione statistica ed informatica dei dati*, in A. DI LUDOVICO, D. NADALI (eds.), *Studi in onore di Paolo Matthiae presentati in occasione del suo sessantacinquesimo compleanno*, Contributi e Materiali di Archeologia Orientale X-Special issue, Roma, Sapienza, 57-95.
- DI LUDOVICO A. 2011, *Experimental Approaches to Glyptic Art Using Artificial Neural Networks. An Investigation into the Ur III Iconological Context*, in E. JEREM, F. REDŐ, V. SZEVEŘÉNYI (eds.), *On the Road to Reconstructing the Past, Proceedings of the 36th Annual Conference on Computer Applications and Quantitative Methods in Archaeology (Budapest 2008)*, Budapest, Archaeolingua, 135-146.
- DI LUDOVICO A. 2018, *Data Description and the Integrated Study of Ancient Near Eastern Works of Art: The Potential of Cylinder Seals*, in V. BIGOT JULOUX, A.R. GANSELL, A. DI LUDOVICO (eds.),

- CyberResearch on the Ancient Near East and Neighboring Regions. Case Studies on Archaeological Data, Objects, Texts, and Digital Archiving*, Digital Biblical Studies 2, Leiden-Boston, Brill, 85-110 (https://doi.org/10.1163/9789004375086_005).
- DI LUDOVICO A., CAMIZ S. 2014, *A Quantitative Approach to Ur III Mesopotamian Figurative Languages: Reflections, Results, and New Proposals*, «Archeologia e Calcolatori», 25, 7-32 (<https://www.archcalc.cnr.it/journal/articles/713>).
- DI LUDOVICO A., CAMIZ S. 2015, *Ancient Mesopotamian Glyptic Products, Statistics and Data Mining: A Research Proposal*, in GILIGNY et al. 2015, 489-496.
- DI LUDOVICO A., CAMIZ S. 2020, *Image Study of Mesopotamian Cylinder Seals through Texts: The Procrustes Transformation Applied to Correspondence Analysis Results*, «Archeologia e Calcolatori», 31.1, 7-32 (<https://doi.org/10.19282/ac.31.1.2020.01>).
- DI LUDOVICO A., CAMIZ S. 2022, *Experimenting Generalized Procrustes Analysis on a Corpus of Scenes Carved on Mesopotamian Cylinder Seals and Encoded in a Textual Data Set*, «Archeologia e Calcolatori», 33.2, 13-32 (<https://doi.org/10.19282/ac.33.2.2022.02>).
- DI LUDOVICO A., CAMIZ S., PIERI G. 2013, *Comparative Use of Mathematical Models in an Investigation on Mesopotamian Cylinder Seals*, in F. CONTRERAS, M. FARJAS, F.J. MELERO (eds.), *Fusion of Cultures. Proceedings of the 38th Annual Conference on Computer Applications and Quantitative Methods in Archaeology (Granada 2010)*, BAR International Series 2494, Oxford, Archaeopress, 495-498.
- DI LUDOVICO A., PIERI G. 2011a, *How to Facilitate Interpretation of Natural Computation Results by Converting Binary Codes of Images Back to Images*, «CiiT International Journal of Artificial Intelligent Systems and Machine Learning», 3, 7, 437-446 (<https://ciitresearch.org/dl/index.php/aiml/article/view/AIML072011006>).
- DI LUDOVICO A., PIERI G. 2011b, *Artificial Neural Networks and Ancient Artefacts: Justifications for a Multiform Integrated Approach Using PST and Auto-CM Models*, «Archeologia e Calcolatori» 22, 99-128 (<https://www.archcalc.cnr.it/journal/articles/589>).
- DI LUDOVICO A., RAMAZZOTTI M. 2008, *Reconstructing Lexicography in Glyptic Art: Structural Relations between the Akkadian Age and the Ur III Period*, in R. D. BIGGS, J. MYERS, M. ROTH (eds.), *Proceedings of the 51st Rencontre Assyriologique Internationale, Held at the Oriental Institute of the University of Chicago (July 2005)*, Studies in Ancient Oriental Civilization 62, Chicago, Oriental Institute, 263-280.
- DIVE J. 1985, *Application de l'analyse des données à différentes étapes de l'étude quantitative du petit outillage de Terra Amata*, in H. DUCASSE (ed.), *Panorama 1985 des traitements de données en archéologie*, Valbonne, ADPCA, 77-98.
- DJINDJIAN F. 1976, *Contribution de l'analyse des données à l'étude de l'outillage de pierre taillée*, Paris, Université de Paris I Panthéon-Sorbonne.
- DJINDJIAN F. 1977a, *Burins de Noailles, burins sur troncature et sur cassure: statistique descriptive appliquée à l'analyse typologique*, «Bulletin de la Société préhistorique française», 74, 5, 145-154 (<https://doi.org/10.3406/bspf.1977.8497>).
- DJINDJIAN F. 1977b, *Étude quantitative des séries aurignaciennes de la Ferrassie par l'analyse des données*, «Bulletin de la Société préhistorique française», 74, 343-361 (<https://doi.org/10.3406/bspf.1977.8457>).
- DJINDJIAN F. 1980a, *Construction de systèmes d'aide à la connaissance en archéologie préhistorique. Structuration et affectation*, Thèse de doctorat, Université de Paris 1 Panthéon-Sorbonne.
- DJINDJIAN F. 1980b, *Faciès chronologiques aurignaciens et périgordiens à la Ferrassie (Dordogne)*, «Les dossiers de l'archéologie», 142, 70-74.
- DJINDJIAN F. (ed.) 1985a, *Rapport d'activités 1985 du Séminaire «Informatique et mathématiques appliquées en Archéologie»*, Paris-GRAQ.
- DJINDJIAN F. 1985b, *Seriation and Toposeriation by Correspondence Analysis*, in VOORRIPS, LOVING 1985, 119-136.
- DJINDJIAN F. 1985c, *Typologie et culture: l'exemple de l'Aurignacien*, in M. OTTE (ed.), *La signification culturelle des industries lithiques*, BAR International Series 239, Oxford, BAR, 338-373.
- DJINDJIAN F. 1986a, *Recherche sur l'Aurignacien du Périgord à partir des données nouvelles de la Ferrassie*, «L'anthropologie», 90, 1, 89-106.

- DJINDJIAN F. 1986b, *Apport des techniques statistiques et informatiques aux méthodes et aux théories de l'archéologie*, «Bulletin de la Société préhistorique française», 83, 10, 372-377 (<https://doi.org/10.3406/bspf.1986.8715>).
- DJINDJIAN F. 1988a, *Les rapports entre les industries magdaléniennes, cresswelliennes et hambourgiennes du Nord de l'Europe*, in M. OTTE (ed.), *De la Loire à l'Oder. Les civilisations du Paléolithique final dans le nord-ouest européen, Actes du colloque de Liège 1985*, BAR International Series 444, Oxford, BAR, 683-706.
- DJINDJIAN F. 1988b, *Improvements in Intra-Site Spatial Analysis Techniques*, in RAHTZ 1988, 95-105.
- DJINDJIAN F. 1989, *Fifteen Years of Contributions of the Data Analysis French School to Quantitative Archaeology*, in RAHTZ, RICHARDS 1989, 193-204.
- DJINDJIAN F. 1991, *Méthodes pour l'archéologie*, Paris, Armand Colin.
- DJINDJIAN F. 1993a, *L'Aurignacien du Périgord: une révision*, «Préhistoire européenne», 3, 29-54.
- DJINDJIAN F. 1993b, *Les origines du peuplement aurignacien en Europe*, in J. PAVÚK (ed.), *Actes du XII^e Congrès international des sciences préhistoriques et protohistoriques (Bratislava 1991)*, Bratislava, Institut Archéologique de l'Académie Slovaque des Sciences, 2, 136-154.
- DJINDJIAN F. 1993-94, *Histoires de burins*, «Bulletin du Centre Genevois d'Anthropologie», 3-4, 3-21.
- DJINDJIAN F. 1996, *Les industries aurignacoïdes en Aquitaine entre 25000 BP et 15000 BP*, in A. PALMA DI CESNOLA, A. MONTET-WHITE, K. VALOCH (eds.), *Actes du 13^e Congrès de l'UISPP, Colloque 11, The Late Aurignacian (Forlì 1996)*, Forlì, ABACO, 6, 41-54.
- DJINDJIAN F. 1997, *L'analyse spatiale de l'habitat pré- et protohistorique perspectives et limites des méthodes actuelles*, in G. AUXIETTE, L. HACHEM, B. ROBERT, A. BOCQUET, *Espaces physiques, espaces sociaux dans l'analyse interne des sites du Néolithique à l'âge du Fer. Actes du 11^e Congrès National des Sociétés Historiques et Scientifiques (Amiens 1994)*, Paris, C.T.H.S., 3-21.
- DJINDJIAN F. 1999, *L'analyse spatiale de l'habitat: un état de l'art*, «Archeologia e Calcolatori», 10, 17-32 (<https://www.archcalc.cnr.it/journal/articles/263>).
- DJINDJIAN F. 2002, *Pour une théorie générale de la connaissance en archéologie*, in DJINDJIAN, MOSCATI 2002, 101-117 (<https://www.archcalc.cnr.it/journal/articles/337>).
- DJINDJIAN F. 2004, *L'art paléolithique dans son système culturel. II. De la variabilité des bestiaires représentés dans l'art pariétal et mobilier paléolithique*, in M. OTTE (ed.), *La spiritualité, Actes du Colloque international (Liège 2003)*, Liège, ÉRAUL, 127-152.
- DJINDJIAN F. 2009, *The Golden Years for Mathematics and Computers in Archaeology (1965-1985)*, in MOSCATI 2009, «Archeologia e Calcolatori», 20, 61-73 (<https://www.archcalc.cnr.it/journal/articles/504>).
- DJINDJIAN F. 2010, *Quantifier les processus archéologiques*, in K.T. BIRÓ (ed.), *Quantitative Methods for the Challenges in 21st Century Archaeology, Proceedings of the Annual Meeting of the UISPP 4th Commission (Budapest, Hungarian National Museum, 2009)*, «Archeologia e Calcolatori», 21, 233-247 (<https://www.archcalc.cnr.it/journal/articles/576>).
- DJINDJIAN F. 2011, *Manuel d'archéologie*, Paris, Armand Colin.
- DJINDJIAN F. 2013a, *Fonctions, significations et symbolismes des représentations animalières paléolithiques*, in J. CLOTTES (ed.), *L'art pléistocène dans le monde. Actes du Congrès IFRAO (Tarascon-sur-Ariège 2010)*, «Palethnologie. Archéologie et sciences humaines», 5 (<https://doi.org/10.4000/palethnologie.5438>).
- DJINDJIAN F. 2013b, *Us et abus du concept de chaîne opératoire dans l'archéologie française contemporaine*, in S. KRAUSZ, A. COLIN, K. GRUEL, I. RALSTON, T. DECHEZLEPRÊTRE (eds.), *Archéologie de l'âge du Fer en Europe. Mélanges offerts à Olivier Buchsenschutz*, Bordeaux, Ausonius, 93-107 (<https://doi.org/10.46608/basic1.9782356134929.11>).
- DJINDJIAN F. 2016, *The Revolution of the Sixties in Prehistory and Protohistory*, in G. DELLEY, M. DÍAZ-ANDREU, F. DJINDJIAN, V.M.FERNÁNDEZ, A. GUIDI, M-A. KAESER (eds.), *History of Archaeology: International Perspectives, Proceedings of the XVII UISPP World Congress (Burgos 2014)*, vol. 11, Oxford, Archaeopress, 125-144 (<https://doi.org/10.2307/jj.15136066>).
- DJINDJIAN F. 2017, *The UISPP Commission IV: Theory, Methods and Computing*, Virtual Museum of Archaeological Computing (<https://archaeologicalcomputing.cnr.it/itineraries/institutions/the-uispp-commission-iv-archaeological-method-and-theory/>).
- DJINDJIAN F., BOSSELIN B. 1994, *Périgordien et Gravettien: l'épilogue d'une contradiction?*, «Préhistoire européenne», 6, 117-132 (<https://hal.science/hal-00201599v1>).

