ARTICLE

Check for updates

A fragment-based approach for computing the long-term visual evolution of historical maps

Remi Petitpierre [1✉], Johannes H. Uhl[2,3], Isabella di Lenardo[1] & Frédéric Kaplan[1]

Cartography, as a strategic technology, is a historical marker. Maps are tightly connected to the cultural construction of the environment. The increasing availability of digital collections of historical map images provides an unprecedented opportunity to study large map corpora. Corpus linguistics has led to significant advances in understanding how languages change. Research on large map corpora could in turn significantly contribute to understanding cultural and historical changes. We develop a methodology for cartographic stylometry, with an approach inspired by structuralist linguistics, considering maps as visual language systems. As a case study, we focus on a corpus of 10,000 French and Swiss maps, published between 1600 and 1950. Our method is based on the fragmentation of the map image into elementary map units. A fully interpretable feature representation of these units is computed by contrasting maps from different, coherent cartographic series, based on a set of candidate visual features (texture, morphology, graphical load). The resulting representation effectively distinguishes between map series, enabling the elementary units to be grouped into types, whose distribution can be examined over 350 years. The results show that the analyzed maps underwent a steady abstraction process during the 17th and 18th centuries. The 19th century brought a lasting scission between small- and large-scale maps. Macroscopic trends are also highlighted, such as a surge in the production of fine lines, and an increase in map load, that reveal cultural fashion processes and shifts in mapping practices. This initial research demonstrates how cartographic stylometry can be used for exploratory research on visual languages and cultural evolution in large map corpora, opening an effective dialogue with the history of cartography. It also deepens the understanding of cartography by revealing macroscopic phenomena over the long term.

[1] Digital Humanities Institute, EPFL, Lausanne, Switzerland. [2] Institute of Behavioral Science, University of Colorado, Boulder, CO, USA. [3] European Commission, Joint Research Centre, Ispra, Italy. ✉email: remi.petitpierre@epfl.ch

Introduction

Why study maps? Maps have played and continue to play a predominant role in modern Western society. The development of cartographic technology spans thousands of years and a considerable evolution separates Ptolemy's Cosmography from the Map of Imola by da Vinci, and contemporary geographic information systems. The map is a strategic administrative and military device, and a crucial asset for exploration, and trade (Hofmann and Tesson, 2012). It is a historical marker that has evolved at the heart of technological innovations in strategic fields, such as astronomy, geometry, printing, geodetic measurement systems, optics, but also chemistry (e.g. synthetic inks), aerospace technology, computer and information science. Many of these innovations directly impact the physical cartographic object or the representation. The art of cartography has also reflected aesthetic considerations throughout the ages.

The map and the landscape are intertwined through various apprehension mechanisms. In the very first degree, maps usually aim to represent a geographical space, in a regulated manner. In this perspective, historical cartography is often used as a source for spatial and environmental history (Liu et al., 2022; Mazier et al., 2015). However, maps are not mere projections of the territory, although they have a similar structure (Korzybski, 1994: 58). In the 1980s, the perspective that the map is also a socially constructed discourse emerged (Harley, 1989). Influenced by the thought of Foucault, Harley sees the map as a power-knowledge practice. This idea is still very popular today (Crampton, 2001; Crampton and Elden, 2007), to the point that some scholars consider that sovereign states as we know them could not have emerged without cartography (Branch, 2014; Schulten, 2012).

The map has also been used as a planning device. As such, it is not only a record of planned interventions, but also entails reflection and theory (Winther, 2020). Seeing a city on a map has contributed to the formation of aesthetic and functional schools of thought, such as hygienism (Geels, 2006). The map also directly influences the reader's use of their environment, through the identification to landmarks (Keil et al., 2020), and orientation. Finally, as an intentional representation rather than a literal projection, the map reflects the cultural apprehension of the environment, through a series of figurative choices (Bertin, 1967), such as highlighting certain monuments, and political boundaries, or the graphical differentiation of particular agricultural surfaces, such as vineyards. Maps can have a profound imaginary or evocative power; some can make use of allegories, e.g., to depict the perceived impetuosity of the sea or the consideration of the rustic charm of the wilderness.

These multiple and complementary perspectives make maps a unique and transversal object of study. As such, archives and libraries have digitized millions of maps, many of which are available in open access online. Yet most works on the history of cartography continue to focus on a few eminent maps or great cartographic series. The spirit of the Annales school and the *Nouvelle Histoire* is struggling to penetrate the field. Through the development of new computational methods, we hope to make it possible to excavate and study the myriad maps that make up the cultural fabric, between the famous instances. The aim of this research is to propose a methodology to identify characteristic formal features in a corpus of maps. The digitization of cartographic collections paves the way for quantitative and corpus approaches. Machine reading techniques arose indeed, proposing to vectorize historical maps automatically and use them for the diachronic study of the territory (Ares Oliveira et al., 2019; Heitzler and Hurni, 2020; Uhl et al., 2022). The possibility of studying large map corpora, or to develop a 'distant reading' (Moretti, 2013) or rather a 'distant viewing' (Arnold and Tilton, 2019) of cartography, however, remains largely unexplored. This article precisely aims to address this gap by developing a methodology for cartographic stylometry. In this context, stylometry refers to the study and analysis of the distinctive visual appearance (or style) of maps. It includes aspects like line work, or texture, that differentiate one map or series of maps from another.

Maps as visual language systems. Corpus linguistics has led to important advances in understanding how languages change over time (Davies, 2012; Hardie and McEnery, 2011; Hilpert and Gries, 2016). However, expanding this approach to other visual language systems could significantly contribute to understanding human expression in a larger sense. In this respect, we believe that cartography is a particularly promising study case. "The map is a written language" (Goode, 1927: 1). More precisely, map-making practices could be considered transforms of natural languages (Jakobson, 1973: 29). In a distributionalist view, maps could then be decomposed into practical analysis units or etic units. In this context, visual computing could help to operationalize pertinent graphical features and capture the systemic organization of these map units. In fact, decomposing the map into smaller tiles is already effectively used by engineers and geographers, for instance to extract and visualize semantic classes (Hosseini et al., 2021, 2022; Petitpierre et al., 2021; Uhl et al., 2018, 2020). (Hiippala and Bateman, 2022) use comparable fragments, or visual tokens, to deploy a distant viewing approach and investigate semiotic modes in diagrams.

More qualitatively, some researchers have attempted to catalog the icons and symbols used in certain maps, such as François de Dainville in his *Langage des géographes* (Dainville et al., 2018). Such elements would be comparable to emic units, or cultural elementary units. However, the resulting listing remains rather limited and subjective, and the approach is hardly scalable. If we are to develop a scholarship of cartographic stylometry, we deem it preferable to already ground it in core theoretical concepts of semiotics, such as the notion of pertinence. Even if a digital methodology on map semiosis is not yet proposed in the scope of this article, we consider it a worthy and natural next step (Casti, 2005).

The system emerging from contrasts. In this context, "pertinence refers to the feature of distinctiveness of structures within a system" (Nöth, 1995: 300). In this perspective, a difference in form, at the lower etic level, can translate into a systemic distinction if a difference is also observed in a higher cultural plane. This is also the intuitive idea underlying contrastive representation learning. Contrastive representation learning is a machine learning strategy used to establish a feature representation, on the base of a set of contrasted sample pairs. Concretely, an unannotated set of images could be used to train a neural network, based on simple logical pairs, such as $Image_A \cong Image_B$ (e.g. they are replicas), and $Image_A \ncong Image_C$. The representation is then optimized to maximize the distance between negative pairs and minimize it between positive ones.

Contrastive representation learning is, therefore, an engaging solution for learning feature representation when one can define a set of opposed samples (e.g. cat vs dog). When it comes to representation, however, it is often uncomfortable to prescribe opposition, in part because of the non-trivial issue to define kinship or replication (Eco, 1992). This has also been acknowledged as one of the main limitations of structuralist approaches. To overcome this obstacle, we propose a mathematical strategy that makes it possible to learn the feature representation of low-level graphical cartographic units, or 'map elements', which we

will define in a few paragraphs, by contrasting higher-level structures, such as map ensembles or series.

One characteristic aspect of cartography is indeed the existence of series. Cartographic series form coherent editorial ensembles and can account for cultural differences, in terms of style, and production practices (tools, techniques, materials). In this regard, cartographic series constitute relevant ensembles, as they exhibit both internal coherence and external contrast. Moreover, contrasting higher-level structures (map series) limits the impact of the subject matter (e.g., geographic space) that could prevail if we were to directly contrast maps or parts of maps. With the approach of joint contrastive representation learning we propose however, the risks of occluding cultural factors are hopefully more limited, and the assumption that the presence of resemblant elementary units in two maps bespeaks stylistic similarity becomes affordable, as it will be discussed in the next section.

Organization of the article. To summarize, in this article, we develop a methodology for cartographic stylometry. In particular, we seek to investigate the relationship between figurative proximity, on the one hand, and time and scale, on the other hand. We tackle several methodological challenges, such as: how to characterize cartographic figuration with computational methods? How to measure the figurative distance between a pair of maps? The results are then discussed in the framework of the historical writing on cartography. In the context of this paper, we use the term 'figuration' to refer to the graphical configuration of features on the map image rather than in a sense of representation or meaning-making.

The structure of the article is as follows. In the next section, we present the corpus, and the proposed approach, step by step. For reasons of clarity of the narration, we favor the explanation of the methodological reasoning over the exposition of the technical implementation details. The technical details on the algorithms and method are duly presented in a separate section, at the very end of the paper, after the conclusion. In the "Results and discussion", we present and discuss the results of the stylometric approach, on a case study focusing on French and Swiss maps. In the "General Discussion and Conclusion", we open the discussion in a more general way and offer some concluding remarks.

Approach

Constitution of the dataset. The corpus we study consists of 10,046 digitized French and Swiss maps available in open access, from the collections of 32 different institutions, the main ones being the Bibliothèque nationale de France (4,817) and the University Library of Zurich (1,452). The selection criteria are focused on the study period (1600–1950), the target scale (1:200 to 1:53,500), and the geographical coverage. The corpus is restricted to European France (75%), and Switzerland (25%). To avoid confusion on the cultural context, maps of France present in Swiss collections are excluded, and vice versa. Apart from these, all digitized cartographic documents made available online by any of the 32 institutions at the time of data collection are considered. When the scale is not indicated in the record, but can be inferred from the map itself, an estimate is made by a human expert. Moreover, for large map series (e.g. cadasters, national topographic maps, atlases), a sampling of 2% to 5% of the corpus (from smaller to larger scale maps) is performed, with a minimum sample of 5 maps per series. According to the metadata, the resulting corpus comprises more than 1800 distinct publishers, and 2700 creators. The complete table of maps and associated metadata is published along with this article Fig. 1.

The maps are geocoded using coordinates harvested from the metadata, when available, or by recognizing the geographic named entities present in the title or description of the map, using natural language processing tools (spaCy: Industrial-strength NLP, 2022). The Nominatim API (Nominatim (2022)) is used to geocode the entities (i.e., to obtain geographic coordinates from placenames); the results are manually checked by a human expert.

Within this corpus, we identify 11 coherent map series, i.e., maps belonging to a series that constitutes an editorial ensemble, and stem from a same production context. These 11 map series should exhibit a variety of scale, time, topic, and geographic subject, as they will be foundational to the operationalization of cartographic figuration. For this purpose, we select three series of cadastral maps of Lausanne: the Ancien Régime cadaster by Melotte (1721–27), the 'Napoleonic' cadaster by Berney (1827–31), and the renovated cadaster by Deluz (1880–86), as well as three series of Napoleonic cadastral maps of the Rhône (1808–68), Côtes d'Armor (1800–66), and Haute-Marne (1800–50) departments. We also include two series of Parisian city atlases, by Jacoubet (1825–40) and Alphand (1860–89). Additionally, we select three topographic series: the 1:40,000 maps of the Etat Major (1831–65), the 1:50,000 national maps of the Institut Géographique National (IGN) (here 1926–63), and the Swiss Federal Topographical Surveys (Swisstopo), starting with the Siegfried maps (1870–1949). These 11 series represent 1063 maps, i.e. around 10% of the corpus (details in Supplementary Material).

Fragmentation. Recent advances in natural language processing have profoundly challenged previous approaches in linguistics to modeling grammar and syntax. Pragmatically, the problem of operationalizing the syntax itself seems to be solved. One of the methodological paradigms that made this possible is the fragmentation of texts into tokens. The second is the statistical association of tokens in texts. In this context, tokens are not even words. They are units consisting of about 4 characters or circa 3/4 of a word. The process of fragmentation of the text into tokens makes it possible to create convenient units, that can be easily represented in a digital space, and manipulated with computational approaches. Interestingly –and one may say counter-intuitively– this fragmented representation does not alter the machine's ability to learn the syntax and grammar of a language, quite the contrary.

Drawing inspiration from 'tokenization', our methodology is based on the 'fragmentation' of the digitized map image into tiles, or elementary map units (mapels). In our perspective, a 'mapel' is an etic unit, analogous to a 'phone' for instance.[1] Mapels are considered physical events, specific instances, cartographic utterances, and the realization of a mapping process. In terms of computer vision, when studying mapels, we study the graphical configuration of map extracts.

Concretely, 800 mapels are cut out automatically from each map in the corpus.[2] Each unit has a size of 50×50 pixels, which permits to cover a minute space of the map.[3] Together, they constitute a summary of the map's figuration. Non-geographic content of the map, as well as areas of the map that are devoid of graphical content, or exhibit a particularly low graphical load, are excluded.[4] Except for these areas, the initial location of the mapels is first drawn randomly, and each one is recentered on the most salient graphical feature (e.g. line, point, icon) close to the initial location. The orientation of the local graphical features is computed, saved, and a rotation is applied to the mapel, so that the graphical content of each unit has a neutral orientation. The computed orientation is stored separately, as an explicit variable attached to the mapel.[5]

The stylometric space of cartography. The prevalent methods for image embedding –to create a digital representation of
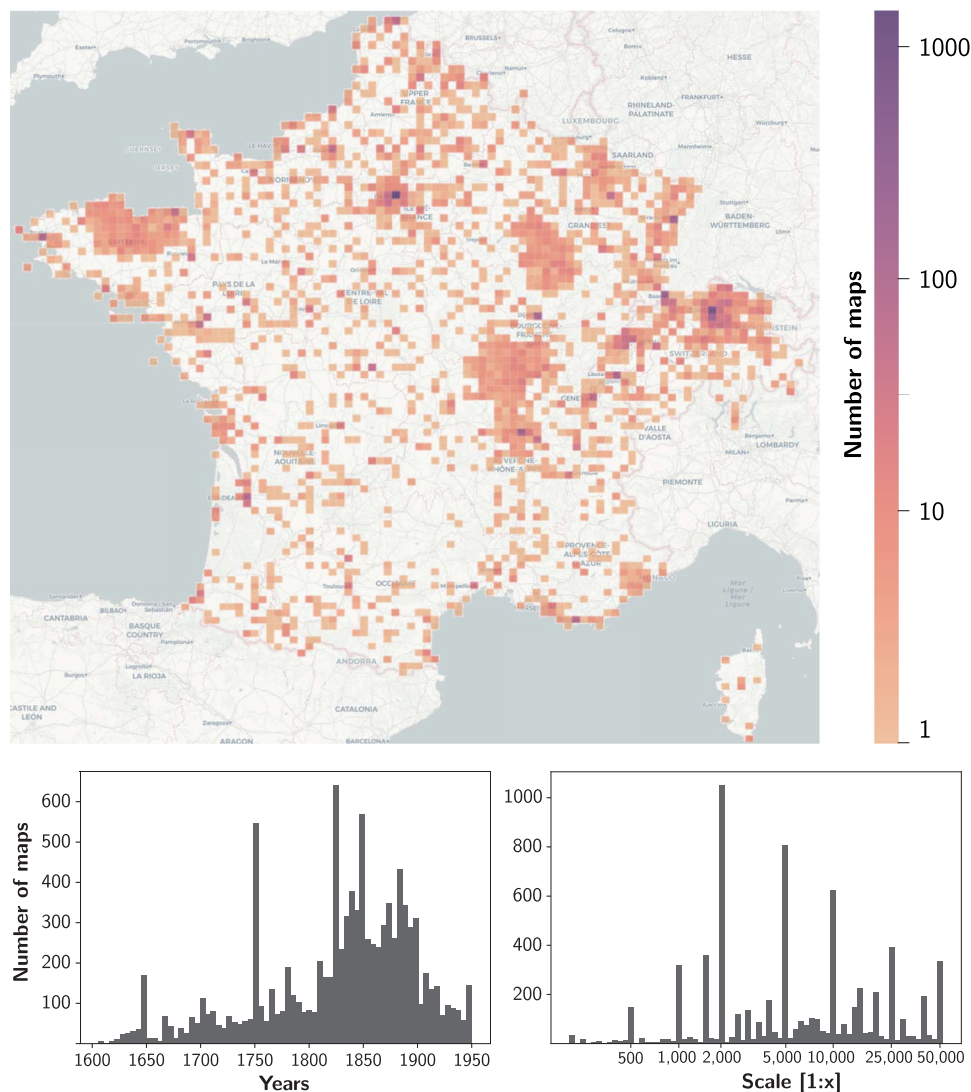
Fig. 1 Descriptive statistics on the final corpus by coverage, publication year, and map scale.

graphical objects– are based on neural networks. However, these methods have inherent limitations, such as producing opaque and non-interpretable features. Additionally, neural networks operate like black boxes, making them incompatible with hermeneutic requirements. Instead, we employ a distinct method, starting with a set of 13 candidate features, based on the literature in digital cartography analysis and computer vision (Barvir and Vozenilek, 2020; Dalal and Triggs, 2005; Herold, 2015: 64–67; Ojala et al., 2002; Otsu, 1979; Scharr, 2000; Uhl et al., 2018). These features cover explicit and complementary aspects of cartographic representation: color (5 candidate features), morphology (1), texture (1), graphical load (3), line thickness (1), and orientation (2). In this study, morphology refers to the form of the visual elements in maps, the shape and the arrangement of graphical features. Color, as a perceptual variable, is a difficult concept to operationalize when it comes to digital images. It is it is commonly encoded as red-green-blue pixel values. As we will discuss, this digital encoding is reductive and impedes the use of color as a pertinent visual feature. Texture refers to the arrangement of pixels in patterns, such as hatchings, grid, or plain patterns. Graphical load measures the density of graphical content in the image. Finally, orientation is computed based on the local dominant orientation of edges.

Methodologically, it is indeed not desirable to impose the features of the problem, since studying these features might

precisely be of interest for the study of cartography. This is why we set up an optimization process to select the most suitable candidates to build a meaningful feature space. To do so, we rely on the principle of distinction between map series, which are coherent sets of maps considered homogeneous cultural, figurative, and editorial ensembles. To construct the digital space of cartographic figuration, we rely on the idea that a conceptual space is constructed by the contrasts and oppositions between the concepts themselves. This is a foundational notion in semiotics. "A semiotic system functions adequately as long as its elements and structures remain sufficiently distinct or differentiated. The semiotic principle which guarantees this functional differentiation is called the principle of pertinence […], pertinence is the property by which a speech sound (a *phone*) is structurally opposed and thus distinct from another speech sound of the same system." (Nöth, 1995: 183). In this research, we base the optimization of the feature space on the conceptually simple idea that the figuration of two maps that belong to the same series is more similar than when the maps belong to distinct map series.

Our research hypotheses postulate that, after optimization, the feature space will enable distinction between map series that represent the same geographical location (1), and those of similar scale (2). For instance, different series representing Paris at a similar scale (e.g. city atlases) should be distinguishable, as they

exhibit stylistic differences. However, we expect the Napoleonic cadastral map series to be indistinguishable from one another due to their identical production process and representation conventions (Hennet, 1811). If these hypotheses hold true, the selected features will be demonstrated as independent from subject, and scale, thereby maximizing their potential to describe cultural and stylistic variables.

Each map is represented by a set of 800 elementary map units (mapels), with each unit tentatively represented by 13 candidate features. In an optimal feature space, mapels of a particular map are close to similar mapels found in maps from the same series. To measure this proximity, we introduce a free parameter k and assume that if the distance d between two mapels m_i, m_j is less than k, the two units can be considered close, or analogous (Eq. 1).

$$(d(m_i, m_j) < k) \Rightarrow (m_i \sim m_j) \tag{1}$$

In this context, k can be defined as the radius of free variation. The variation in the features, within this radius, are so small that they do not disrupt the relation of analogy between two mapels.

Next, the stylometric distance D between a map A and a map B can be defined as the proportion of mapels of A that are analogous to at least one mapel of B, plus the proportion of mapels of B that are analogous to at least one mapel of A (Eq. 2).[6]

$$D(A, B) = \left[\sum_{m_i \in A} \left(\min_{m_j \in B} d(m_i, m_j) \right) < k \right] + \left[\sum_{m_i \in B} \left(\min_{m_j \in A} d(m_i, m_j) \right) < k \right] \tag{2}$$

Hence, when maps A and B belong to the same series, many mapels should be shared by both maps. The stylometric distance D represents the proportion of mapels common to both maps.

Furthermore, we make use of information on map series to which individual maps belong, to identify an *optimal feature space*. This problem represents the inverse of a clustering problem. In this case, the clusters are known (map series), and we are trying to find the features space that best matches this data and allows us to distinguish the clusters. Translated into mathematical terms, the optimization problem requires maximizing the median distances $\mu_D(S_a, S_b)_{a \neq b}$ between maps of different series S, and minimizing the median distances $\mu_D(S_a, S_b)_{a=b}$ between maps of the same series. To keep the space structured and homogeneous, the optimization is constrained by minimizing the deviations of distances within the same series $\sigma_D(S_a, S_b)_{a=b}$, as well as the deviations of distances between two different series $\sigma_D(S_a, S_b)_{a \neq b}$. The aim is to avoid the creation of a sparse, inhomogeneous space, or the impact of

outstanding maps. For the mean noted $\bar{\cdot}$:

$$\max \frac{\overline{\mu_D(S_a, S_b)_{a \neq b}} - \overline{\mu_D(S_a, S_b)_{a=b}}}{\overline{\sigma_D(S_a, S_b)_{a \neq b}} + \overline{\sigma_D(S_a, S_b)_{a=b}}} \tag{3}$$

To solve this optimization problem, the feature space can be deformed, either by adding and removing some of the 13 candidate features, by modifying their relative weight, or their internal parameters. The exploration of these different combinations is performed using a genetic algorithm. Genetic algorithms are a powerful family of algorithms used to find the optimal combination of parameters in a problem. They mimic an evolutionary mechanism, with several iterations called 'generations', where the best performing parameters are kept from one generation to the next. At the end of the evolutionary optimization process, the best combination of parameters is retained. The radius of free variation k is dynamically optimized for each solution, using a line search.

At this stage, the feature space enables us to efficiently contrast the cartographic representation of the various series. We will discuss this aspect in more depth in the section "Exploration of the stylometric map space" of the results, but we can already indicate that the mathematical optimization allows us to select 6 features among the 13 candidates, and to assign them a weight between 0.05 and 1 (details in Table 1). These features thus compose a 6-dimensional space, where some dimensions have a greater impact on the distance between two mapels. In this space, the distance d between two mapels is indicative of the figurative proximity between them. Moreover, the distance is no longer derived from an arbitrary set of graphical metrics, but directly from the notion of distinction between map series. Mapels that are separated by a distance smaller than k, the radius of free variation, are analogous (Eq. 1).

Interpretation of the feature space. In this respect, mapels lying in the same hypersphere of radius k can be considered variants of the same canonical form. However, this canonic form does not constitute an emic unit (a "mapeme") in the sense of linguistics. An "emic unit, such as a phoneme or morpheme, is an invariant form obtained from the reduction of a class of variant forms to a limited number of abstract units" (Nöth, 1995: 183). The distinction between emic units should also reflect the possibility to alter meaning. In the present case, this second necessary condition is not demonstrated. In this regard, the nature of the relationship between mapels and their canonical form is rather that of a token to its type. Therefore, we prefer to call these canonical units 'mapotype'. A mapotype, in this context, is the archetype

Table 1 Candidate features, and their resulting weight after optimization.