- DJINDJIAN F., DE CROISSET E. 1976, *Étude typométrique d'une série de deux cents bifaces moustériens de Tabatière (Dordogne) par l'analyse des données*, in *Thèmes spécialisés, IX^e Congrès UISPP (Nice 1976)*, Nice, Prétrages, 38-50.
- DJINDJIAN F., DUCASSE H. (eds.) 1987, *Data Processing and Mathematics Applied to Archaeology. Mathématiques et Informatique appliquées à l'Archéologie*, «PACT», 16.
- DJINDJIAN F., LEREDDE H. 1980, *Traitement automatique des données en archéologie*, «Les dossiers de l'archéologie», 42.
- DJINDJIAN F., MOSCATTI P. (eds.) 2002, *XIV UISPP Congress (Liège-Belgium 2001). Proceedings of Commission IV Symposia. Data Management and Mathematical Methods in Archaeology*, «Archeologia e Calcolatori», 13 (<https://www.archcalc.cnr.it/journal/volumes/13>).
- DJINDJIAN F., NOIZET H., COSTA L., POUGET F. (eds.) 2008, *Webmapping dans les sciences historiques et archéologiques, Actes du Colloque international (Paris 2008)*, «Archeologia e Calcolatori», 19 (<https://www.archcalc.cnr.it/journal/volumes/19>).
- DJINDJIAN F., VIGNERON E. 1980, *L'analyse des données au service de l'archéologie préhistorique*, «Bulletin de la Société préhistorique française», 77, 6, 177-180 (<https://doi.org/10.3406/bspf.1980.5265>).
- DJINDJIAN F., VITALI V. 1992, *Reconstructing Stratigraphy: A Discrete Sampling Approach*, in G. LOCK, J. MOFFETT (eds.), *Computer Applications and Quantitative Methods in Archaeology 1991*, BAR International Series 577, Oxford, Tempus Reparatum, 179-182.
- DO V.H., CANZAR S. 2021, *A Generalization of t-SNE and UMAP to Single-Cell Multimodal Omics*, «Genome Biology», 22, 130, 1-9 (<https://doi.org/10.1186/s13059-021-02356-5>).
- DOLUKHANOV P.M., KOZŁOWSKI J.E., KOZŁOWSKI S.K. 1980, *Multivariate Analysis of Upper Palaeolithic and Mesolithic Stone Assemblages*, Krakow, Universitatis Jagellonicae.
- DONHO D. 2017, *50 Years of Data Science*, «Journal of Computational and Graphical Statistics», 26, 745-766 (<https://doi.org/10.1080/10618600.2017.1384734>).
- DORAN J., HODSON F.R. 1966, *A Digital Computer Analysis of Palaeolithic Flint Assemblages*, «Nature», 210, 688-689 (<https://doi.org/10.1038/210688a0>).
- DORAN J., HODSON F.R. 1975, *Mathematics and Computers in Archaeology*, Edinburgh, Edinburgh University Press.
- DORRITY M.W., SAUNDERS L.M., QUEITSCH C., FIELDS S., TRAPNELL C. 2020, *Dimensionality Reduction by UMAP to Visualize Physical and Genetic Interactions*, «Nature Communications», 11, 1537, 1-6 (<https://doi.org/10.1038/s41467-020-15351-4>).
- DRAP P. 2012, *Archéologie sous-marine: relevé et réalité virtuelle pour l'étude de sites inaccessibles*, in F. GILIGNY, L. COSTA, F. DJINDJIAN, P. CIÉZAR, B. DESACHY (eds.), *Actes des 2^{èmes} Journées d'Informatique et Archéologie de Paris – JIAP 2010 (Paris 2010)*, «Archeologia e Calcolatori», Suppl. 3, 427-439 (<https://www.archcalc.cnr.it/supplements/articles/659>).
- DUCASSE H. 1986, *La Néo-Archéo-Informatique ou les possibilités de faire du neuf avec de l'ancien*, «Bulletin de la Société préhistorique française», 83, 10, 291-293 (<https://doi.org/10.3406/bspf.1986.8699>).
- DUMONCEL J. 2017, *Analyse morphométrique 3D de structures anatomiques pour la paléanthropologie. Modélisation et simulation*, Toulouse, Université Paul Sabatier-Toulouse III.
- EDZARD D.O. 1976-1980, *Keilschrift*, in *Reallexikon der Assyriologie*, 5, Berlin-New York, De Gruyter, 544-568.
- EISENMANN V., TURLOT J.-C. 1978, *Sur la taxinomie du genre Equus*, «Les Cahiers de l'analyse des données», 3, 2, 179-201 (https://www.numdam.org/item/CAD_1978__3_2_179_0/).
- EMMERT-STREIB F., DEHMER M. 2019, *Defining Data Science by a Data-Driven Quantification of the Community*, «Machine Learning and Knowledge Extraction», 1, 235-251 (<https://doi.org/10.3390/make1010015>).
- ESCOFIER B. 1969, *L'analyse factorielle des correspondances*, «Cahiers du Bureau universitaire de recherche opérationnelle. Série Recherche», 13, 25-59.
- ESCOFIER B. 2003, *Analyse des correspondances. Recherches au cœur de l'analyse des données*, Rennes, Presses universitaires de Rennes.
- ESCOFIER B., PAGÈS J. 1984, *L'analyse factorielle multiple*, «Cahiers du Bureau universitaire de recherche opérationnelle. Série Recherche», 42, 3-68.
- ESCOFIER B., PAGÈS J. 2008, *Analyses factorielles simples et multiples. Objectifs, méthodes et interprétation*, Paris, Dunod.

- ESTER M. 1981, *A Column-Wise Approach to Seriation*, «American Antiquity», 46, 3, 496-512 (<https://doi.org/10.2307/280597>).
- FALKENSTEIN A. 1964, *Das Sumerische*, Handbuch der Orientalistik 1.2.1, Leiden, Brill.
- FÉNELON J.-P. 1981, *Qu'est-ce que l'analyse des données?*, Paris, Lefonen.
- FISHER R.A. 1936, *The Use of Multiple Measurements in Taxonomic Problems*, «Annals of Eugenics», 7, 2, 179-188 (<https://doi.org/10.1111/j.1469-1809.1936.tb02137.x>).
- FISHER R.A. 1940, *The Precision of Discriminant Function*, «Annals of Eugenics», 10, 422-429 (<https://doi.org/10.1111/j.1469-1809.1940.tb02264.x>).
- FOTHERINGHAM S., ROGERSON P. 1994, *Spatial Analysis and GIS*, London, Taylor & Francis.
- FRIEDRICH J. 1954, *Entzifferung verschollener Schriften und Sprachen*, Verständliche Wissenschaft 51, Berlin-Heidelberg, Springer.
- FROMAGEOT-LANIEPCE V. 2019, *Les pratiques de la recherche en archéologie à l'heure du numérique. L'évolution de la recherche d'information et de la publication de 1955 à nos jours*, «Archeologia e Calcolatori», Suppl. 12 (<https://doi.org/10.19282/ACS.12.2019>).
- GALE N.H. 1991, *Copper Oxhide Ingots: Their Origins and their Place in the Bronze Age Metals Trade in the Mediterranean*, in N.H. GALE (ed.), *Bronze Age Trade in the Mediterranean. Papers Presented at the Conference Held at Rewley, Oxford, in December 1989*, Jonsered, P. Åström, 197-239.
- GALE N.H., STOS-GALE Z.A. 1982, *Bronze Age Copper Sources in the Mediterranean: A New Approach*, «Science», 216, 11-19 (<https://doi.org/10.1126/science.216.4541.11>).
- GALLAY A. 1989, *Logicism: A French View of Archaeological Theory Founded in Computational Perspective*, «Antiquity», 63, 27-39 (<https://doi.org/10.1017/S0003598X00075554>).
- GALLAY A. 2007, *25 ans de logicisme en archéologie: quel bilan?*, in J. EVIN (ed.), *Congrès du centenaire: un siècle de construction du discours scientifique en Préhistoire. 26^e Congrès préhistorique de France (Avignon 2004)*, Paris, Société Préhistorique Française, 23-36.
- GALLAY A., GARDIN J.-C. 2009, *Les méthodes logicistes en archéologie. Perspectives et limites*, in B. WAL-LISER (ed.), *La cumulativité du savoir en sciences sociales*, Paris, Éditions de l'ÉHÉSS, 111-161 (<https://doi.org/10.4000/books.editionschess.20562>).
- GARDIN J.-C. 1955, *Problèmes de la documentation*, «Diogenes», 11, 107-124.
- GARDIN J.-C. 1958, *Four Codes for the Description of Artifacts: An Essay in Archeological Technique and Theory*, «American Anthropologist», 60, 2, 335-357 (<https://doi.org/10.1525/aa.1958.60.2.02a00090>).
- GARDIN J.-C. 1959, *On the Coding of Geometrical Shapes and Other Representations, with Reference to Archaeological Documents*, in *Proceedings of the International Conference on Scientific Information (Washington, D.C. 1958)*, Washington, D.C., National Academy of Sciences, 107-124.
- GARDIN J.-C. 1966, *Éléments d'un modèle pour la description des lexiques documentaires*, «Bulletin des bibliothèques de France», 5, 171-182.
- GARDIN J.-C. 1967, *Methods for the Descriptive Analysis of Archaeological Material*, «American Antiquity», 32, 13-30 (<https://doi.org/10.2307/278775>).
- GARDIN J.-C. (ed.) 1970, *Archéologie et calculateurs. Problèmes sémiologiques et mathématiques. Actes du Colloque international (Marseille 1969)*, Paris, Éditions du CNRS.
- GARDIN J.-C. 1971, *Archéologie et calculateurs: nouvelles perspectives*, «Revue internationale des sciences sociales», 23, 204-218 (https://unesdoc.unesco.org/ark:/48223/pf0000003802_fre).
- GARDIN J.-C. 1976, *Codes pour l'analyse des formes de poteries*, Paris, Éditions du CNRS.
- GARDIN J.-C. 1977, *Conclusion*, in MÉTHODES CLASSIQUES 1977, 313-318.
- GARDIN J.-C. 1978, *Code pour l'analyse des ornements*, Paris, Éditions du CNRS.
- GARDIN J.-C. 1980, *Archaeological Constructs. An Aspect of Theoretical Archaeology*, Cambridge, Cambridge University Press.
- GARDIN J.-C. 1981, *Le analisi dei discorsi*, Napoli, Liguori Editore.
- GARDIN J.-C. 1996, *La révolution cognitive et l'archéologie*, in MOSCATI 1996a, 1221-1230 (<https://www.archcalc.cnr.it/journal/articles/227>).
- GARDIN J.-C. 1997, *Le questionnement logiciste et les conflits d'interprétation*, «Enquête», 5, 35-54 (<https://doi.org/10.4000/enquete.1043>).
- GARDIN J.-C. 2002, *Les modèles logico-discursifs en archéologie*, in DJINDJIAN, MOSCATI 2002 (<https://www.archcalc.cnr.it/journal/articles/332>).