Method type	Method	Interpretation	Weight
Color distribution	RGB Histogram	Distribution of color intensities across the red-green-blue channels.	–
	RGB Peak color	Most frequent color intensity in each channel.	–
	RGB Standard deviation	Dispersion of color intensities.	–
	RGB Skew	Asymmetry of color intensity distribution.	–
	RGB Kurtosis	Acuteness of color intensity distribution.	–
Morphology	HOG	Histogram of local edge orientations, relatively to the dominant edge orientation.	1.
Texture	LBP	Texture patterns (e.g. hatched, grid, plain, etc.), based on the local arrangement of pixels.	0.1
Graphical Load	Otsu's dark pixel proportion	Density of pixels with dark content.	–
	ED2	Square root of the density of edges.	0.15
	Num Connected Components	Number of graphically connected features in the image.	0.25
Line width	Skeleton ratio	Content ratio before and after thinning the mapel to single pixel width (skeletonization).	0.25
Orientation	HOG orientation	Dominant edge orientation of the mapel, before its neutralization.	0.05
	Binned HOG pattern	Regroup the edge orientations into discrete bins*.	–

*vertical (±π), horizontal (±π/2, ±3π/2), diagonal (±π/4, ±3π/4), regular oblique (±π/6, ±2π/6, ±4π/6, ±5π/6), and irregular oblique (all other orientations).
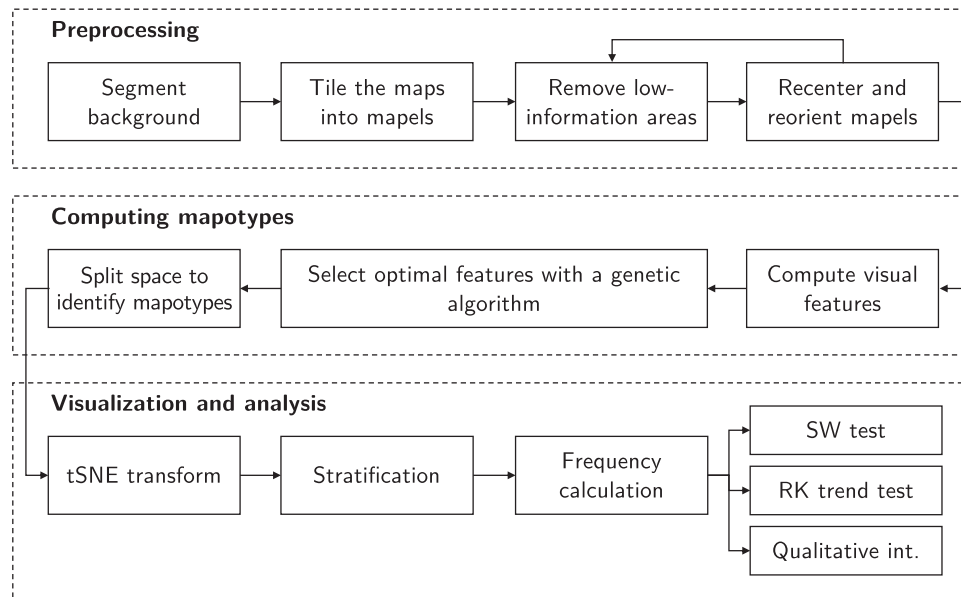
Fig. 2 Flow-chart summarizing the method.

obtained from the reduction of a class of variant cartographic utterances. Albeit we can intuit that mapotypes are cultural units, because of the contrastive process they stem from, what distinguishes them from emic units seems to be the scale at which differentiation operates. While emic units are determined by their direct functional role, such as their capacity to alter meaning, mapotypes are determined by their joint ability to reflect cultural contrasts. In this context, for instance, mapotypes make it possible to operationalize contrasts between cartographic series.

We use the concept of mapotype to construct meaningful and pragmatical observation units for the study of the evolution of cartographic figuration. To do so, we iteratively subdivide the feature space, until we obtain cells of radius $r < k$. These cells represent our functional units of analysis, or mapotypes.

Analysis. For visualization purposes, the mapel closest to the geometric center of the mapotype is used to represent the latter in a geometric space. To facilitate visualization in 2D, we reproject the mapotypes using t-stochastic neighbor embedding (t-SNE, Maaten and Hinton, 2008), a nonlinear dimensionality reduction method that preserves neighborhood relations when possible. The projected mapotypes are then constrained on a grid (Klingemann, 2020). This visualization we call 'mapotypic mosaic' constitutes the basis for most of the visual-analytical results presented in Section "Results and discussion". Our approach to the analysis of the results is inspired by a distributionalist-variationist perspective. We operate a uniform stratification (temporal, scalar) of the corpus and observe the relative distribution of mapotypes by stratum. The Fig. 2 summarizes the main methods deployed for the analysis of the results, as well as the steps of the methodology described in this section.

Results and discussion
Proof of concept. As discussed in Section "The stylometric space of cartography", the optimization process has the objective of maximizing contrasts, thus distance, between cartographic series. This process yields an optimum of 3.59 for the Eq. 3, and the resulting distances are presented in Fig. 3. The stylometric distance clearly differentiates between the maps belonging to the same series, on the diagonal axis, and those belonging to distinct series. The only exception to this trend is observed for the

Napoleonic cadasters, and this result verifies our research hypotheses (see Sections "The stylometric space of cartography" and "Interpretation of the feature space"). Moreover, maps representing the same geographical subject (Paris, Lausanne), but belonging to different series are correctly distinguished. The metric is also robust at various scales and map types, from cadastral plans to national topographic maps, and city atlases.

Until now, the notions of elementary map unit (mapel) and mapotype may have seemed rather theoretical and abstract. We have stated that replication is difficult to circumscribe when it comes to representation, as each instance is somehow unique. In Fig. 4, we reveal how mapels look like, and how much variation is found within a mapotype. As will be discussed in the next section, the color itself is not a feature of the stylometric map space. Apart from that, we note that mapotypes generally capture resemblance, although a certain amount of free variation remains. This seems particularly true for most complex and cluttered mapels. Overall, as a complement to the statistical validation offered by Fig. 2, this outcome is sufficiently consistent to validate a distributional approach.

Exploration of the stylometric map space. The features selected to construct the stylometric map space, after optimization, are listed, with their relative weight and possible interpretation, in Table 1. In general, morphology seems to be the most pertinent feature (weight = 1.0), far ahead of features related to graphical load (0.25, 0.15), line thickness (0.25), or texture (0.1). The importance of the orientation seems minimal (0.05), as expected. No feature related to the distribution of colors was retained.

In total, our methodology yields 55,599 mapotypes. Only 0.1% have less than 5 occurrences in the corpus and more than half of them (53%) have more than a 100 and account for 85% of the mapels in the corpus (see Supplementary material). An overview of all the mapotypes observed is presented in the mapotypic mosaic (Fig. 5). Note that this space is not globally continuous. The t-SNE algorithm used for visualization tends to preserve neighborhood relationships at the expense of the linearity of the dimensional reduction. Overall, we can notice from Fig. 5 that the method is highly sensitive to different kinds of lines (contrasted, dark, thick, colored, interrupted, or continuous, e.g. A3, A4, B4), textures (hatchings, dotted patterns, grids, e.g.
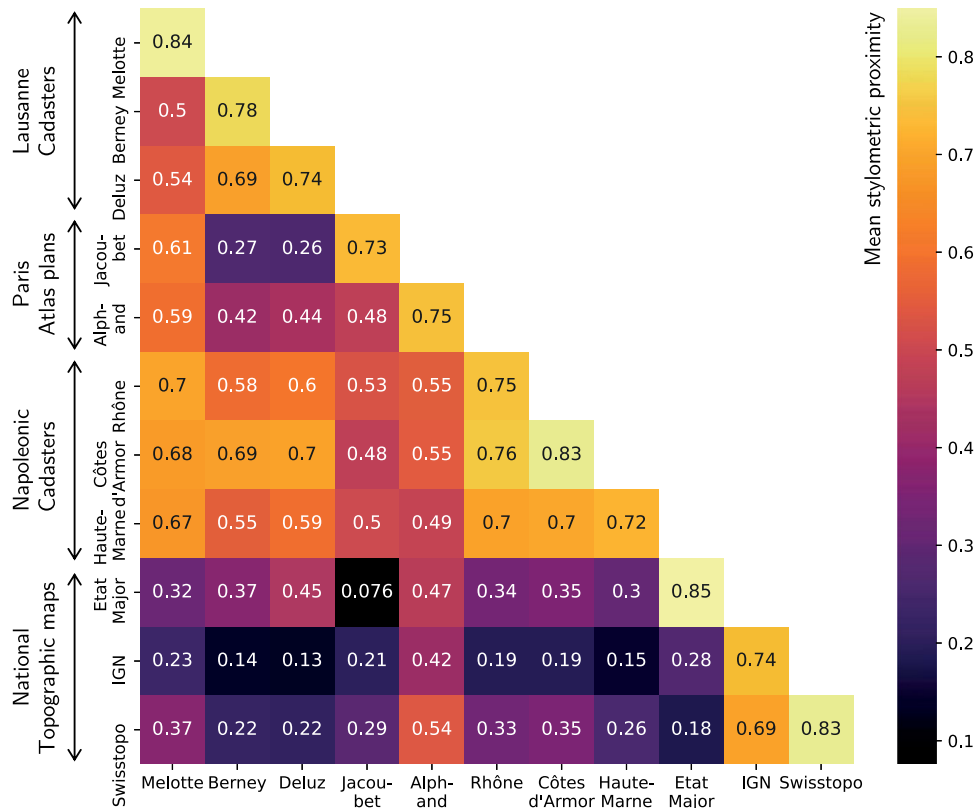
Fig. 3 Matrix of average stylometric proximity between the maps in the selected cartographic series.

A1, B3), as well as other graphical elements (texts, colored gradients, e.g. C4).

The fact that no descriptive feature of color distribution is retained suggests that contrary to our intuition, color might not be a very strong explanatory cue to operationalize stylometric proximity. On one hand, this intriguing result may indicate that color is not specific enough. Indeed, similar shades can be found in various unrelated maps and contexts, which is also true for some texture types. On the other hand, the identity of color is only a mental concept derived from the Gestalt principle of similarity. Recent psychophysical and cognitive research suggest that color saliency, contrasts, as well as high-level knowledge might be more defining factors of perceived identity or constancy of color (Maule et al., 2023; Tajima and Komine, 2015). Especially for hand-colored maps, the physical color as measured by the camera can be quite different from one area of the map to another or from one map to another, within the same series. This is true even when color is perceived as identical, hence the difficulty of applying color-based segmentation methods to cartographic documents, as experienced in previous work (Leyk and Boesch, 2009; Liu et al., 2016). However, despite the absence of features specifically encoding color, it appears that the feature space differentiates between colored and uncolored features (e.g. Figures 5, C4). This is interesting because it could indicate that more than the shade itself, it is the fact of being colored or not that is decisive in the characterization of the cartographic representation.

Due to the optimization methodology used to select features, the occupation of the space itself can be interpreted as a marker for determining which graphical elements are most important for the analysis of figuration. Thus, the fact that a large proportion of the space is simply occupied by different kinds of lines indicates that these minute differences in linework may play an important role in characterizing cartographic representation, which is

consistent with the literature. For instance, line drawing is one of the main cues that allow experts to distinguish different techniques (e.g. engraving, etching, lithography) or tools used to create the map (Gascoigne, 1986).

Cartography in the Conjuncture. The previous results demonstrated the potential of our approach to distinguish between map series and provided a first insight into mapotypes. We now seek to use them as tools to dive into the evolution of cartography over periods of a few decades each, in a medium-term timeframe close to what Fernand Braudel referred to as the *conjuncture* (Braudel, 1996; Santamaria and Bailey, 1984). Fig. 6 is based on the projection used for creating the mapotypic mosaic (Fig. 5) and permits to visualize the relative distribution of each mapotype, within time, and scale-based strata. Tab 2. presents the W statistic ($n = 55{,}599$) of the Shapiro-Wilk normality test (henceforth called SW test), as adapted by (Royston, 1995), applied to the relative distribution of mapotypes per time and scale stratum. The SW test is considered the most powerful normality test (Razali and Wah, 2011). Lower W statistics indicate the occurrence of extremely low or high-frequent mapotypes (i.e., departure from normality in the relative frequency distribution), thus a more characteristic figuration in a particular stratum.[8] Put simply, a lower value indicates that the distribution for this stratum is not uniform and thus some types are clearly over- or under-represented.

1600–1817: From geographic iconography to the symbolic map. We can note in Fig. 6 that cartography occupies a relatively stable stylometric subspace from 1600 to 1817. In the first time stratum (1605–1730), mapotypes in the lower right (B3–4, C4, D1–4) are dominating. These brighter areas correspond to figurations strongly characteristic of this period. Gradually, from 1730 onwards, the occupation of the feature space becomes more diffuse, with fewer bright areas. This dilution of the characteristic
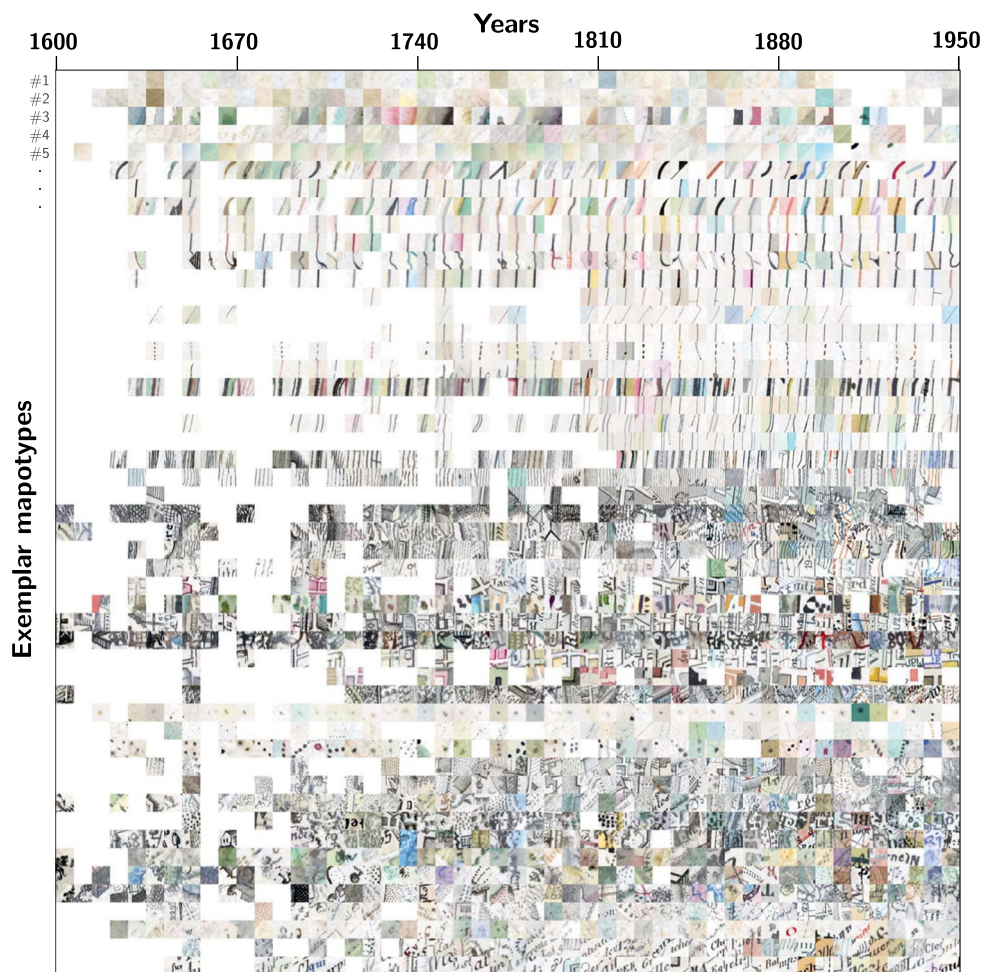
Fig. 4 Qualitative evaluation of an exemplar sample of mapotypes. Each row corresponds to a mapotype, and each tile corresponds to a mapel within that mapotype. For each mapotype, a sample of mapels is drawn at random from the corresponding year range subset. When no instance of the mapotype is present for a given year range, the space is left blank.

idiosyncrasy is also observed statistically (Table 2). Indeed, the W statistic of the SW test is increasing for the time slices 1731–1770 and 1771–1817, implying a more normal distribution of the relative frequencies per stratum. The 17th century and the beginning of the 18th century are characterized by an over-representation of handcrafted shading. This corresponds for instance to irregular hatching and diamond-shaped grid textures observed in Fig. 7a, b. The lines are often relatively thick and dark. There were also hand-painted icons, representing trees for example.

From 1731 onwards, more abstract textures, such as colored hatching, became comparatively more frequent. They adopted light pastel tones. The turn of the century (1770–1817, Fig. 7c) was a continuation of this trend of abstraction. Some characteristic motifs of this period included the abstraction of wooded areas or orchards by small circles (°) and vineyards by crossed-out S ($). Hatching became more regular and were also used to represent the approximate relief.

The studied period starts after the popularization, during the 16th century, of copperplate engraving by Italian mapmakers (Woodward, 1996). Prior to this, the predominant technique was xylogravure (woodcut). A certain legacy of this period remained, such as the aesthetic of the thick, dark line. Deeply marked lines could extend the longevity of the copper plate. The pursuit of deep black pigments, obtained from wine-black or from white wine and beer is also well documented in secondary sources (Woodward,

1996). In the 17th century, textures seemed to be used primarily for shading figurative elements (e.g. buildings). At this time, cartography still blended strongly with iconographic engraving and the adoption of a less abstract, oblique perspective was still very popular, as in the works of the Swiss cartographer Matthaeus Merian, for instance, in which buildings are stylized but clearly recognizable (Merian, 1638). This early ambiguity between iconography and cartography is one of the reasons why the 16th century was not included in the corpus. The period 1731–1817 truly materialized the transition to a more abstract cartographic representation. For example, the rather iconic representation of wooded areas and trees was replaced by a more abstract, symbolic one. The foliage of the trees became circles, the buildings or the colored furrows of the crops became hatching. The use of shading hatching was seemingly altered to represent topography and natural relief. This result can also be intuited from previous work, such as that by Dainville. Although, in that case, the punctual examples and counter-examples made it difficult to ascertain the presence of a general trend (Dainville et al., 2018).

1818–1834: The boom of cadastral survey. The early 19th century transition is the most marked in the corpus. The characteristic areas are highlighted in Figs. 6, A2–3, which corresponds to thin lines. The line, simple and uncluttered, became in fact the central element of the cartographic representation. We also notice in Fig. 7d the presence of colored borders, notably characteristic carmine pinks, but also shades of orange or blue. The distribution
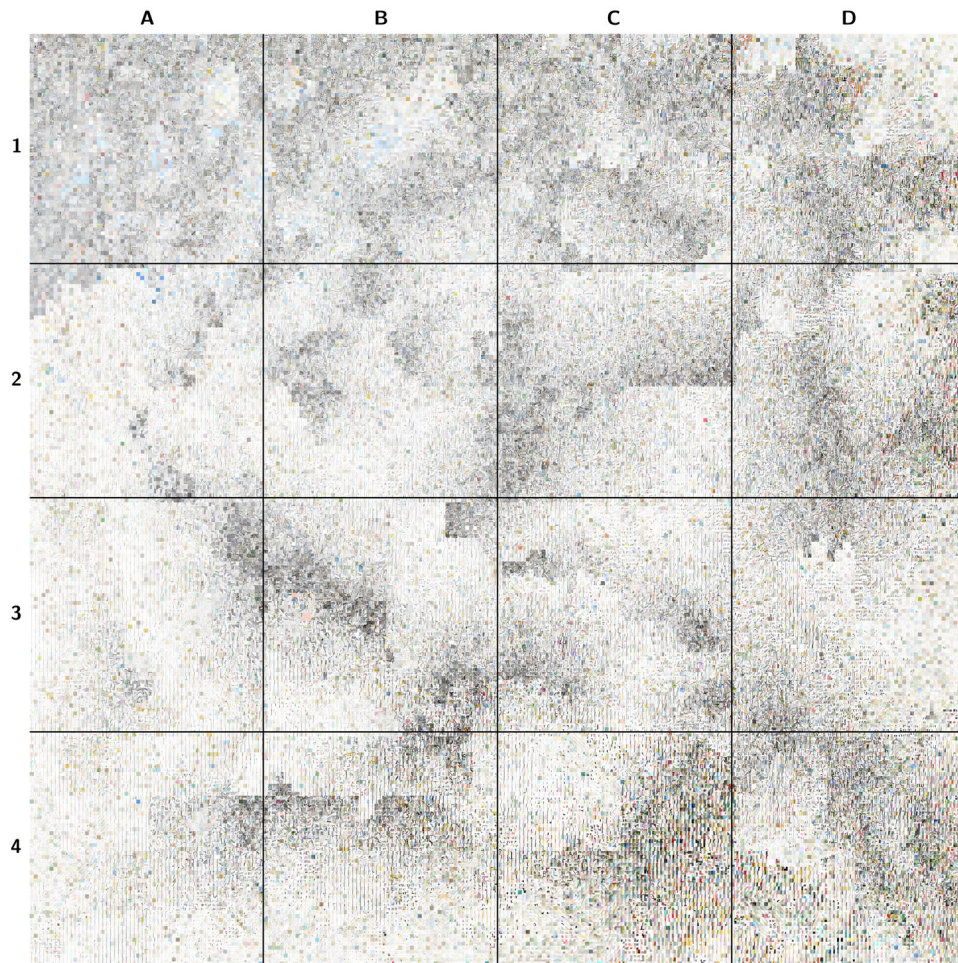
Fig. 5 Mapotypic mosaic. Constrained t-SNE visualization of all the mapotypes observed in the corpus. The graticule (A1–D4) is used in the text as a coordinate system to refer to specific zones of the image. The image enlargements in full resolution are available online[7].An enlargement of each image zone (A1–D4) is also provided in the supplementary material.

of mapotypes in this period also strongly coincides with their distribution among large-scale maps (1:1801–1:2500). The stratum 1818–1834 is also the period with the most characteristic figuration, according to the SW test (W = 0.81).

This period 1818–1834 was marked by the advent of the so-called 'Napoleonic' cadaster, of which a number of plates are found in the corpus. The production of the Napoleonic cadaster, which also extended to Switzerland, was a monumental undertaking that involved the surveying and recording of more than 100 million parcels in France. Established primarily for fiscal reasons, it also materialized the increasing influence of the administration (Clergeot, 2007). The cadastre was not only a powerful *dispositif*, reinforcing the control of authority over the population and the territory. In the context of post-revolutionary France and the First French Empire, the cadastre was presented as a way to introduce an equal and fair taxation system. The first attempts to establish a cadastre for the new empire began in 1807, but they led to very uneven results. Therefore, in 1811, the 'Recueil méthodique des lois, décrets, règlemens, instructions et décisions sur le cadastre de la France; approuvé par le Ministre des Finances' was published (Hennet, 1811). The document was supposed to organize and regulate the cadastration of the territory by the surveyors, and establish the mechanisms for the control of the surveys by a verifying engineer. The document also defines the scale of the cadastral maps, as well as the rules of surveying, the methods for the calculation of triangulation, and for the