- GARDIN J.-C. 2003, *Archaeological Discourse, Conceptual Modelling and Digitalisation: An Interim Report of the Logicist Program*, in M. DOERR, A. SARRIS (eds.), *The Digital Heritage of Archaeology. Proceedings of the 30th CAA Conference (Heraklion, Crete 2002)*, Athens, Hellenic Ministry of Culture, 5-11 (<http://dx.doi.org/10.15496/publikation-3417>).
- GARDIN J.-C., GUILLAUME O., HERMAN P.-Q., HESNARD A., LAGRANGE M.-S., RENAUD M., ZADORA-RIO E. 1987, *Systèmes experts et sciences humaines. Le cas de l'archéologie*, Paris, Eyrolles.
- GARDIN J.-C., LAGRANGE M.-S. 1975, *Essais d'analyse du discours archéologique*, Valbonne, Centre de recherches archéologiques.
- GARDIN J.-C., LÉVY F. 1962, *Le SYNTOL Syntagmatic Organization Language*, in C.M. POPPLEWELL (ed.), *Information Processing 1962, Proceedings of IFIP Congress 62 (Munich 1962)*, Amsterdam, North-Holland, 279-283.
- GASSE F. 1986, *East African Diatoms and Water pH*, in J.P. SMOLL, R.W. BATTARBEE, R.B. DAVIS, J. MERILÄINEN (eds.), *Diatoms and Lake Acidity*, Dordrecht, Junk, 149-168.
- GASSE F., TÉKAÏA F. 1979, *La paléocéologie des diatomées: évolution des lacs de l'Afar central*, «Les Cahiers de l'Analyse des données», 4, 1, 81-94 (https://www.numdam.org/item/CAD_1979__4_1_81_0/).
- GATTIGLIA G. 2022, *A Postphenomenological Perspective on Digital and Algorithmic Archaeology*, «Archeologia e Calcolatori», 33.2, 319-334 (<https://doi.org/10.19282/ac.33.2.2022.17>).
- GATTIGLIA G. 2025, *Managing Artificial Intelligence in Archeology. An Overview*, «Journal of Cultural Heritage», 71, 225-233 (<https://doi.org/10.1016/j.culher.2024.11.020>).
- GERHARD E. 1840-1867, *Etruskische Spiegel*, I-V, Berlin, Reimer (<https://doi.org/10.11588/diglit.5021>).
- GIFI A. 1980, *Niet-Lineaire Multivariate Analyse*, Leiden, University of Leiden.
- GIFI A. 1990, *Nonlinear Multivariate Analysis*, New York, John Wiley & Sons.
- GILIGNY F. 1990, *La reconnaissance des formes céramiques: une approche typologique formalisée*, «Histoire & Mesure», 5, 12, 89-107.
- GILIGNY F. 2008, *Une méthode d'analyse spatiale des rejets domestiques par les remontages céramiques. Étude d'un site néolithique jurassien: Chalain station 3*, in DJINDJIAN et al. 2008, 173-186 (<https://www.archcalc.cnr.it/journal/articles/486>).
- GILIGNY F., DJINDJIAN F., COSTA L., MOSCATI P., ROBERT S. (eds.) 2015, *CAA 2014. 21st Century Archaeology. Concepts, Methods and Tools. Proceedings of the 42nd Annual Conference on Computer Applications and Quantitative Methods in Archaeology (Paris 2014)*, Oxford, Archaeopress.
- GINOUVÈS R. 1971, *Archéographie, archéométrie, archéologie. Pour une informatique de l'archéologie gréco-romaine*, «Revue Archéologique», 93-126.
- GINOUVÈS R. 1988, *Perspectives actuelles de l'archéologie classique*, «Classica, Revista de Sociedade Brasileira de Estudos Clasicos», 1, 111-126 (<https://doi.org/10.24277/classica.v1i1.648>).
- GINOUVÈS R., GUIMIER-SORBETS A.-M. 1978, *La constitution des données en archéologie classique*, Paris, Éditions du CNRS.
- GIUDICE F. 1993, *Le rotte commerciali dei vasi attici dal VI al IV sec. a.C. Analisi quantitativa e qualitativa*, «Archeologia e Calcolatori», 4, 181-196.
- GIUDICE F., BARBERI S., BARRESI S., RANDAZZO M. 1999, *L'archivio ceramografico dell'Università di Catania ed il Progetto Post-Paralipomena*, «Archeologia e Calcolatori», 10, 79-88 (<https://www.archcalc.cnr.it/journal/articles/267>).
- GOB A. 1986, *Utilisation du SAS, logiciel d'analyse des données, en archéologie*, «Archeo-Log», 1, 59-78.
- GOB A. 1988, *Multivariate Analysis of Lithic Industries*, in RUGGLES, RAHTZ 1988, 15-24.
- GOLDMANN K. 1973, *Zwei Methoden chronologischer Gruppierung*, «Acta Praehistorica et Archaeologica», 3, 1-34 (<https://doi.org/10.11588/apa.1972.0.67902>).
- GOODFELLOW I.J., POUGET-ABADIE J., MIRZA M., XU B., WARDE-FARLEY D., OZAIR S., COURVILLE A., BENGIO Y. 2014, *Generative Adversarial Networks*, arXiv: 1406.2661, 1-9 (<https://doi.org/10.48550/arXiv.1406.2661>).
- GOODMAN L.A. 1970, *The Multivariate Analysis of Qualitative Data: Interactions Among Multiple Classifications*, «Journal of the American Statistical Association», 65, 329, 226-256 (<https://doi.org/10.1080/01621459.1970.10481076>).
- GRAHAM I.D. 1980, *Spectral Analysis and Distance Methods in the Study of Archaeological Distributions*, «Journal of Archaeological Science», 7, 105-130 ([https://doi.org/10.1016/S0305-4403\(80\)80015-X](https://doi.org/10.1016/S0305-4403(80)80015-X)).

- GRAHAM J.M., ROE D. 1970, *Discrimination of British Lower and Middle Palaeolithic Handaxe Groups Using Canonical Variates*, «World Archaeology», 1, 321-337 (<https://doi.org/10.1080/00438243.1970.9979450>).
- GREENACRE M.J. 1984, *Theory and Applications of Correspondence Analysis*, London, Academic Press.
- GREENACRE M.J. 1986, *SIMCA: A Program to Perform Simple Correspondence Analysis*, «American Statistician», 51, 230-231 (<https://doi.org/10.2307/2684547>).
- GUENOCHÉ A., HESNARD A. 1983, *Typologie d'amphores romaines par une méthode logique de classification*, «Computers and the Humanities», 17, 185-198 (<https://doi.org/10.1007/BF02252805>).
- GUENOCHÉ A., TCHERNIA A. 1978, *L'analyse descriptive dans la construction d'un modèle typologique des amphores. Dressel 20 (DR 20)*, in BORILLO 1978, 166-181.
- GUERMANDI M.P. 1993, *Gli archeologi classici di fronte al computer: l'esempio di Francia e Italia*, in *Les archéologues et l'archéologie, Colloque de Bourg-en-Bresse (Archives) 1992*, «Caesarodunum», 27, 252-270.
- GUERMANDI M.P., MIGNANI S., MONTANARI A. 1994, *Il trattamento dei dati iconografici nell'ambito della ceramica attica e l'esempio del progetto Spina*, in MOSCATI 1994a, 111-140.
- GUIDI A. 2015, *Quantitative Methods in Italian Archaeology: A Review*, «Archeologia e Calcolatori», 26, 45-58 (<https://www.archcalc.cnr.it/journal/articles/782>).
- GUILBAUD M. 1987, *Élaboration d'un cadre morphotechnique pour l'étude du débitage en typologie analytique de quelques industries des gisements de Saint-Césaire (Charente-Maritime) et de Quincay (Vienne)*, in B. VANDERMEERSCH (ed.), *Préhistoire en Poitou-Charentes: problèmes actuels*, Paris, CTHS, 101-113.
- GUIOT J. 1984, *Deux méthodes d'utilisation de l'épaisseur des cernes ligneux pour la reconstitution de paramètres climatiques anciens*, «Palaeo», 45, 347-368 ([https://doi.org/10.1016/0031-0182\(84\)90011-7](https://doi.org/10.1016/0031-0182(84)90011-7)).
- GUTTMAN L. 1941, *The Quantification of a Class of Attributes: A Theory and Method of Scale Construction*, in P. HORST (ed.), *The Prediction of Personal Adjustment*, New York, Social Science Research Council, 319-348.
- HAMOND E., HESNARD A. 1977, *Problèmes de documentation et de description relatifs à un corpus d'amphores romaines*, in MÉTHODES CLASSIQUES 1977, 18-33 (www.persee.fr/doc/efr_0000-0000_1977_act_32_1_4686).
- HARDY-GUILBERT C., DJINDJIAN F. 1980, *Organisation des décors de stuc sur l'arc de portail de la mosquée de Bust (Afghanistan)*, «Les dossiers de l'archéologie», 42, 88-93.
- HE K., ZHANG X., REN S., SUN J. 2015, *Deep Residual Learning for Image Recognition*, arXiv:1512.03385, 1-12 (<https://doi.org/10.48550/arXiv.1512.03385>).
- HESNARD A. 1987, *Cas n. 4: détermination de l'origine d'amphores romaines*, in GARDIN et al. 1987, 139-167.
- HESNARD A., LEMOINE C. 1981, *Les amphores du Cécube et du Falerne. Prospections, typologie, analyses*, «Mélanges de l'École Française de Rome, Antiquité», 93, 1, 243-295 (<https://doi.org/10.3406/mefr.1981.1276>).
- HESSE A. 1984, *La répartition céramique à la surface de l'enclos SAV2 de l'île de Sai d'après l'analyse des correspondances*, «Cahiers de Recherche de l'Institut de Papyrologie et d'Égyptologie de Lille», 6, 7-67.
- HILL M.O. 1974, *Correspondence Analysis: A Neglected Multivariate Method*, «Journal of the Royal Statistical Society», 23, 3, 340-354 (<https://doi.org/10.2307/2347127>).
- HIRSCHFELD H.O. 1935, *A Connection between Correlation and Contingency*, «Proceedings of the Cambridge Philosophical Society», 31, 520-524 (<https://doi.org/10.1017/S0305004100013517>).
- HODDER I., ORTON C. 1976 *Spatial Analysis in Archaeology*, Cambridge, Cambridge University Press.
- HODSON F.R. 1969, *Searching for Structure within Multivariate Archaeological Data*, «World Archaeology», 1, 299-320 (<https://doi.org/10.1080/00438243.1969.9979430>).
- HODSON F.R. 1971, *Numerical Typology and Prehistoric Archaeology*, in HODSON, KENDALL, TAUTU 1971, 30-45.
- HODSON F.R., KENDALL D.G., TAUTU P. (eds.) 1971, *Mathematics in the Archaeological and Historical Sciences. Proceedings of the Anglo-Romanian Conference (Mamaia 1970)*, Edinburgh, Edinburgh University Press.
- HODSON F.R., SNEATH P.H.R., DORAN J.E. 1966, *Some Experiments in the Numerical Analysis of Archaeological Data*, «Biometrika», 53, 3-4, 311-324 (<https://doi.org/10.1093/biomet/53.3-4.311>).