recognition of owners. It also lays down the principles for the division of the commune into sections and the 'tableau d'assemblage'. The Recueil is also the document that founds the concept of urban cadastre (Bourillon, 2007). The instructions related to the figuration occupy a modest part of the document. As an administrative document with a very limited edition, the cadaster was not engraved, but hand drawn. The characteristic lines of this period, traced with Indian ink, were used to delineate the parcel of land, as well as the administrative boundaries, effectively and precisely. Furthermore, it was set that "the limits of each section are marked by a border of a different color for each" (Hennet, 1811: 97), which is indeed observed in the results. Carmine pink was the official color for representing the built environment, but the law also regulated the use of green, red, blue, and 'water green' (*vert d'eau*). What some cartographers call by extension 'Napoleonic cadastre' is in reality a set of documents produced in various administrative contexts, whose realization began during the Napoleonic period but continued until the middle or the end of the 19th century, according to the definition. It is not geographically circumscribed to France or the First French Empire, but also includes Belgium, the Netherlands, Switzerland, Spain, Northern Italy, and sometimes Austria-Hungary. These documents are characterized by a relative adherence to the rules described in the 'Recueil méthodique'. In fact, the first instances of the figurative principles laid down in the Recueil precede the publication of the latter. For example the
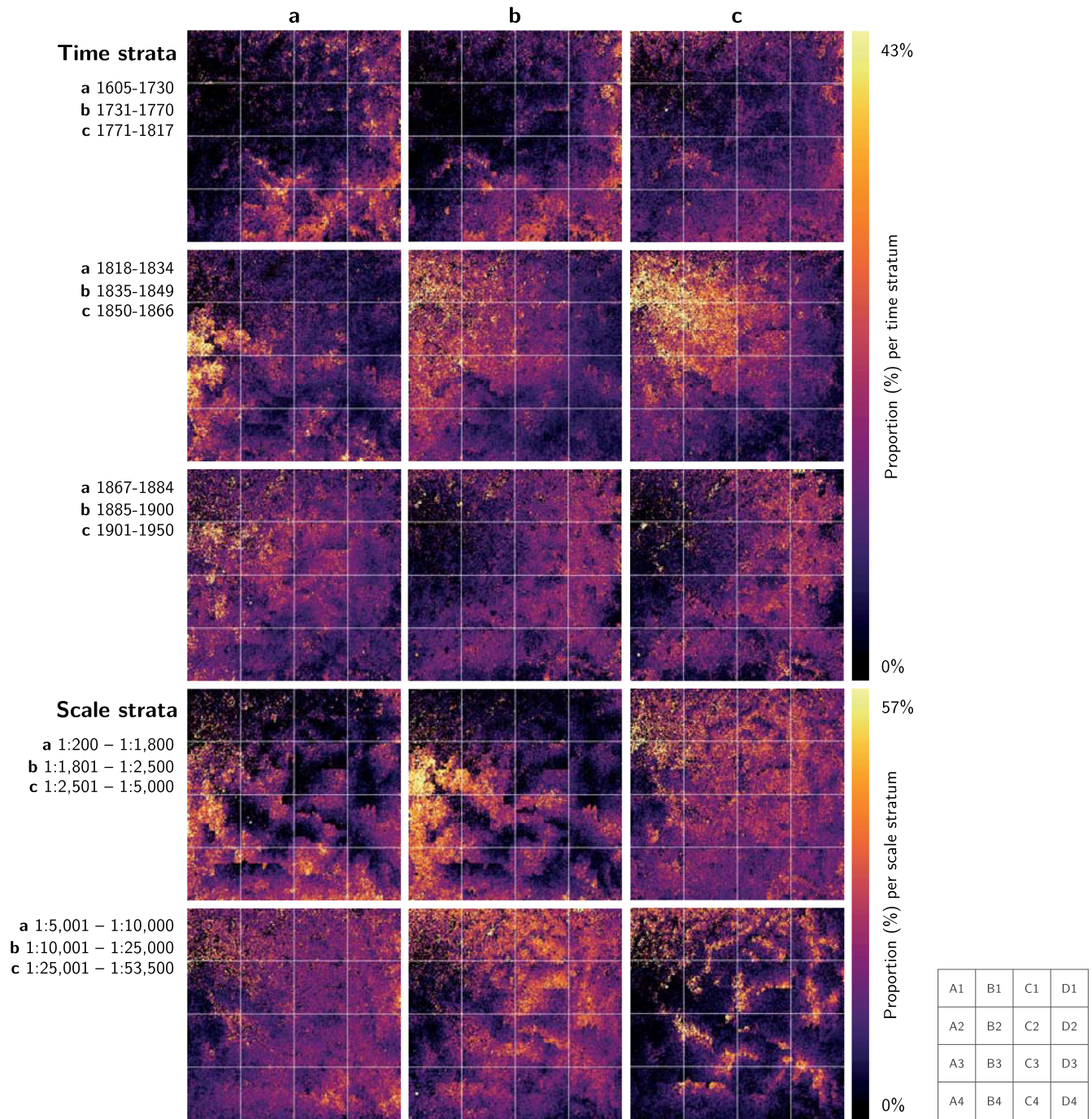
Fig. 6 Evolution of the relative distribution of the mapotypes through time, and by scale. The upper part of the figure represents the relative distribution of types, by time strata. The lower part of the figure represents the same distribution, by scale strata. Each block depicts the density map for a specific year or scale stratum, defined by uniform stratification (1/9th, 1/6th). The corresponding stratum is indicated on the left of each block row. The brighter areas correspond to a relatively high representation of the corresponding figurative subspace. The graticule in the lower-right corner can be used to make the connection between each block and the Fig. 5.

austrian cadastre of Venice, drawn up in 1808, is often considered to be one of the first Napoleonic cadastre (di Lenardo et al., 2021). It is also important to mention that other legal documents were sometimes established at a local level to regulate the surveying of the Napoleonic cadastre. For example, in the Swiss canton of Vaud, which includes the city of Lausanne, it is the 'Instructions données par le Département des Finances pour la levée des plans et l'établissement du cadastre, ensuite de la décision du Conseil d'Etat du 6 décembre 1826' (Soulier and Berdez, 1827). This

document repeats almost point by point the ideas of the Recueil. However, there are some interesting variations. For example, it states that water should be represented in blue, not green.

The cadastration campaign in France was interrupted shortly after the publication of the Recueil, first by the economic crisis and instability caused by the Napoleonic Wars, and then by the fall of the Empire (Vivier, 2007). Under the Bourbon Restoration, the idea that the cadastre allowed for an egalitarian taxation throughout the territory was highly contested. The administration

of the cadastre was reduced to a trickle, before being re-established by a regulation in 1827. It was not until this date, and even more so under the July Monarchy, that most of the territory was surveyed, which explains the chronology observed in our results. The highly regulated production context of the Napoleonic cadastre, in terms of drawing conventions, also explains why these are the least well differentiated series, according to the results presented in Fig. 3, and why the figuration of the maps in this time stratum is particularly characteristic.

Based on the quantitative methodology we describe in this article, (Petitpierre, 2023) explores how these legal instructions and the related graphical utterances were replicated in various local contexts, in France, but also in Switzerland, Belgium, and the Netherlands. According to their results, the visual consistency of the Napoleonic cadasters is at its maximum when observing the first generation of cadasters, established until the 1820s, and progressively decreases for the later ones ('renovated' cadastres, established in the second half of the 19th century). Urban cadasters seem also to be more consistent with each other, compared to rural ones, which is consistent with research on the diffusion of innovations (Lengyel et al., 2020). (Petitpierre, 2023)'s case study on the Napoleonic cadasters also shows how mapels can be used to compare the representation of various semantic classes, such as buildings, water, and the road network, in different cadastral cases.

Table 2 Statistic W of the Shapiro-Wilk normality test [0–1] applied to the relative distribution of mapotypes per time and scale stratum.

Year stratum	W	Scale stratum	W
1605–1730	0.87	1:200–1:1,800	0.88
1731–1770	0.89	1:1,801–1:2,500	0.87
1771–1817	0.92	1:2,501–1:5,000	0.91
1818–1834	0.81	1:5,001–1:10,000	0.92
1835–1849	0.84	1:10,001–1:25,000	0.94
1850–1866	0.84	1:25,001–1:53,500	0.82
1867–1884	0.82		
1885–1900	0.92		
1901–1950	0.87		

Lower W values indicate departure from normality, thus a somehow more characteristic figuration in a particular stratum.
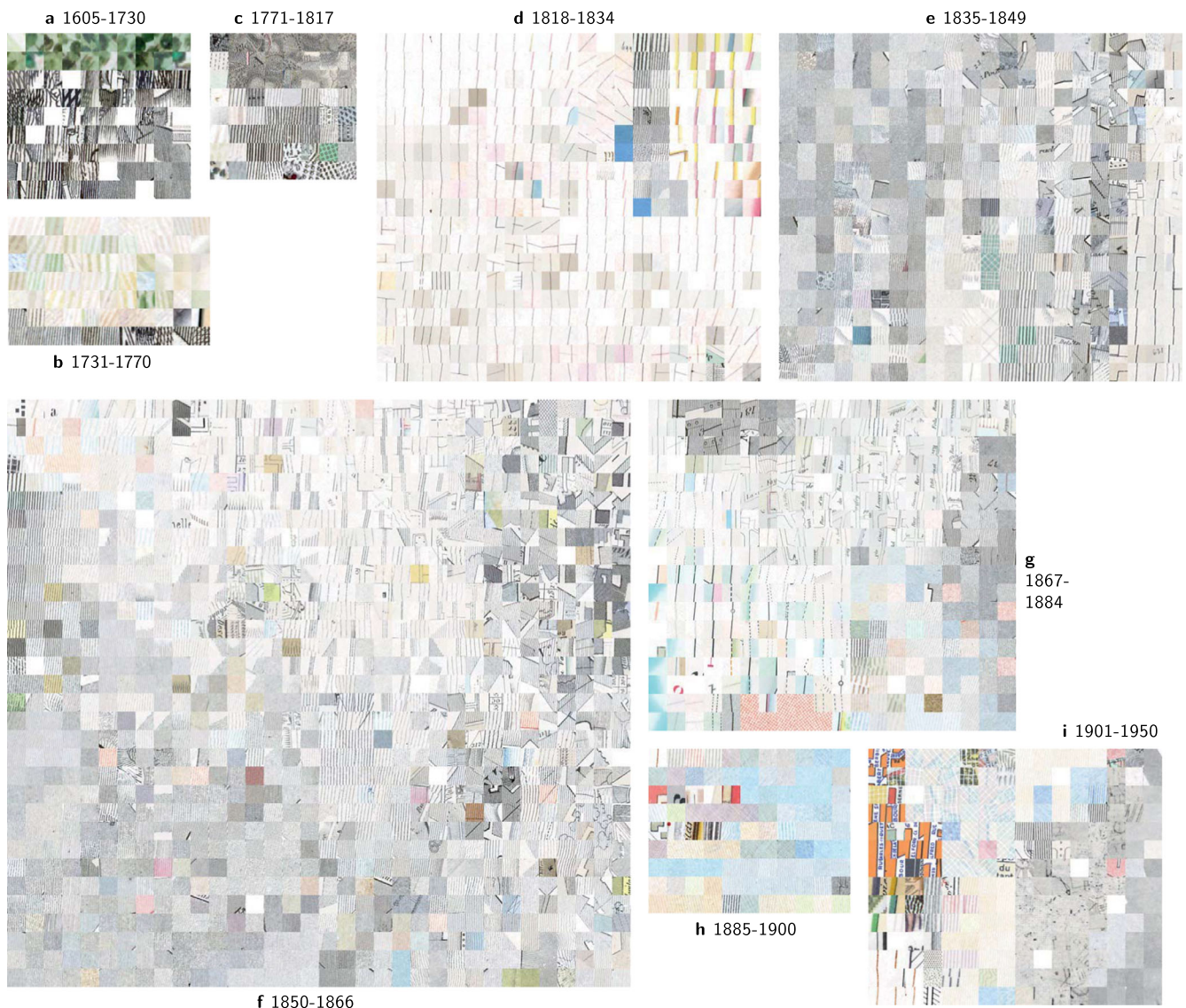

Fig. 7 Top 5% most characteristic mapotypes by time strata. a 1605–1730; **b** 1731–1770; **c** 1771–1817; **d** 1818–1834; **e** 1835–1849; **f** 1850–1866; **g** 1867–1884; **h** 1885–1900; **i** 1901–1950.

1835–1884: Cartography in the Industrial Revolution. The years 1835–1849 and 1850–1866 enabled for the exploration of new figurative spaces (Figs. 6, A1, 2, B1, 2), and were characterized by a marked return of hatched textures (Fig. 7e, f). This time, however, the patterns were more regular, the lines were generally finer and the spacing between them could be shorter. We also observe in the bottom-left part of Fig. 7e shading lines of varied identical length. From 1850 onwards, color prints also became visibly more frequent.

These developments, which took place in the mid-19th century, were likely due to the propagation of technological innovations that occurred several years earlier, in the late 18th and early 19th centuries. A first innovation was the development, thanks to the advances in mechanics that preceded the French Industrial Revolution, of rocker mechanisms and mechanical rulings (Robinson, 1975). These tools hastened the engraving of straight lines and hatching. Even shading is somewhat rationalized. In the map of the Etat Major, for instance, shading lines are in fact based on surveyed contours curves (Arnaud, 2022: 138–149); the length of the shading lines is inversely proportional to the slope steepness. Another innovation was the invention of lithography, which, by exploiting the chemical antagonism between hydrophilic and hydrophobic substances, made it possible to draw the maps more naturally, directly on the plate. This innovation, detailed in 1818 by the German Alois Senefelder (Senefelder, 1818) and translated into French in 1819, considerably accelerated plate drawing compared to copperplate engraving, while offering a more precise and finer line than etching. However, if this method was popularized in Germany as early as 1809, with the Bavarian cadastral campaign, its diffusion in the French-speaking world was slower (Ristow, 1975). It was not until 1826 that a commission of the Société de Géographie emphasized the advantages of this new method, although it still considered copperplate engraving to be "more perfect" (Jomard et al., 1826). In 1837, the Frenchman Godefroy Engelmann also presented a method of chromolithography that became very popular and made a major contribution to the development of color printing (Ristow, 1975). Our results (Fig. 7g) indicate that the quantitative impact of these technological innovations was not immediate. Indeed, it is reflected especially in the period 1835–1866, and even rather 1850–1866 for color printing, during which the production really took off. This time delay between the technical innovation and its dissemination is not surprising. Indeed, our definition of the characteristic idiosyncrasy of the figuration in a certain period is above all quantitative and thus tends to reflect substantial and momentous developments, rather than avant-garde movements. The *Carte géologique générale de la France*, for instance, was printed first in color in 1842 (Rossi, 2005). By 1855, when this map was showcased at the Universal Exposition, 3,000 copies had been printed. Printing large-format maps in color remained a challenge until the mid-19th century. The first sheets printed in colour by the *Dépôt de la Guerre*, for instance, had to wait until the 1860s (Arnaud, 2022: 76). Color printing, then six times cheaper than manual colorization, finally enabled the spread of color in mapmaking (Cook, 1995).

The period 1867–1884 continued and strengthens the mid-century trends. We also observe more texts, such as street names, as well as the strong representation of regular dashed lines, which make this period very characteristic, in terms of figuration (Table 2). This period was marked, among other things, by the regular publication of the Parisian atlases, in particular the Alphand plans, from 1867, towards the end of the Haussmann works. Some mapels of these maps are represented among the most characteristic mapotypes of the period.

1885–1950: Democratization. During the later part of the 19th century and the first half of the 20th century, color became completely prevalent. Almost all the characteristic mapotypes of the period 1885–1900 are colored (Fig. 7h), such as the fine blue hatching used in the Siegfried series of national maps in Switzerland to represent lakes, or the brown contours lines, whose presence also reflects the technical progress of the 19th century in terms of relief depiction (Hurni, 2008), but also in terms of geodetic survey. In the first half of the 20th century, we note the presence of mapotypes from urban maps, containing textual information on street names. A graphical and typographical evolution can also be observed in this respect with the development of a more modern aesthetic. Finally, we also notice the presence of halftone patterns consisting of a grid of small dots in grayscale.

The production of the national topographic maps in the 19th century, such as the maps of the Etat Major in France, and the Siegfried maps in Switzerland, represented a masterpiece that incarnates the rise of nation states. In the 19th century indeed, cartography is thought to have played an essential role in the construction of modern nation states (Branch, 2014; Schulten, 2012). The production of the Siegfried national maps for instance, was the subject of intense debates among the scientific community and the military institutions. Its production was transferred from lithography to more costly but precise copperplate engraving beginning in 1910, resulting in particularly fine textures and sharp contours (Feldman, 2015). The creation of national topographic series is also an opportunity to undertake new geodetic surveys. These increasingly accurate elevation measurements were plotted as contour lines. Altough this technique for representing land relief dates back to the end of the 18th century (Arnaud, 2022: 63), it widespread use was slowed down until sufficient elevation data was available. The beginning of the 20th century also reflects another paradigmatic shift in cartography. The map was no longer limited to a military or administrative tool. Indeed, following the technical advances of the 19th century which enabled a vertiginous rise in production, and the increasing mobility, the touristic or orientation map became popular, creating new economic opportunities (Akerman and Nekola, 2023; Olson, 2010). The cartography of this period also shows the arrival of photography, and the development of aviation. These technologies led to the appearance of photomechanical printing processes, such as rotogravure or halftone screens. These techniques produce very convincing halftone patterns that are reflected in the image by the presence of regular screens composed of small dots, as observed in Fig. 7i (Cook, 2015). These photographic techniques were used, for example, to map mountain relief in the Alps, either by aerial photograph (Imhof, 1982: 48–52), or by photoengraving from a relief model (Pearson, 2015: 1260–1262).

Macroscopic trends. In this final result section, we look at some of the macroscopic trends that can observed in the longer term. These trends can be derived from the observation of the interpretable features we have calculated, or simply from the relative distributions of mapotypes, the functional units we defined in section "Interpretation of the feature space" as "the archetype obtained from the reduction of a class of variant cartographic utterances".

Disruptive evolution regarding scale. The issue of scale also plays an important role in the characterization of map figuration (Fig. 6). We notice that mapotypes corresponding to 1:200–1:2,500 maps occupy very distinct figurative spaces compared to 1:25,001–1:53,500 maps. As evidenced by the high value of the SW test statistic (Table 2), the stratum of 1:2,501–1:25,000 maps partly bridges the gap between the characteristic large-scale maps and

the equally characteristic very small-scale maps. Fig. 8 shows that the correlation (as measured by Kendall's tau coefficient, Kendall, 1938) of this stratum with other scale strata is weak, but above average, which supports a bridging role hypothesis.[9]

For the 1:25,001–1:53,500 maps, the distribution is concentrated in very bright clusters in Fig. 6. A visual investigation of the mapotypes informs about the strong predominance of lines at large scale (1:200–1:2500), with lesser graphical load, while the small-scale maps (1:25,001–1:53,500) show a higher graphical load, which seems not only due to a higher concentration of information, but also to the presence of many textures depicting relief and topography with shading (see also Supplementary Material). Figure 8 suggests that these figurative differences develop during the period studied. At the beginning of the period indeed, the results show a positive, albeit weak association between the cartographic figuration at various scales. However, from the end of the 18th century, the association becomes more marked and negative. This indicates that the corresponding stylometric map spaces became increasingly distinct.

This disruptive evolution can intuitively be understood as a phenomenon of differentiation, in which one genre, the map, separated into two distinct sub-genres, specializing differently for small- and large-scale maps. This result is interesting, firstly because of its temporality, since this phenomenon, beginning between 1771 and 1817, was concomitant with the appearance of

the Napoleonic cadaster, and secondly and more importantly with the invention of the concepts of 'small-scale' and 'large-scale' themselves, as Edney demonstrates (Edney, 2019: 171). While he postulates a generalization of the principles of large-scale mapmaking, and foremost the idea of scale, to small-scale maps, we observe a figurative divergence. This is not incompatible, as figuration could have become a novel way of distinguishing these newly conceived categories, that became otherwise united by making practices. This scale-related differentiation seems persistent. Indeed, even if it weakens slightly at the end of the century, it is far from returning to its previous values. These results suggest that cartographic figuration, initially independent of scale (hence the low initial correlation), diverged durably at the turn of the 19th century (negative correlation). This dichotomy is not surprising for those studying historical and contemporary cartography. Until the end of the 20th century and the arrival of web mapping, which has somewhat blurred the boundaries, the style and semiology of large-scale maps (e.g. cadasters) and smaller-scale topographic maps differed evidently. However, our results suggest that such a dichotomy may not be justified for earlier maps.

Fashions and macroscopic trends on representation. As the selected features are interpretable (Table 1), they provide informative metrics on the macroscopic evolution of the corpus (Fig. 9). The Regional Kendall statistical test (RK, Helsel and Frans, 2006) assesses the significance of a regional trend.[10] In the present case, 'regions' are not geographical, but rather scale-based strata. Indeed, the measured variables –graphical load, line width, number of components– are not expected to be independent from scale. Therefore, we constitute 3 equilibrated 'regions' for large-, medium- and small-scale maps (1:200–1:2500, 1:2501–1:10,000, 1:10,001–1:53,500), to assess that the trends observed are not due to a variation in map scale, but rather to a general trend in cartography. The results of the statistical analysis are summarized in Table 3.

These macroscopic evolutions (Fig. 9) are consistent with the effects observed in the previous section. They highlight technological and cultural processes that first favored the production of finer lines, with a peak in the first half of the 19th century, before the line was reinforced again. The process of abstraction of geographical elements, or shift from a rather iconic to a more symbolic representation, which took place from the 17th to the beginning of the 19th century (Fig. 7a–c), seems accompanied by a stability of the perceived graphical load,[11] despite the increase in the number of components. This suggests an improved readability of the map. However, from 1807 onwards, the graphical load increased again, accompanied from 1834 by the
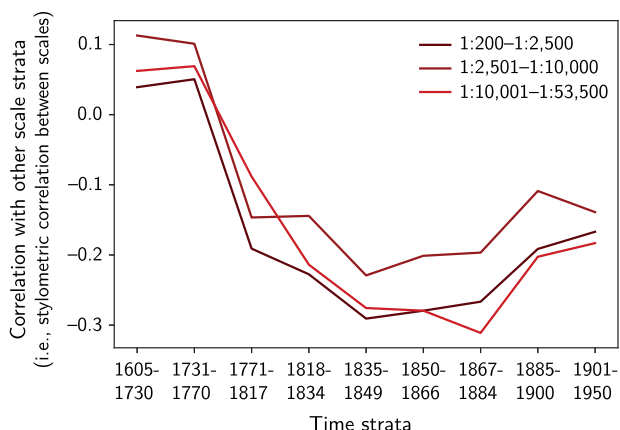

Fig. 8 Evolution of mean Kendall's correlation tau coefficient between the relative distribution of mapotypes for the 3 scale strata, through time. Kendall's tau measures the degree of similarity the distributions, i.e., the stylistic correlation between different map scales. All $p_{val} < 8 \cdot 10^{-4} < 0.05$. $n = 55,599$.
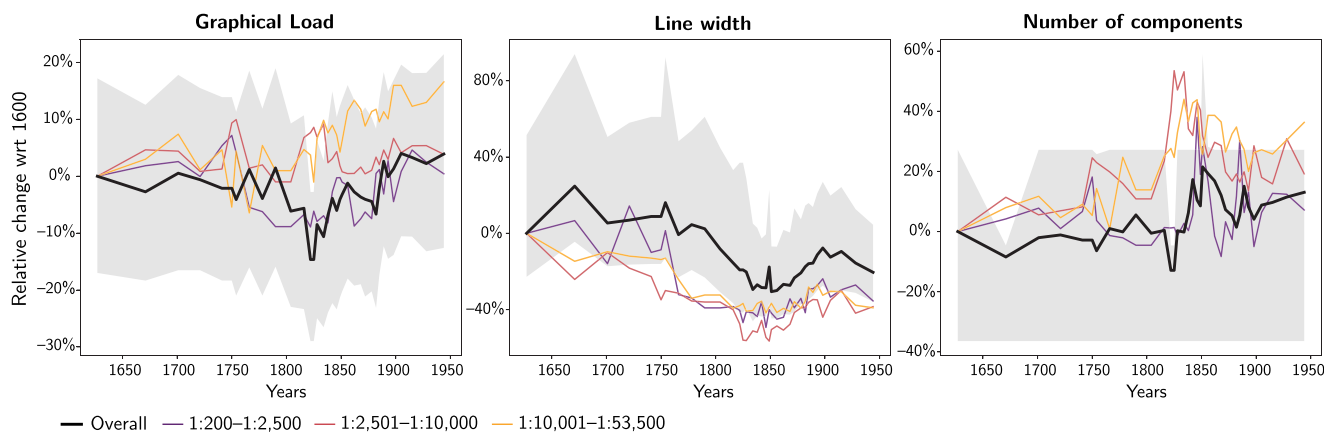

Fig. 9 Evolution of graphical load, line width, and number of components through time, overall and per scale-based strata. For the overall evolution, the median line, and the 25–75th percentiles interval are represented.