- HODSON F.R., TYERS P.A. 1988, *Data Analysis for Archaeologists: The Institute of Archaeology Packages*, in RAHTZ 1988, 28-41.
- HOLMES S., JOSSE J. 2017, *50 Years of Data Science, Discussion*, «Journal of Computational and Graphical Statistics» 26, 4, 768-769 (<https://doi.org/10.1080/10618600.2017.1385471>).
- HOTELLING H. 1933, *Analysis of a Complex of Statistical Variables into Principal Components*, «Journal of Educational Psychology», 24, 498-520 (<https://doi.org/10.1037/h0070888>).
- HOTELLING H. 1936, *Relations between Two Sets of Variates*, «Biometrika», 28, 3-4, 321-377 (<https://doi.org/10.1093/biomet/28.3-4.321>).
- HOURS F. 1976, *L'épipaléolithique du Liban: résultats acquis en 1975*, in *Colloque III, Actes du IX^e Congrès UISPP (Nice 1976)*, Paris, Éditions du CNRS, 107-130.
- HOWELLS W.W. 1966, *Cranimetry and Multivariate Analysis*, «Papers of the Peabody Museum of Archeology and Ethnology», 57, 1-43.
- HOWELLS W.W. 1973, *Cranial Variation in Man: A Study by Multivariate Analysis of Patterns of Difference among Recent Human Populations*, Papers of the Peabody Museum of Archeology and Ethnology 67, Cambridge, Mass., Harvard University.
- IACOVELLA A. 1996, *Observations sur la distribution des tombes dans une nécropole grecque d'époque archaïque: le cas de la nécropole Ouest de Mégara Hyblaea*, in MOSCATI 1996a, 373-384 (<https://www.archcalc.cnr.it/journal/articles/152>).
- IACOVELLA A. 1997, *Études des proximités dans l'espace funéraire: le cas de la nécropole occidentale de Mégara Hyblaea*, «Archeologia e Calcolatori», 8, 67-102 (<https://www.archcalc.cnr.it/journal/articles/236>).
- IACOVELLA A., AUDA Y. 1994, *Nécropoles de Sicile: étude de l'utilisation des espaces funéraires dans le temps (du IX^e au I^{er} s. av. J.C.)*, in MOSCATI 1994a, 69-86.
- IHM P. 1978, *Statistik in der Archäologie. Probleme der Anwendung, allgemeine Methoden, Seriation und Klassifikation*, «Archaeo-Physika», 9, Köln, Rheinland-Verlag-Bonn, Habelt.
- IHM P. 1983, *Korrespondenzanalyse und Seriation*, «Archäologische Informationen», 6, 1, 8-21 (<https://doi.org/10.11588/ai.1983.1.27644>).
- IMBRIE J., KIPP N.G. 1971, *A New Micropaleontological Method for Quantitative Paleoclimatology: Application to a Late Pleistocene Caribbean Core*, in R.F. FLINT, K.K. TURRIKIAN (eds.), *The Late Cenozoic Glacial Ages*, New Haven, Yale University Press, 71-181.
- IWATSUBO S. 2023, *My Recollections of People in the World of Data Science*, in E.J. BEH, R. LOMBARDO, J.G. CLAVEL (eds.), *Analysis of Categorical Data from Historical Perspectives. Essays in Honour of Shizubiko Nishisato*, Singapore, Springer, 39-55.
- JAMBU M., LEBEAUX M.O. 1983, *Cluster Analysis for Data Analysis*, Amsterdam, North Holland.
- JOHNSON I. 1977, *Local Density Analysis: A New Method for Quantitative Spatial Analysis*, in S. LAFLIN (ed.), *Computer Applications in Archaeology 1977. Proceedings of the Annual Conference*, Birmingham, University of Birmingham, 90-98.
- JOHNSON I. (ed.) 1994, *Methods in the Mountains. Proceedings of UISPP Commission IV Meeting (Mount Victoria, Australia, 1993)*, Sydney University Archaeological Methods Series 2, Sydney.
- KAMPFMEYER U. 1986, *ARCOS et la classification des céramiques assistée par ordinateur*, «Archéologie et Ordinateurs», 8, 3-23.
- KAY P.A. 1979, *Multivariate Statistical Estimates of Holocene Vegetation and Climate Change, Forest-Tundra Transition Zone*, «Quaternary Research», 11, 1, 125-140 ([https://doi.org/10.1016/0033-5894\(79\)90073-5](https://doi.org/10.1016/0033-5894(79)90073-5)).
- KELLY-BUCCELLATI M. 1977, *Towards the Use of Quantitative Analysis in Mesopotamian Sphragistics*, «Mesopotamia», 12, 41-52 (<https://doi.org/10.6084/m9.figshare.8293991>).
- KELLY-BUCCELLATI M. 1979-1980, *Iconographic and Formal Analysis of Mesopotamian Cylinder Seals: System Rescription*, in *Colloque International d'études ugaritiques à l'occasion du Cinquantenaire des fouilles archéologiques à Ras Shamra (UGARIT) (Lattaquié 1979)*, «Annales Archéologiques Arabes Syriennes. Revue d'archéologie et d'histoire», 29-30, Special issue, 273-284 (<https://doi.org/10.6084/m9.figshare.8294000>).
- KELLY-BUCCELLATI M., ELSTER E. 1973, *Statistics in Archaeology and its Application to Ancient Near Eastern Data*, in G. BUCCELLATI (ed.), *Gelb Volume: Approaches to the Study of the Ancient Near*

- East*, «Orientalia», 42, Special Issue, Roma, Pontificium Institutum Biblicum, 195-211 (<https://doi.org/10.6084/m9.figshare.8293979>).
- KENDALL D.G. 1969, *Some Problems and Methods in Statistical Archaeology*, «World Archaeology», 1, 1, 68-76 (<https://doi.org/10.1080/00438243.1969.9979428>).
- KENDALL D.G. 1970, *A Mathematical Approach to Seriation*, «Philosophical Transactions of the Royal Society of London. Series A», 269, 125-134 (<http://doi.org/10.1098/rsta.1970.0091>).
- KENDALL D.G. 1971, *Seriation from Abundance Matrices*, in HODSON, KENDALL, TAUTU 1971, 215-252.
- KENDALL D.G. 1985, *Mathematical Statistics in the Humanities, and Some Related Problems in Astronomy*, in A.C. ATKINSON, S.E. FIENBERG (eds.), *A Celebration of Statistics. The ISI Centenary Volume*, New York, Springer, 393-408 (https://doi.org/10.1007/978-1-4613-8560-8_17).
- KINGMA D.P., WELING M. 2013, *Auto-Encoding Variational Bayes*, arXiv: 1312.6114, 1-14 (<https://doi.org/10.48550/arXiv.1312.6114>).
- KINTIGH K.W., AMMERMANN A.J. 1982, *Heuristic Approaches to Spatial Analysis in Archaeology*, «American Antiquity», 47, 31-63 (<https://doi.org/10.2307/280052>).
- KOHONEN T. 1989, *Self-Organization and Associative Memory*, Berlin, Springer (<https://doi.org/10.1007/978-3-642-88163-3>).
- KOMAR P., BRUGHMANS T., BORISOVA E. 2025, *Consumption Trends, Trading Patterns and Economic Development in Italy Across Centuries: Data Analysis of Roman Amphorae in a Long-Term Perspective*, «Journal of Archaeological Method and Theory», 32, 21 (<https://doi.org/10.1007/s10816-024-09686-1>).
- KRIZHEVSKY A., SUTSKEVER I., HINTON G.E. 2012, *ImageNet Classification with Deep Convolutional Neural Networks*, in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 1097-1105.
- KRUSKAL J.B. 1964, *Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis*, «Psychometrika», 29, 1-27 (<https://doi.org/10.1007/BF02289565>).
- KURTZ D. 2009, *www.beazley.ox.ac.uk. From Apparatus of Scholarship to Web Resource. The Beazley Archive 1970-2008*, in MOSCATI 2009, 37-46 (<https://www.archcalc.cnr.it/journal/articles/502>).
- LAGRANGE M.-S., BONNET CH. 1978, *Les chemins de la "memoria". Nouvel essai d'analyse du discours archéologique*, Paris, Éditions du CNRS.
- LAGRANGE M.-S., MONTEIRO RODRIGUES M.C. 1994, *La représentation d'un raisonnement en archéologie. Un exemple: analyse logiciste et système expert*, in MOSCATI 1994a, 333-353.
- LAGRANGE M.-S., RENAUD M. 1983a, *Simulation d'un raisonnement archéologique. Description de l'application d'un système expert: le système SNARK*, in H. DUCASSE (ed.), *Panorama 1983 des traitements de données en archéologie*, Valbonne, A.P.D.C.A., 31-64.
- LAGRANGE M.-S., RENAUD M. 1983b, *L'interprétation des documents figurés en archéologie et histoire de l'art: essai de simulation sur ordinateur*, in F. LISSARRAGUE, F. THELAMON F. (eds.), *Image et céramique grecque. Actes du Colloque (Rouen 1982)*, Rouen, Université de Rouen, 43-66.
- LAGRANGE M.-S., RENAUD M. 1985, *Intelligent Knowledge-Based Systems in Archaeology: A Computerized Simulation of Reasoning by Means of an Expert System*, «Computers and the Humanities», 19, 37-52 (<https://doi.org/10.1007/BF02259616>).
- LAPLACE G., MERINO J.M. 1979, *Application de la typologie analytique et structurale à l'étude du processus d'azilianisation: la série phylétique de la grotte Urriaga en Pays Basque*, in D. DE SONNEVILLE-BORDES (ed.), *La fin des temps glaciaires en Europe. Chronostratigraphie et écologie des cultures du Paléolithique final, Actes du Colloque (Talence 1977)*, Paris, Éditions du CNRS, vol. 2, 694-710.
- LAURIÈRE J.-L. 1986, *Un langage déclaratif: SNARK*, «Technique et science informatiques», 3, 141-172.
- LAURIN B., ROUSSEAU D.D. 1985, *Analyse multivariée des associations malacologiques d'Achenheim. Implications climatiques et environnementales*, «Bulletin de l'Association française pour l'étude du Quaternaire», 22, 1, 21-30 (<https://doi.org/10.3406/quate.1985.1523>).
- LEBART L. 1978, *Programme d'agrégation avec contraintes*, «Les Cahiers de l'analyse des données», 3, 3, 275-287 (https://www.numdam.org/item/CAD_1978__3_3_275_0/).
- LEBART L., MORINEAU A. 1982, *SPAD. Système Portable pour l'Analyse des Données*, Paris, CISIA.
- LEBART L., MORINEAU A., BÉCUE M. 1988, *SPAD.T. Système Portable pour l'Analyse des Données Textuelles. Manuel de l'utilisateur*, Paris, CISIA.

- LEBART L., MORINEAU A., LAMBERT T. 1987, *SPAD.N – Système Portable pour l'Analyse des Données. Manuel de l'utilisateur*, Paris, CISIA.
- LEBART L., MORINEAU A., PIRON M. 1995, *Statistique exploratoire multidimensionnelle*, Paris, Dunod.
- LEBART L., MORINEAU A., TABARD N. 1977, *Techniques de la description statistique. Méthodes et logiciels pour l'analyse de grands tableaux*, Paris, Dunod.
- LEBART L., MORINEAU A., WARWICK K.M. 1984, *Multivariate Descriptive Statistical Analysis. Correspondence Analysis and Related Techniques for Large Matrices*, Chichester, Wiley.
- LEBART L., PINCEMIN B., POUDAT C. 2019, *Analyse des données textuelles*, Québec, Presses de l'Université du Québec.
- LEBART L., PIRON M. 2012, *Pratique de l'analyse des données numériques et textuelles avec Dtm-Vic*, L2C.
- LEBART L., SALEM A. 1988, *Analyse statistique des données textuelles*, Paris, Dunod.
- LEBART L., SALEM A. 1994, *Statistique textuelle*, Paris, Dunod.
- LEBART L., SALEM A., BERRY L. 1998, *Exploring Textual Data*, Text, Speech, and Language Technology 4, Dordrecht, Kluwer (<https://doi.org/10.1007/978-94-017-1525-6>).
- LEBART L., SAPORTA G. 2014, *Historical Elements of Correspondence Analysis and Multiple Correspondence Analysis*, in J. BLASIUS, M. GREENACRE (eds.), *Visualization and Verbalization of Data*, New York, Chapman and Hall, 31-44 (<https://doi.org/10.1201/b16741>).
- LECUN Y., BENGIO Y., HINTON G. 2015, *Deep Learning*, «Nature», 521, 436-444 (<https://doi.org/10.1038/nature14539>).
- LECUN Y., BOSER B., DENKER J.S., HENDERSON D., HOWARD R.E., HUBBARD W., JACKEL L.D. 1989, *Backpropagation Applied to Handwritten Zip Code Recognition*, «Neural Computation», 1, 541-551 (<https://doi.org/10.1162/neco.1989.1.4.541>).
- LEGOUX R. 1980, *Analyse chronologique relative d'une vaste nécropole: l'exemple du site de Bulles (Oise)*, in P. PERIN (ed.), *La datation des tombes mérovingiennes. Historique-Méthodes-Applications*, Genève, Librairie Droz, 284-307.
- LEMOINE C. 1982, *Méthodes de classification et classement de céramiques. Exemple d'application*, in H. DUCASSE (ed.), *Panorama 1981 des applications informatiques en archéologie*, Valbonne, A.P.D.C.A., 86-106.
- LEREDDE H., PERIN P. 1980, *Les plaques-boudes mérovingiennes*, «Les dossiers de l'archéologie», 42, 83-87.
- LE RIDER G. 1975, *Code pour l'analyse des monnaies*, Paris, Éditions du CNRS.
- LERMAN I.C. 1970, *Les bases de la classification automatique*, Paris, Gauthier-Villars.
- LERMAN I.C. 1981, *Classification et analyse ordinaire des données*, Paris, Dunod.
- LERMAN I.C., LEREDDE H. 2022, *Seriation in Combinatorial and Statistical Data Analysis*, Cham, Springer (<https://doi.org/10.1007/978-3-030-92694-6>).
- LIVERANI M. 2005, *Imperialism*, in S. POLLOCK, R. BERNBECK (eds.), *Archaeologies of the Middle East. Critical Perspectives*, Malden, Blackwell, 223-243.
- MADSEN T. 1985, *Numerisk dataanalyse for arkæologer*, Århus, Århus Universitet.
- MADSEN T. 1988, *Multivariate Archaeology, Numerical Approaches to Scandinavian Archaeology*, Jutland Archaeological Society Publications 21, Aarhus, Aarhus University Press.
- MALLET GAUTHIER S., CHOULAKIAN V. 2015, *Taxicab Correspondence Analysis of Abundance Data in Archaeology: Three Case Studies Revisited*, «Archeologia e Calcolatori», 26, 77-94 (<https://www.archcalc.cnr.it/journal/articles/784>).
- MARCHESINI S. 2004, *Seriazione ed epigrafia: l'impiego di BASP (The Bonn Archaeological Software Package) nello studio di iscrizioni*, in MOSCATI 2004a, 257-266 (<https://www.archcalc.cnr.it/journal/articles/383>).
- MARCHESINI S. 2010, *L'alfabeto atestino. Determinazione cronologica delle iscrizioni dopo analisi con strumenti informatici (The Bonn Archaeological Software)*, «Incidenza dell'Antico», 8, 127-142.
- MARION S. 2002, *Recherches sur l'âge du Fer en Île-de-France. Analyse des sites fouillés entre le Hallstatt final et La Tène finale: chronologie et société*, Thèse de doctorat, Université de Paris I Panthéon-Sorbonne.
- MARSAL E. 2014, *The Beginnings of Sumerology (I). From Early Sketches to a First Complete Grammar*, «Aula Orientalis», 32, 2, 283-297.
- MARSAL E. 2015, *The Beginnings of Sumerology (II). From Delitzsch's Grammar to Adam Falkenstein*, «Aula Orientalis», 33, 2, 255-269.

- MASIERO C. 1987, *Problemi di classificazione di reperti archeologici: una applicazione sulla ceramica cassita*, in D'ANDRIA 1987, 199-212.
- MATSKEVITCH S., SHARON I. 2018, *A Conceptual Framework for Archaeological Data Encoding*, in V. BIGOT JULOUX, A.R. GANSELL, A. DI LUDOVICO (eds.), *CyberResearch on the Ancient Near East and Neighboring Regions. Case Studies on Archaeological Data, Objects, Texts, and Digital Archiving*, Digital Biblical Studies 2, Leiden-Boston, Brill, 25-59 (https://doi.org/10.1163/9789004375086_003).
- MATTHEWS R. 2003, *The Archaeology of Mesopotamia*, London-New York, Routledge.
- MAZZOCATO S., DAFFARA C., MARCHIORO G., MENEGAZZI A. 2021, *Optical Microprofilometry Optimized for Surface Analysis and 3D Printing of Archaeological Objects*, in CARVALE 2021, 227-236 (<https://doi.org/10.19282/ac.32.2.2021.21>).
- MCCULLOCH W.S., PITTS W. 1943, *A Logical Calculus of the Ideas Immanent in Nervous Activity*, «Bulletin of Mathematical Biophysics», 5, 115-133 (<https://doi.org/10.1007/BF02478259>).
- MCINNES L., HEALY J., MELVILLE J. 2020, *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*, arXiv:1802.03426, 1-63 (<https://doi.org/10.48550/arXiv.1802.03426>).
- MEILÀ M., ZHANG H. 2023, *Manifold Learning: What, How, and Why*, arXiv: 2311.03757, 1-33 (<https://doi.org/10.48550/arXiv.2311.03757>).
- MELLO E., MONNA D., ODDONE M. 1988, *Discriminating Sources of Mediterranean Marbles: A Pattern Recognition Approach*, «Archaeometry», 30, 1, 102-108 (<https://doi.org/10.1111/j.1475-4754.1988.tb00438.x>).
- MÉTHODES CLASSIQUES 1977, *Actes du Colloque Méthodes classiques et méthodes formelles dans l'étude des amphores (Rome 1974)*, Rome, École Française de Rome.
- MIKE K., HAZZAN O. 2023, *What is Data Science?*, «Communications of the ACM», 66, 12-13 (<https://doi.org/10.1145/3575663>).
- MILLETT M. (ed.) 1979, *Pottery and the Archaeologist*, London, University College London.
- MOBERG C.A. 1971, *Archaeological Context and Mathematical Methods*, in HODSON, KENDALL, TAUTU 1971, 551-562.
- MODRZEWSKA I., ODDONE M., PIANETTI F., TARONI G. 1993, *Anfore spagnole nel Veneto. Progetto di ricerca e primi prodotti*, «Archeologia e Calcolatori», 4, 127-135.
- MODRZEWSKA I., TARONI G., PIANETTI F. 2010, *Un'anfora frammentaria dalla Laguna di Venezia*, «Archeologia e Calcolatori», 21, 201-210 (<https://www.archcalc.cnr.it/journal/articles/572>).
- MOHEN J.P. 1980, *L'âge du Fer en Aquitaine*, Mémoires de la SPF 14, Paris, Société préhistorique française.
- MOHRI M., ROSTAMIZADEH A., TALWALKAR A. 2012, *Foundations of Machine Learning. Adaptive Computation and Machine Learning*, London, Cambridge Mass, The MIT Press.
- MOOREY P.R.S. 1991, *A Century of Biblical Archaeology*, Cambridge, Lutterworth.
- MOSCATI P. 1984, *Ricerche matematico-statistiche sugli specchi etruschi*, Contributi del Centro Linceo Interdisciplinare di Scienze Matematiche e loro Applicazioni 66, Roma, Accademia Nazionale dei Lincei.
- MOSCATI P. 1986, *Analisi statistiche multivariate sugli specchi etruschi*, Contributi del Centro Linceo Interdisciplinare di Scienze Matematiche e loro Applicazioni 74, Roma, Accademia Nazionale dei Lincei.
- MOSCATI P. 1987a, *Archeologia e Calcolatori*, Firenze, Giunti.
- MOSCATI P. 1987b, *Metodi matematico-statistici negli studi etruschi: il caso degli specchi*, in D'ANDRIA 1987, 277-301.
- MOSCATI P. 1990, *L'analisi quantitativa nell'archeologia di epoca storica*, «Archeologia e Calcolatori», 1, 39-80 (<https://www.archcalc.cnr.it/journal/articles/7>).
- MOSCATI P. (ed.) 1994a, *Choice, Representation and Structuring of Archaeological Information*, «Archeologia e Calcolatori», 5 (<https://www.archcalc.cnr.it/journal/volumes/5>).
- MOSCATI P. 1994b, *Un gruppo di urne etrusche di produzione volterrana: prospettive di analisi quantitativa*, in MOSCATI 1994a, 87-110.
- MOSCATI P. 1995a, *Quantitative Analysis of Etruscan Cinerary Urns*, in J. HUGGETT, N. RYAN (eds.), *CAA94. Computer Applications and Quantitative Methods in Archaeology 1994*, BAR International Series 600, Oxford, Tempus Reparatum, 101-104.
- MOSCATI P. 1995b, *Méthodes quantitatives et problèmes iconographiques*, in L. VALDÉS, I. ARENAL, I. PUJANA (eds.), *Aplicaciones Informáticas en Arqueología: Teorías y Sistemas (Saint-Germain-en-Laye 1991)*, vol. 1, Bilbao, Denboraren Argia, 317-329.