Table 3 Statistical evaluation of the trends observed in the macroscopic evolution of the variables (Fig. 9), using a Regional Kendall test.

Feature	Time range	Trend	n	p_{val} ($\alpha = 0.05$)	S
Graphical load	1600–1807	None observed	7	0.06 > 0.05	23
	1808–1950	Increasing	28	$5 \cdot 10^{-8}$	479
Line width	1600–1834	Decreasing	13	$8 \cdot 10^{-10}$	−175
	1835–1950	Increasing	22	$3 \cdot 10^{-9}$	365
Number of components	1600–1950	Increasing	35	$1 \cdot 10^{-3}$	404

The data are distributed in 35 temporal strata, and the test is performed jointly on the median of each scale-based strata (1:200–1:2500, 1:2501–1:10,000, 1:10,001–1:53,500). The test is bilateral, with H_0 corresponding to an absence of monotonic trend. Kendall's S statistic is an indicator of the magnitude of the trend.

thickening of the line, as well as the continuous increase in the density of graphical components. This phenomenon also coincides with an unprecedented increase in cartographic production, at the same period.

One possible hypothesis, based on the analysis of the macroscopic trends (Fig. 9), as well as the most characteristic mapotypes for the period (Fig. 7), would be that the map could have reached, at the end of the 18th century, a certain state of sophistication, which might in turn have favored its use as a device to spatialize other kinds of data, for example administrative or military. To fulfill these new functional requirements, specific and additional informational layers, such as topography, infrastructure, or property ownership, could be superimposed on geographic data. These additional geometric objects may have increased the graphical load and the density of components. Furthermore, at this time, the advances in cartographic printing techniques led to an unprecedented increase in production. We suspect that the sudden abundance of maps in certain social circles could also have fueled its complexification since the cartographic literacy of readers might have increased. At the same time, the improvement of printing methods also provided the technical requirements for creating more regular and finer patterns. These innovations seem to translate into a fashion of the fine line.

This trend seems to fade at the end of the 19th century, when the lines became again thicker, and more visible. This is evident in Fig. 7, where, at the beginning of the 20th century, mapotypes showing strong and contrasted lines are clearly visible.

General Discussion and Conclusion

To conclude, in this article, we present a new stylometric methodology applied to cartographic information. Our approach draws inspiration from structural linguistics, as well as distributionalism. We take as a case study a large and diverse corpus of more than 10,000 French and Swiss maps. The maps are decomposed into elementary map units (mapels). A set of interpretable features are selected, first based on the literature, and later by optimizing the representation to keep only those features that best contrast the cartographic series, and thus capture their cultural differences. This optimized feature representation makes it possible to group mapels according to their type. We call the resulting class 'mapotype'.

The feature representation that results from this first approach effectively differentiate the various cartographic series, which confirms its effectiveness. Morphology seems to be the most relevant feature in the construction of this representation, followed by graphical load, line thickness, and texture. Counter-intuitively, color

does not seem to stand out. The diachronic analysis of mapotypic distributions permit to identify key shifts or stabilities corresponding to important cultural and historical changes. The investigation of our corpus highlights a slow process of abstraction of the representation of the environment in the 17th and 18th centuries, which appears to favor symbolic representation over iconic ones. From the 19th century onwards, the boom of the cadaster reflects the rise of the administration, and the technological developments in mechanics and chemistry, in the context of the industrial revolution, especially the invention of mechanical rulings and lithography. During the 19th century, a peak in the production of finer lines is observed, supported by these technical developments and by cultural processes. We also evidence a macroscopic trend towards increased map load, which can be interpreted as a sophistication of the map, used as a support for the representation of various data (e.g. topographic, ownership, etc.). Later, the rise of nation-states is visible, with the production of color-printed national topographic maps, as well as the arrival of photography, aviation, and the growing tourism.

We see the approach we present in this paper not as a definitive tool, but rather as an initial proof of concept for cartographic stylometry. Pragmatically, we propose a computational fragment-based methodology for the visual study of cartography and demonstrate that it permits an effective dialogue with the history of cartography. This paper has therefore, above all, a confirmatory role, with respect to the methodology, by demonstrating that the results it produces integrate effectively with the existing literature on cartography. However, it has also made it possible, incidentally, to underline phenomena that are little or not theorized. This is the case of the disruptive evolution of maps according to scale, which seems to take place at the turn of the 19th century and cause a lasting split in the figuration of small- and large-scale maps, with the appearance of two sub-genres. The possibility of detecting such visual linguistic evolutions on large map corpora is original.

Another notable observation derived from our analysis pertains to the evident shift in cartographic representation at the same time. This 'epistemological rupture', to coin the term introduced by Gaston Bachelard and Louis Althusser (Althusser, 2005; Bachelard, 2001) is characterized by a substantial discontinuity in the distribution of mapotypes, which indicates a momentous change in the prevailing paradigm. The scale of this rupture also resonates with Michel Foucault's work (Foucault, 1976), as he identified a profound shift in the way knowledge is constructed at that time, which he interprets as the transition from a classical to a modern episteme. Our findings similarly suggest an evolution at an epistemological level. Between 1600 and 1730, maps in our corpus are characterized by handcrafted shading, figurative elements, and a focus on visual aesthetics. This approach echoes a worldview that values direct resemblance and interpretation of the world. As time progresses, we note a transition to more abstract forms of representation. Instead of iconographic representation, cartographers began using simpler symbols to represent geographical features. This shift seems to reflect a changing understanding of knowledge as more abstract, categorized, and systematized. By the early 19th century, information overlay and data mapping, with a growing number of components, and soon graphical load, reflects an increasingly abstract and systemic understanding of geographical space.

In this regard, the scholarship of cartographic stylometry for which we advocate, is not only about retracing aesthetical ideals or technological progress. Notwithstanding its interspersion with technical innovations, the narrative that we attempt to construct allows us to read, in the multitude and beyond the eminent instances, the way in which the visual language of the mapmakers, evolves. Further, the changes we observe when studying maps do not only concern the way in which cartographers represent the world, but also how they perceive and idealize it.

A computational approach to cartography is not an end in itself. To propose a *mise en abyme*, we can draw an analogy between the computation of a formal feature space and a mapping process, with several implications: it is a subjective, reductive, incomplete, partial, idealized, constructed discourse on cartography (Crampton, 2001; Edney, 2019; Harley, 1989). However, like maps themselves, it offers the advantage of enabling discussion on territories that are beyond the purview of a single observer. In this regard, maps are astonishing devices that are connecting us to large and complex abstract ideas that could not be conceived without them. Our study focused on a varied corpus of more than 10,000 maps. As a comparison, it is already more than all the maps present in the five volumes of the *History of Cartography* series, whose analysis and discussion required the work of hundreds of the best experts in the field. The comparison is not about the nature of the knowledge produced, which is necessarily richer and more detailed in the second case, but merely about scalability. As we write, not tens of thousands but closer to a million[12] maps have already been digitized by national libraries and map collections throughout the world. We can content ourselves with studying one percent of all maps and ignoring the majority, or we can try and invent new methods of approaching this vastly unexplored mass. We advocate for the second option. Computational methods can be used in complement to qualitative close interpretation of the documents, in particular for the confirmation of formulated research hypotheses on larger corpora, or the exploration of macroscopic trends. For instance, we could cross-validate the theoretical taxonomies of cartographic modes, such as the one proposed by Edney (Edney, 2019: 33). Finally, the methodology for cartographic stylometry we propose is based on particular details (mapels) embedded in individual maps, which may themselves be part of larger ensembles. This multiscalar trackable approach means that we can move back and forth between these different scales, and thus bridge qualitative and quantitative cartographic studies.

To conclude, as intriguing as the initial results proposed in this article might be, the design space for this approach still remains largely unexplored. The implementation of this method raised many questions. For example, at what scale (or scales) does the style of the map reside, and how do we define the size of the mapel based on this? What is the optimal number of mapels to represent a map? Should mapotypes be able to overlap? Is it desirable to extend the concept of map series to improve the representativeness of the model? What about authorship? On a purely technical level, one could also integrate other candidate features, for example by considering alternative color spaces (Banda et al., 2011; Ebi et al., 1994; Herrault et al., 2013; Salvatore and Guitton, 2004). In addition, various mathematical strategies for optimizing the feature representation could be explored, including differentiable ones. In future work, we expect to address some of these questions. We also seek to integrate semantics into the model, to foster a discussion of map semiosis. Furthermore, while maintaining our focus on cartography as a priority, we seek to build an approach as generic as possible, as it seems to us that the methodology could also be transposed to other domains. Future research should replicate the methodology to other corpora, to prove its genericity, and improve the specificity of the formal representation. Moreover, as the method demonstrated sound results in relation to the historical writing on cartography, we believe that it will now be possible to delve deeper into more complex cultural mechanisms, such as the propagation of cultural conventions over the long term and across large areas, a measure of evolutionary complexity in cartography, and the relationship between the signs and the perception of the environment itself.

Complement on techniques and methods

Segmentation of the map background. The geographic content of the map is separated from the background (map frame, title block, legend, scale, scanner background, etc.). For this step, a randomly drawn sample of 1061 maps, reduced to an area of 10^6 pixels (approx. 1000 × 1000 pixels) is manually annotated, i.e., the actual geographic area within the scanned map image is delineated. This annotated sample is divided into train, validation, and test subsets (70/20/10%) and is then used to train an OCRNet neural network (Yuan et al., 2020), implemented in mmsegmentation (MMSegmentation: OpenMMLab Semantic Segmentation Toolbox and Benchmark, 2022), in order to segment the maps into geographic vs. background map content. OCRNet uses an HRNet-V2p-W48 encoder, and is pre-trained on the Cityscapes-160k dataset (Cordts et al., 2016), containing urban scenes in first person view. The network is trained by stochastic gradient descent (lr=0.01), using the cross entropy as loss, until convergence. The trained model achieves an intersection over union (IoU) of 87.7% for geographic content segmentation (pixel accuracy 90.5%).

Exclusion of the image areas with a low graphical load. To limit the implicit impact of low information areas, the graphical load of each tile is computed, using ED2, an approach based on edge density. This metric was proposed by (Barvir and Vozenilek, 2020), who have evaluated and compared several metrics for computing the graphical load of maps, against a human panel. ED2 is the metric that appears to be the best proxy for measuring the perceived graphical load, according to their study. The computation of ED2 is done using a Scharr filter (Scharr, 2000) to detect edges, and applying a square root operator to the result. Then, areas with very low graphical load (ED2 < 2.75) are excluded. This criterion is applied once again after the normalization of the mapels.

Normalization of the mapels. The mapels are recentered on the most salient graphical feature close to their initial randomly drawn location. This is done by computing the pixel with the lowest value in the 50×50 area centered around the initial location and setting it as the new center. To neutralize the orientation, the Histogram of Oriented Gradient (HOG, (Dalal and Triggs, 2005)) is computed on each mapel (Eq. 4a). As its name indicates, the output of HOG is a histogram of the local orientation of graphical features, with 180 bins i.e. one per degree. The maximal peak in the histogram, corresponding to the dominant edge orientation, is stored as the mapel's orientation (Eq. 4b).

$$HOG(m_i) = \{h_0, h_1, ..., h_{179}\} \qquad (4a)$$

$$orientation(m_i) = arg \max_{h_j} HOG(m_i) \qquad (4b)$$

Then, the mapel's orientation is neutralized by simply locally rotating the image by its orientation.

Computation of candidate stylometric features. The candidate stylometric features are computed on each mapel. First the RGB color histogram is computed on each channel separately. Two different values are tested for the number of bins per channel: 6 and 9, resulting in an 18-dimensional, and another 27-dimensional vector for this feature[13]. The other features relative to color distribution (peak color, standard deviation, skew, kurtosis) are computed on an extensive 256-bins histogram, for each channel separately (resulting on a 3-dimensional vector for each feature). The peak color is calculated by finding the histogram maximal peak (argmax). The use of these 4 features for the computational analysis of map images was first proposed by (Uhl et al., 2018). The HOG (Dalal and Triggs, 2005) is computed on the whole mapel, in grayscale, setting the number of bins to 12 for

the orientation histogram, and testing three different parameters for the cell size (5, 10, and 25). The output counts three 12-dimensional vectors for this feature.