- MOSCATI P. (ed.) 1996a, *III International Symposium on Computing and Archaeology (Roma 1995)*, «Archeologia e Calcolatori», 7 (<https://www.archcalc.cnr.it/journal/volumes/7/>).
- MOSCATI P. 1996b, *Archeologia quantitativa: nascita, sviluppo e "crisi"*, in MOSCATI 1996a, 579-590 (<https://www.archcalc.cnr.it/journal/articles/166>).
- MOSCATI P. 1997, *Metodologie archeologiche fra tradizione e informatizzazione: la classificazione dei materiali*, in T. ORLANDI (ed.), *Discipline umanistiche e informatica. Il problema della formalizzazione*, Contributi del Centro Linceo Interdisciplinare di Scienze Matematiche e loro Applicazioni 70, Roma, Accademia Nazionale dei Lincei, 151-157.
- MOSCATI P. 2000, *Dal mondo della statistica applicata*, «Archeologia e Calcolatori», 11, 396-400 (<https://www.archcalc.cnr.it/journal/articles/312>).
- MOSCATI P. 2003, *Dal mondo della statistica applicata*, «Archeologia e Calcolatori», 14, 340-344 (<https://www.archcalc.cnr.it/journal/articles/367>).
- MOSCATI P. (ed.) 2004a, *New Frontiers of Archaeological Research. Languages, Communication, Information Technology*, «Archeologia e Calcolatori», 15 (<https://www.archcalc.cnr.it/journal/volumes/15>).
- MOSCATI P. 2004b, *Per la descrizione computerizzata delle urne volterrane. Problemi di formalizzazione*, in M. FANO SANTI (ed.), *Studi di archeologia in onore di Gustavo Traversari*, Roma, G. Bretschneider, 647-655.
- MOSCATI P. (ed.) 2009, *La nascita dell'informatica archeologica. Atti del Convegno Internazionale (Roma, Accademia Nazionale dei Lincei, 2008)*, «Archeologia e Calcolatori», 20 (<https://www.archcalc.cnr.it/journal/volumes/20>).
- MOSCATI P. 2013, *Jean-Claude Gardin (Parigi 1925-2013). Dalla meccanografia all'informatica archeologica*, «Archeologia e Calcolatori», 24, 7-24 (<https://www.archcalc.cnr.it/journal/articles/660>).
- MOSCATI P. 2015, *Towards a History of Archaeological Computing: An Introduction*, in GILIGNY et al. 2015, 9-16.
- MOSCATI P. 2018, *Quantifying Archaeology: An Interview with Frank Roy Hodson*, Virtual Museum of Archaeological Computing (<http://archaeologicalcomputing.cnr.it/itineraries/protagonists/quantifying-archaeology-an-interview-with-frank-roy-hodson/>).
- MÜLLER J., ZIMMERMANN A. (eds.) 1997, *Archäologie und Korrespondenzanalyse. Beispiele, Fragen, Perspektiven*, Leidorf, Espelkamp.
- MURTAGH F. 2008, *Origins of Modern Data Analysis Linked to the Beginnings and Early Development of Computer Science and Information Engineering*, «Electronic Journal for History of Probability and Statistics», 4, 2 (<https://doi.org/10.48550/arXiv.0811.2519>).
- MUSSINO A. 1983, *L'analisi statistico sociale di grandi matrici di dati: applicazione dell'analisi delle corrispondenze in una ricerca sociale*, «Sociologia e Ricerca Sociale», 10, 69-99.
- NAVARRO P., CINTAS C., LUCENA M., FUERTES J.M., DELRIEUX C., MOLINOS M. 2021, *Learning Feature Representation of Iberian Ceramics with Automatic Classification Models*, «Journal of Cultural Heritage», 48, 65-73 (<https://doi.org/10.1016/j.culher.2021.01.003>).
- NISHISATO S., BEH E.J., LOMBARDO R., CLAVEL J.C. 2021, *Modern Quantification Theory. Joint Graphical Display, Biplots, and Alternatives*, Singapore, Springer (<https://doi.org/10.1007/978-981-16-2470-4>).
- OLCESE G. (ed.) 1994, *Ceramica romana e archeometria: lo stato degli studi. Atti delle Giornate internazionali di Studio (Castello di Montegufoni, Firenze 1993)*, Firenze, Edizioni All'Insegna del Giglio.
- ORLANDI T. 1993, *Sulla codifica delle fonti archeologiche*, «Archeologia e Calcolatori», 4, 27-38.
- ORTON C. 1980, *Mathematics in Archaeology*, Cambridge, Cambridge University Press.
- ORTON C. 1982, *Computer Simulation Experiments to Assess the Performance of Measures of Quantity of Pottery*, «World Archaeology», 14, 1, 1-20 (<https://doi.org/10.1080/00438243.1982.9979846>).
- ORTON C. 2010, *Fit for Purpose? Archaeological Data in the 21st Century*, in K. BIRÓ (ed.), *Quantitative Methods for the Challenges in 21st Century Archaeology, Proceedings of the Annual Meeting of the UISPP 4th Commission (Budapest, Hungarian National Museum, 2009)*, «Archeologia e Calcolatori», 21, 249-260 (<https://www.archcalc.cnr.it/journal/articles/577>).
- ORTON C. 2017, *Degrees of Freedom and other Episodes in an Archaeological Life*, Woodston, Peterborough, Book Printing UK.
- ORTON C., TYERS P. 1990, *Statistical Analysis of Ceramic Assemblages*, «Archeologia e Calcolatori», 1, 81-110 (<https://www.archcalc.cnr.it/journal/articles/8>).

- ORTON C., TYERS P., VINCE A. 1993, *Pottery in Archaeology*, Cambridge, Cambridge University Press.
- PACE A., BURSICH D. 2021, *New Perspectives on Documenting Attic Pottery*, in V. BALDONI (ed.), *From Pottery to Context. Archaeology and Virtual Modelling*, «Archeologia e Calcolatori», 32.2, 83-94 (<https://doi.org/10.19282/ac.32.2.2021.08>).
- PARISOTTO S., LEONE N., SCHÖNLIEB C.B., LAUNARO A. 2022, *Unsupervised Clustering of Roman Potsherds via Variational Autoencoders*, «Journal of Archaeological Science», 142, 105598, 1-12 (<https://doi.org/10.1016/j.jas.2022.105598>).
- PEACOCK D.P.S. 1971, *Roman Amphorae in Pre-Roman Britain*, in M. JESSON, D. HILL (eds.), *The Iron Age and its Hillforts. Papers Presented to Sir Mortimer Wheeler*, Southampton, Southampton University Archaeology Society, 161-188.
- PEACOCK D.P.S. 1977, *Roman Amphorae: Typology, Fabric and Origins*, in MÉTHODES CLASSIQUES 1977, 261-278.
- PEARSON K. 1901, *On Lines and Planes of Closest Fit to Systems of Points in Space*, «The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science», 2, 559-572 (<https://doi.org/10.1080/14786440109462720>).
- PEARSON K. 1904, *On the Theory of Contingency and its Relation to Association and Normal Correlation*, London, Dulau & Co.
- PETIT-MAIRE N., LESSERTISSEUR J., NAKACHE J.P. 1972, *Étude des caractères cranio-faciaux dans l'Ordre des Primates. Apport de l'analyse factorielle des correspondances et de l'analyse discriminante*, «Bulletins et Mémoires de la Société d'anthropologie de Paris», 9, 4, 269-291 (<https://doi.org/10.3406/bmsap.1972.2054>).
- PETROGNANI S. 2010, *À propos du traitement stylistique du cheval et du bison dans l'art pariétal paléolithique anté-magdalénien franco-ibérique*, «Bulletin de la Société Préhistorique Ariège-Pyrénées», 64, 165-174.
- PICON M. 1984, *Le traitement des données d'analyse*, in T. HACKENS, M. SCHVÖRER (eds.), *Datation-caractérisation des céramiques anciennes: Cours intensif européen (Bordeaux-Talence 1981)*, Paris, Éditions du CNRS, 379-399.
- PION P. 1997, *Les habitats laténiens tardifs dans la vallée de l'Aisne. Contribution à la périodisation de la fin du second âge du Fer en Gaule nord-orientale: La Tène C2 - période augustéenne précoce, II^e-I^{er} siècle av. J.-C.*, Thèse de doctorat, Université de Paris 1 Panthéon-Sorbonne.
- PLETS G., HUIJNEN P., VAN OEVEREN B. 2021, *Excavating Archaeological Texts: Applying Digital Humanities to the Study of Archaeological Thought and Banal Nationalism*, «Journal of Field Archaeology», 46, 5, 289-302 (<https://doi.org/10.1080/00934690.2021.1899889>).
- POLLARD A.M. (ed.) 1992, *New Developments in Archaeological Science. Joint Symposium of the Royal Society and the British Academy*, Oxford, Oxford University Press.
- POLLOCK S., BERNBECK R. 2005, *Introduction*, in S. POLLOCK, R. BERNBECK (eds.), *Archaeologies of the Middle East. Critical Perspectives*, Malden, Blackwell, 1-10.
- RAGNO R. 2023, *The Spoil Project. Assessing the Rate of Excavators' Accidental Ceramic Discard at the Archaeological Site of Siponto*, «Archeologia e Calcolatori», 34.2, 69-86 (<https://doi.org/10.19282/ac.34.2.2023.04>).
- RAHTZ (ed.) 1988, *Computer and Quantitative Methods in Archaeology 1988*, BAR International Series 446, Oxford, BAR.
- RAHTZ S., RICHARDS J. (eds.) 1989, *Computer Applications and Quantitative Methods in Archaeology 1989. CAA89*, BAR International Series 548, Oxford, BAR.
- RASCHKA S. 2022, *Machine Learning with PyTorch and Scikit-Learn: Develop Machine Learning and Deep Learning Models with Python*, Birmingham, Packt Publishing Limited.
- READ D. 1994, *What Do We Need to Know to Do Quantitative Archaeology?*, in JOHNSON 1994, 1-6.
- RENFREW C. 1982, *Towards an Archaeology of Mind*, Cambridge, Cambridge University Press.
- RENFREW C. 1994, *Towards a Cognitive Archaeology*, in C. RENFREW, E.B.W. ZUBROW (eds.), *The Ancient Mind. Elements of Cognitive Archaeology*, Cambridge, Cambridge University Press, 3-12.
- RINGROSE T.J. 1988, *Correspondence Analysis as an Exploratory Technique for Stratigraphic Abundance Data*, in RUGGLES, RAHTZ 1988, 3-14.
- ROBINSON N.S. 1951, *A Method for Chronologically Ordering Archaeological Deposits*, «American Antiquity», 15, 301-313 (<https://doi.org/10.2307/276978>).

- ROSATI R., QUARTILI L., GUERMANDI M.P. 1989, *La ceramica attica nel Mediterraneo. Analisi computerizzata della diffusione. Le fasi iniziali (630-560 a.C.)*, Bologna, Cooperativa Libreria Universitaria Editrice.
- ROSENBLATT F. 1958, *The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain*, «Psychological Review», 65, 386-408 (<https://doi.org/10.1037/h0042519>).
- ROßBERGER E., KURMANGALIEV A. 2023, *Integrating Data, Creating Knowledge from Ancient West Asian Seals and Sealings: The DigANES/ACAWAI-CS Projects*, in V. BIGOT JULOUX, A. DI LUDOVICO, S. MATSKEVITCH (eds.), *The Ancient World Goes Digital. Case Studies on Archaeology, Texts, Online Publishing, Digital Archiving, and Preservation*, Digital Biblical Studies 6, Leiden-Boston, Brill, 112-136 (https://doi.org/10.1163/9789004527119_006).
- ROßBERGER E., KURMANGALIEV A., OTTO A. 2018, *Erstellung eines Digitalisierungskonzeptes: Digitalisierung und Verschlagwortung altorientalischer Roll- und Stempelsiegel-Glyptik (DigANES): Schlussbericht*, Leibniz, Informationszentrum Technik und Naturwissenschaften Universitätsbibliothek (<https://doi.org/10.2314/GBV:1029903549>).
- ROSSI I. 2020, *Qualitative and Quantitative Approaches in Digital Epigraphy*, in A. CARVALE, P. MOSCATI (eds.), *Logic and Computing. The Underlying Basis of Digital Archaeology. Proceedings of the MetroArchaeo 2019 Special Session (Florence 2019)*, «Archeologia e Calcolatori», 31.2, 145-156 (<https://doi.org/10.19282/ac.31.2.2020.14>).
- ROUANET H. 2008, *Entretien de H. Rouanet avec Philippe Bonnet sur "Benzécéri et l'Analyse des Données"*, «Journ@l Electronique d'Histoire des Probabilités et de la Statistique», 4, 2.
- ROUANET H., LE ROUX B. 1993, *Analyse des données multidimensionnelles*, Paris, Dunod.
- ROUX M. 1968, *Un algorithme pour construire une hiérarchie particulière*, Thèse de 3^e cycle, Université de Paris.
- ROUX M. 1979, *Estimation des paléoclimats d'après l'écologie de foraminifères*, «Les Cahiers de l'analyse des données», 4, 1, 61-79 (https://www.numdam.org/item/CAD_1979__4_1_61_0/).
- ROUX M., ROBERT J., JAMBU M. 1976, *Analyse de données sur l'art préhistorique*, «Les Cahiers de l'analyse des données», 1, 1, 61-70 (https://www.numdam.org/item/CAD_1976__1_1_61_0/).
- ROVA E. 1994, *Ricerche sui sigilli a cilindro vicino-orientali del periodo di Uruk/Jemdet Nasr*, *Oriens Antiqui Collectio* 20, Roma, Istituto C.A. Nallino.
- ROVA E. 1995, *Metodi di codifica e analisi per lo studio di immagini strutturate: un'applicazione ai sigilli cilindrici del periodo di Uruk/Jemdet Nasr*, «Archeologia e Calcolatori», 6, 7-34.
- ROVA E. 1996, *A Method for Analyzing the Composition Structure of Images: A Test on Near-Eastern Cylinder Seals*, in A. BIETTI, A. CAZZELLA, I. JOHNSON, A. VOORRIPS (eds.), *Commission IV Sessions at the XIII UISPP Conference. Theoretical and Methodological Problems*, vol. 1, Forlì, Abaco, 27-34.
- ROVA E. 2000, *Automatic Classification of Archaeological Materials: Possibilities and Limits. An Example (Uruk/Jamdat Nasr Cylinder Seals from Greater Mesopotamia)*, in E. ROVA (ed.), *Patavina Orientalia Selecta*, History of the Ancient Near East, Monographs 4, Padova, Sargon, 1-20.
- ROVA E. 2005, *Animali ed ibridi nel repertorio iconografico della glittica del periodo di Uruk*, in E. CINGANO, A. GHERSETTI, L. MILANO (eds.), *Animali tra zoologia, mito e letteratura nella cultura classica e orientale. Atti del Convegno (Venezia 2002)*, Quaderni del Dipartimento di Scienze dell'Antichità e del Vicino Oriente, Università Ca' Foscari di Venezia I, Padova, Sargon, 13-32.
- RUGGLES C.L.N., RAHTZ S.P.Q. (eds.) 1988, *Computer and Quantitative Methods in Archaeology 1987*, BAR International Series 393, Oxford, BAR.
- SAKAI M., SAKURAI A., LU S., OLANO J., ALBRECHT C.M., HAMANN H.F., FREITAG M. 2024, *AI-Accelerated Nazca Survey Nearly Doubles the Number of Known Figurative Geoglyphs and Sheds Light on their Purpose*, «Proceedings of the National Academy of Sciences», 121, e2407652121, 1-9 (<https://doi.org/10.1073/pnas.2407652121>).
- SALOMÉ M.-R. 1965, *Projet de code pour l'analyse des représentations figurées sur les vases grecs*, Paris, CADA.
- SALOMÉ M.-R. 1980, *Code pour l'analyse des représentations figurées sur les vases grecs*, Paris, Éditions du CNRS.
- SANTAMARIA A. 2023, *From Images to Signs: Cretan Hieroglyphic and Linear A in Context*, PhD Diss., Bologna, Alma Mater Studiorum.
- SANTORO BIANCHI S., FABBRI B. (eds.) 1997, *Il contributo delle analisi archeometriche allo studio delle ceramiche grezze e comuni. Il rapporto forma/funzione/impasto, Atti della I Giornata di Archeometria della ceramica (Bologna 1997)*, Imola, University Press Bologna.

- SAPORTA G. 2006, *Probabilités, analyse des données et statistiques*, Paris, Éditions Technip.
- SAPORTA G. 2019, *50 Years of Data Analysis: From Exploratory Data Analysis to Predictive Modeling and Machine Learning*, in C. SKIADAS, BOZEMAN J., *Data Analysis and Applications 1. Clustering and Regression, Modeling-estimating, Forecasting and Data Mining*, ISTE-Wiley, xv-xxix.
- SAUVET G., SAUVET S. 1979, *Fonction sémiologique de l'art pariétal animalier franco-cantabrique*, «Bulletin de la Société préhistorique française», 76, 10-12 (E&T), 340-354 (<https://doi.org/10.3406/bspf.1979.5161>).
- SCHILD R. 1979, *Chronostratigraphie et environnement du Paléolithique final en Pologne*, in D. DE SONNEVILLE-BORDES (ed.), *La fin des temps glaciaires en Europe. Chronostratigraphie et écologie des cultures du Paléolithique final, Actes du Colloque (Talence 1977)*, Paris, Éditions du CNRS, 799-820.
- SCHUCHMANN S. 2019, *Analyzing the Prospect of an Approaching AI Winter* (<https://doi.org/10.13140/RG.2.2.10932.91524>).
- SCOLLAR I., WELDNER B., HERZOG I. 1985, *A Portable Seriatim Package with Dynamic Memory Allocation in PASCAL*, in VOORRIPS, LOVING 1985, 149-160.
- SEMERARO G. 1990, *Note sulla distribuzione delle ceramiche di importazione greca nel Salento in età arcaica. Aspetti metodologici*, «Archeologia e Calcolatori», 1, 111-163 (<https://www.archcalc.cnr.it/journal/articles/9>).
- SEMERARO G. 2004, *Forma e funzione: osservazioni sul rapporto fra nuovi sviluppi dell'archeologia e il linguaggio descrittivo*, con *Appendice di F. Notarstefano*, in MOSCATI 2004a, 161-183 (<https://www.archcalc.cnr.it/journal/articles/378>).
- SEMERARO G. 2016, *Form, Function and Descriptive Analysis in Archaeology*, in F. DJINDJIAN, P. MOSCATI (eds.), *Dossier: Jean-Claude Gardin (1925-2015)*, «Les Nouvelles de l'archéologie», 144, 26-28 (<https://doi.org/10.4000/nda.3472>).
- SHENNAN S. 1988, *Quantifying Archaeology*, Edinburgh, Edinburgh University Press.
- SHEPARD R.N. 1962, *The Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function*, «Psychometrika», 27, 125-140 (<https://doi.org/10.1007/BF02289630>).
- SIGNORE O., MISSIKOFF O., MOSCATI P. 2005, *La gestione della conoscenza in archeologia: modelli, linguaggi e strumenti di modellazione concettuale dall'XML al Semantic Web*, «Archeologia e Calcolatori», 16, 291-319 (<https://www.archcalc.cnr.it/journal/articles/415>).
- SIMONYAN K., ZISSERMAN A. 2015, *Very Deep Convolutional Networks for Large-Scale Image Recognition*, arXiv: 1409.1556, 1-14 (<https://doi.org/10.48550/arXiv.1409.1556>).
- SLACHMUYLDER J.L. 1985, *Seriation by Correspondence Analysis for Mesolithic Assemblages*, in VOORRIPS, LOVING 1985, 137-148.
- SOHL-DICKSTEIN J., WEISS E.A., MAHESWARANATHAN N., GANGULI S. 2015, *Deep Unsupervised Learning Using Nonequilibrium Thermodynamics*, arXiv: 1503.03585, 1-18 (<https://doi.org/10.48550/arXiv.1503.03585>).
- SOKAL R.R., SNEATH P.H.A. 1963, *Principles of Numerical Taxonomy*, San Francisco-London, Freeman.
- SPAULDING A.C. 1953, *Statistical Techniques for the Discovery of Artifact Types*, «American Antiquity», 18, 305-313 (<https://doi.org/10.2307/277099>).
- STANKE G. 1993, *Image Acquisition and Image Analysis of Archaeological Fragments*, in BOARDMAN, KURTZ 1993 307-314.
- STEHLI P., ZIMMERMANN A. 1980, *Zur Analyse neolithischer Gefäßformen*, «Archaeo-Physika», 7, 147-177.
- STRECK M.P. 2006-2008, *Rawlinson, Sir Henry Creswicke*, in *Reallexikon der Assyriologie*, 11, Berlin-New York, De Gruyter, 277.
- SUN Z., LIU S., ZHANG J., CHEN K., KAUFMAN B. 2023, *Resolving the Complex Mixing History of Ancient Chinese Bronzes by Manifold Learning and a Bayesian Mixing Model*, «Journal of Archaeological Science», 151, 105728, 1-9 (<https://doi.org/10.1016/j.jas.2023.105728>).
- TENENHAUS M. 1998, *La régression PLS. Théorie et pratique*, Paris, Editions Technip.
- THURSTONE L.L. 1947, *Multiple Factor Analysis*, Chicago, University of Chicago Press.
- TOMEI F. 2024, *Least-Cost Path Analysis for the Reconstruction of the Communication Networks between Thasian Amphorae Workshops and Other Sites in the 4th-3rd Century BC*, «Archeologia e Calcolatori», 35.1, 269-284 (<https://doi.org/10.19282/ac.35.1.2024.18>).

- TOOSI A., BOTTINO A., SABOURY B., SIEGEL E., RAHMIM A. 2021, *A Brief History of AI: How to Prevent Another Winter (A Critical Review)*, «PET Clinics», 16, 449-469 (<https://doi.org/10.1016/j.cpet.2021.07.001>).
- TORGERSON W.S. 1958, *Theory and Methods of Scaling*, New York, Wiley.
- TRIGGER B.G. 1996, *Storia del pensiero archeologico*, Biblioteca di Storia 57, Firenze, La Nuova Italia.
- TRINKL E., KARL S., LENGAUER S., PREINER R., SCHRECK T. 2024, *Cross-Modal Search and Exploration of Greek Painted Pottery*, in M. HOSTETTLER, A. BUHLKE, C. DRUMMER, L. EMMENEGGER, J. REICH, C. STÄHELI (eds.), *The 3 Dimensions of Digitalised Archaeology*, Cham, Springer, 109-128 (https://doi.org/10.1007/978-3-031-53032-6_7).
- TRUST ISSUES 2023, *Trust Issues. Artificial Intelligence, Human Cognition, and Archaeology (Universität Augsburg, 2023)*, Proceedings not published (<https://www.uni-augsburg.de/de/fakultaet/philhist/professuren/kunst-und-kulturgeschichte/klassische-archaologie/forschung/trust-issues-artificial-intelligence-human-cognition-and-archaeo/>).
- UNIVERSITY OF SOUTHAMPTON 2014, *Roman Amphorae: A Digital Resource*, York, Archaeology Data Service (<https://doi.org/10.5284/1028192>).
- VAN DER MAATEN L., HINTON G. 2008, *Visualizing Data Using t-SNE*, «Journal of Machine Learning Research», 9, 2579-2605.
- VAN DER MAATEN L., POSTMA E., VAN DEN HERIK J. 2009, *Dimensionality Reduction: A Comparative Review*, «Journal of Machine Learning Research», 10, 13.
- VANDERPLAS J. 2016, *Python Data Science Handbook: Essential Tools for Working with Data*, Beijing-Boston-Farnham-Sebastopol-Tokyo, O'Reilly.
- VAN HELDEN D., MIRKES E., TYUKIN I., ALLISON P. 2022, *The Arch-I-Scan Project: Artificial Intelligence and 3D Simulation for Developing New Approaches to Roman Foodways*, «Journal of Computer Applications in Archaeology», 5, 1, 78-95 (<https://doi.org/10.5334/jca.92>).
- VAN VARK G.N., HOWELLS W.W. (eds.) 1984, *Multivariate Statistical Methods in Physical Anthropology. A Review of Recent Advances and Current Development*, Dordrecht, Reidel Publishing.
- VASWANI A., SHAZEER N., PARMAR N., USZKOREIT J., JONES L., GOMEZ A.N., KAISER L., POLOSUKHIN I. 2023, *Attention Is All You Need*, arXiv:1706.03762, 1-14 (<https://doi.org/10.48550/arXiv.1706.03762>).
- VENABLES W.N., RIPLEY B.D. 2002, *Modern Applied Statistics with S*, New York, Springer (<https://doi.org/10.1007/978-0-387-21706-2>).
- VITALI V. 1989, *Archaeometric Provenance Studies: An Expert System Approach*, «Journal of Archaeological Science», 4, 383-391 ([https://doi.org/10.1016/0305-4403\(89\)90013-7](https://doi.org/10.1016/0305-4403(89)90013-7)).
- VITALI V., GIFFORD J.A., DJINDJIAN F., RAPP G. 1992, *A Formalized Approach to Analysis of Geoarchaeological Sediment Samples: The Location of the Early Punic Harbor at Carthage, Tunisia*, «Geoarchaeology», 7, 6, 545-581 (<https://doi.org/10.1002/gea.3340070604>).
- VITALI V., LAGRANGE M.-S. 1988, *An Expert System for Provenance Determination of Archaeological Ceramics*, in RAHTZ 1988, 369-375.
- VO-PHAMHI J., LEIDWANGER J. 2020, *A New Computational Method to Quantify Morphological Standardization and Variation within Ceramic Assemblages*, «Archeologia e Calcolatori», 31.1, 55-76 (<https://doi.org/10.19282/ac.31.1.2020.03>).
- VOLLE M. 1997, *Analyse des données*, Paris, Economica.
- VOORRIPS A. 1996, *Information Science in Archaeology: A Short History and Some Recent Trends*, in MOSCATI 1996a, 303-312 (<https://www.archcalc.cnr.it/journal/articles/146>).
- VOORRIPS A., LOVING S.H. (eds.) 1985, *To Pattern of the Past*, «PACT» 11.
- VUAILLAT D., MASSONIE J.-P. 1974, *Essai d'une typologie des haches à ailerons médians par une méthode mathématique*, «Dialogues d'Histoire Ancienne», 1, 29-49 (<https://doi.org/10.3406/dha.1974.1363>).
- WAYENBERG J., CAPRIOTTI M. 2014, *Investigating Greek Painted Iconology by Using Artificial Neural Networks. Maenads and Satyrs on Athenian Red-Figure Pottery*, in M. RAMAZZOTTI (ed.), *ARCHEOSEMA. Artificial Adaptive Systems for the Analysis of Complex Phenomena. Collected Papers in Honour of David Leonard Clarke*, «Archeologia e Calcolatori», Suppl. 6, 263-290 (<https://www.archcalc.cnr.it/supplements/articles/764>).
- WHALLON R. 1972, *The Computer in Archaeology: A Critical Survey*, «Computers and the Humanities», 7, 1, 29-45 (<https://doi.org/10.1007/BF02403759>).

- WHALLON R. 1984, *Unconstrained Clustering for the Analysis of Spatial Distributions in Archaeology*, in H. HIETALA (ed.), *Intrasite Spatial Analysis in Archaeology*, Cambridge, Cambridge University Press, 242-277.
- WICKHAM H., GROLEMUND G. 2016, *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*, Sebastopol, O'Reilly, CA.
- WIDEMANN F., LAUBENHEIMER F., ATTAS M., FONTES P., GRUEL K., LEBLANC I., LLERES J. 1979, *Analytical and Typological Study of Gallo-Roman Workshops Producing Amphorae in the Area of Narbonne*, «Archaeo-Physika», 10, 317-341.
- WILCOCK J.D., SHENNAN S.J. 1975, *Shape and Style Variation in Central German Bell Beakers. A Computer Assisted Study*, «Science and Archaeology», 15, 17-31.
- WILKINSON E.M. 1971, *Archaeological Seriation and the Travelling Salesman Problem*, in HODSON, KENDALL, TAUTU 1971, 276-283.
- WITTEN I.H., FRANK E., HALL M.A.A. 2011, *Data Mining: Practical Machine Learning Tools and Techniques*, Burlington, MA, Morgan Kaufmann (3rd ed.).

In line with the objectives of the Series, this volume explores the fifty-year evolution of *Analyse des Données* over the past fifty years, starting with Jean-Paul Benzécri's seminal 1973 work, which laid its theoretical and methodological foundations. Central to this journey is the transformative impact of data analysis on archaeology and the strategic importance of computers for processing vast amounts of data simultaneously. Benzécri's data-driven approach emphasized pattern discovery through exploratory data analysis and anticipated the rise of Data Science. The book features contributions from scholars directly involved in this evolution: François Djindjian highlights the role of the French school and its influence on prehistoric archaeology; Paola Moscati examines the reception and dissemination of multivariate analysis techniques, with a particular focus on their application in classical archaeology; Alessandro Di Ludovico and Lorenzo Cardarelli address contemporary topics rooted in this tradition, including Textual Data Analysis and the shift toward Data Science. A comprehensive bibliography closes the volume, providing a valuable resource for future research.

François Djindjian
Paola Moscati

ANALYSE DES DONNÉES AND ARCHAEOLOGY FIFTY
YEARS LATER: FROM DATA ANALYSIS TO DATA SCIENCE

€ 34,00

ISSN 1723-4565

ISBN 978-88-9285-374-4

e-ISBN 978-88-9285-375-1



30 anni di informatica archeologica – 3

ANALYSE DES DONNÉES AND ARCHAEOLOGY FIFTY YEARS LATER: FROM DATA ANALYSIS TO DATA SCIENCE

François Djindjian and Paola Moscati

with

Lorenzo Cardarelli and Alessandro Di Ludovico