The linear binary pattern (LBP, (Ojala et al., 2002)) was computed on the black and white image, which is obtained by applying Otsu's (Otsu, 1979) thresholding algorithm for binarization. The number of bins is set to 12, as for the HOG, and three different values are tested for the radius parameter (2, 3, and 4), accounting for various texture scales, and resulting in three 12-dimensional vectors.

For the graphical load, three approaches are assessed. The first involves a simple computation of the proportion of dark pixel in the binary image, after Otsu's thresholding. The second, ED2 is based on edge density, as explained already. The last one involves counting the number of dark connected components after Otsu's thresholding. For instance, a single line in a mapel will result in a value of 1, a double line in a value of 2.

To compute the line width, the image is first binarized, then thinned (i.e. skeletonized) using the algorithm proposed by (Zhang and Suen, 1984). The width is obtained by calculating the ratio before and after thinning. Finally, two methods are tested to operationalize the orientation. The first is using the value computed in the previous section, i.e. the maximum peak of the HOG prior to the neutralization of orientation. The second corresponds to the binned HOG pattern method, proposed by (Petitpierre, 2020) for the analysis of historical maps. This method also relies on computing the HOG, with 24 bins. The orientations are then regrouped into discrete angle bins: vertical ($\pm\pi$), horizontal ($\pm\pi/2$, $\pm 3\pi/2$), diagonal ($\pm\pi/4$, $\pm 3\pi/4$), regular oblique ($\pm\pi/6$, $\pm 2\pi/6$, $\pm 4\pi/6$, $\pm 5\pi/6$), and irregular oblique (all other orientations).

Features selection with a genetic algorithm. The optimization process (see Eq. 3) relies on the genetic algorithm implemented by (Solgi, 2020). The candidate features are normalized before optimization. The population size at each epoch is 20. The ratio of parents retained from one generation to another is 4/20. These 'parents' combinations are sometimes (with a 50% probability) recombined uniformly within each other with a single crossover point. The genetic algorithm is trained until reaching a plateau, which occurs after 80 generations.

Iterative subdivision of the feature space into mapotypes. To obtain the mapotypes, the feature space is iteratively subdivided, until obtaining cells of radius $r < k$, where k is the radius of free variation. The iterative division of the space is performed using a KMeans algorithm. Each cell is iteratively divided until the radius r between the center of the cell and the outermost mapel is less than k. The center is defined, in the sense of KMeans, as the geometric center of the samples in the cell. KMeans' K parameter, corresponding to the number of clusters or subdivisions at each split, is set to 4.

Data availability
The complete list of maps in the corpus, as well as the associated metadata table, are released along with this article (https://github.com/RPetitpierre/cartographic-stylometry/tree/main/corpus). The metadata table also includes a hyperlink to access all the digitised maps available in open access online via a IIIF protocol or similar (9569 documents, 95.2% of the total). 244 maps can be downloaded manually from the websites of the corresponding institutions, free of charge. The remaining 234 documents can be directly requested from the corresponding public archives, free of charge.

Code availability
The segmentation of the map background relies on the open-source implementation of OCRNet by OpenMMLab (MMSegmentation: OpenMMLab Semantic Segmentation Toolbox and Benchmark, 2022). The optimisation process relies on the genetic algorithm of Ryan M. Solgi (Solgi, 2020), freely available on Github. All major algorithms described in the article, including those for fragmenting the map into mapels, computing stylometric features, optimizing and splitting the feature space to identify mapotypes are made open-source and released along with this article (https://github.com/RPetitpierre/cartographic-stylometry/tree/main/code).

Notes
1 Here we understand etic unit, in the sense of semiotics, as defined for instance by Nöth in his *Handbook of semiotics*: "The etic unit, the phone or *morph*, consists of individual and contextual variants of this abstract unit [NB: the emic unit]" (Nöth, 1995: 183).

2 This number of 800 is defined arbitrarily. It is considered sufficient to adequately cover multiple areas of each map image, and account for the overall figuration. Taking the same number of 800 samples on each map results in a symmetrical and equitable measure of their figurative content. This also ensures a comparable representation of each map in the data, regardless of the size.

3 Previous studies aiming at semantic classification of map patches considered patch sizes ranging from 48×48 px to 100×100 px (Hosseini et al., 2022; Uhl et al., 2018, 2020). This is indicative of the order of magnitude of graphical features on the image. In the present case, we focus on style rather than semantics. Knowing that semantic classification requires, in addition to the graphical features themselves, to capture visual context, we assume that the minimal unit of style might lie at the lower end of this range. Further studies would be welcomed to challenge this assumption.

4 The empty areas all resemble each other, except maybe for some peculiarities of the support. We observed that not removing them significantly alters the possibility to analyze other stylometric features. Indeed, as the proportion of empty space in the map is a ubiquitous factor for resemblance between two maps, the optimization process tends to overfocus on that aspect and neglect the actual figurative content of mapels. To avoid that, we prefer to exclude the areas that exhibit a particularly low graphical load, and separately store their proportion in maps as an explicit data, that can be studied purposely.

5 For the technical description of the steps in this paragraph, please refer to the Complement on Techniques and Methods, at the end of the article.

6 The distance $D(A, B)$ can also be intuited as a pointwise minimum Hausdorff distance (Hausdorff, 1914).

7 https://github.com/RPetitpierre/cartographic-stylometry/tree/main/Figure_5

8 Note that for $n > 5000$, the validity of the p_{val} is not guaranteed, as it can reach very small orders of magnitude. However, the W statistic remains correct, and this is what we are interested in here.

9 The use of this rank-based method is justified since the relationship between relative distributions is not linear. This is easily verified as a polynomial regression of degree ≥ 2 yields a lower R^2 compared to a simple linear regression. $R^2 = \frac{u}{v}$, with $u = \sum_{\forall i}(Y_i - \hat{Y}_i)^2$, the residual sum of squares and $v = \sum_{\forall i}(Y_i - \overline{Y_i})^2$, the total sum of squares. With $\hat{Y}_i$ the predicted value, Y_i the actual value, and $\overline{Y}_i$, its mean. $\overline{R^2} = 0.85$ for linear regression, $\overline{R^2} = 0.83$ for polynomial of degree $n = 2$, $\overline{R^2} = 0.78$ for $n = 3$, $\overline{R^2} = 0.74$ for $n = 15$.

10 It is valid for $n > 3$.

11 The ED2 metric, which we use to measure graphical load, is an effective proxy for estimating perceived graphical map load, as was demonstrated by (Barvir and Vozenilek, 2020) in a panel experiment.

12 Just 13 of the largest digital map collections already amounts to more than 840,500 documents available online: National Library of Scotland 200,000+, David Rumsey 115,300+, Stanford Library 97,600+, British Library 73,000+, United States Geological Survey 72,700+, Bibliothèque Nationale de France 69,200+, Library of Congress 55,800+, Biblioteche Italiane 33,700+, Charles University 28,500+, New York Public Library 28,200+, Deutsche Fotothek 25,600+, University of Princeton 20,500+, Perry Castaneda Library 20,400+.

13 Note that the retention of more than one of these values is allowed in the optimization process described in the next section. This is also the case for the two other parametric candidate algorithms, LBP and HOG.

References

Akerman JR and Nekola P (2023) Mapping Movement in American History and Culture. Available at: https://mappingmovement.newberry.org/ (accessed 5 December 2023)

Althusser L (2005) *For Marx*. London; New York: Verso

Ares Oliveira S, di Lenardo I, Tourenc B, et al. (2019) A deep learning approach to Cadastral Computing. In: *Digital Humanities Conference*, Utrecht, Nederlands, 8 July 2019. Available at: https://dev.clariah.nl/files/dh2019/boa/0691.html (accessed 5 November 2020)

Arnaud J-L (2022) *La carte de France – Histoire et techniques*. Marseille: Parenthèses. Available at: https://shs.hal.science/halshs-03688318 (accessed 12 December 2023)

Arnold T, Tilton L (2019) Distant viewing: analyzing large visual corpora. Digit Scholarsh Hum 34(Supplement_1):i3–i16

Bachelard G (2001) *The Formation of the Scientific Mind: A Contribution to a Psychoanalysis of Objective Knowledge*. 1st edition. Manchester: Clinamen Press Ltd

Banda S, Agarwal A, Rao CR, et al. (2011) Contour layer extraction from colour topographic map by feature selection approach. In: *2011 IEEE Symposium on Computers Informatics*, March 2011, pp. 425–430

Barvir R, Vozenilek V (2020) Developing versatile graphic map load metrics. ISPRS Int J Geo-Inf 9(12):705

Bertin J (1967) *Sémiologie Graphique: Les Diagrammes, Les Réseaux, Les Cartes*. Paris: Gauthier-Villars

Bourillon F (2007) 1807-1850, l'histoire des cadastres urbains. In: *Cent Millions de Parcelles En France: 1807, Un Cadastre Pour l'Empire*. Editions Publi-Topex. Paris, pp. 92–101

Branch J (2014) *The Cartographic State: Maps, Territory, and the Origins of Sovereignty*. 1st edition. New York: Cambridge University Press

Braudel F (1996) *The Mediterranean and the Mediterranean World in the Age of Philip II, Vol. 1*. First Edition. Berkeley, California: University of California Press

Casti E (2005) Towards a Theory of Interpretation: Cartographic Semiosis. Cartogr: Int J Geogr Inf Geovis 40(3):1–16

Clergeot P (2007) *Cent Millions de Parcelles En France: 1807, Un Cadastre Pour l'Empire*. Paris: Editions Publi-Topex

Cook KS (1995) From false starts to firm beginnings: Early colour printing of geological maps. Imago Mundi 47(1):155–172

Cook KS (2015) Photomechanical Processes. *The History of Cartography, Volume 6: Cartography in the Twentieth Century* Monmonier M (ed.). The History of Cartography. Chicago: University of Chicago Press. Available at: https://press.uchicago.edu/ucp/books/book/chicago/H/bo4149962.html (accessed 12 December 2023)

Cordts M, Omran M, Ramos S, et al. (2016) The Cityscapes Dataset for Semantic Urban Scene Understanding. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 3213–3223. Available at: https://openaccess.thecvf.com/content_cvpr_2016/html/Cordts_The_Cityscapes_Dataset_CVPR_2016_paper.html (accessed 13 September 2022)

Crampton JW (2001) Maps as social constructions: power, communication and visualization. Prog Hum Geogr 25(2):235–252

Crampton JW, Elden S (2007) Space, Knowledge and Power: Foucault and Geography. Routledge, Aldershot, England; Burlington, Vermont

Dainville F de, Souchon C, Richard H, et al. (2018) *Le langage des géographes: Termes, signes, couleurs des cartes anciennes*. Illustrated édition. Paris: Comité des travaux historiques et scientifiques - CTHS

Dalal N and Triggs B (2005) Histograms of Oriented Gradients for Human Detection. In: *International Conference on Computer Vision & Pattern Recognition (CVPR '05)*, 20 June 2005, pp. 886–893. IEEE Computer Society. Available at: https://hal.inria.fr/inria-00548512 (accessed 13 June 2020)

Davies M (2012) Expanding horizons in historical linguistics with the 400-million word Corpus of Historical American English. Corpora 7(2):121–157

di Lenardo I, Barman R, Pardini F, et al. (2021) Une approche computationnelle du cadastre napoléonien de Venise. Humanités numériques 3

Ebi N, Lauterbach B, Anheier W (1994) An image analysis system for automatic data acquisition from colored scanned maps. Mach Vis Appl 7(3):148–164

Eco U (1992) La production des signes. biblio essais. Le Livre de Poche, Paris

Edney MH (2019) *Cartography: The Ideal and Its History*. Chicago: University of Chicago Press. Available at: https://press.uchicago.edu/ucp/books/book/chicago/C/bo34094164.html (accessed 5 December 2023)

Feldman H-U (2015) Bundesamt für Landestopographie. *The History of Cartography, Volume 6: Cartography in the Twentieth Century* Monmonier M (ed.). The History of Cartography. Chicago, IL: University of Chicago Press. Available at: https://press.uchicago.edu/ucp/books/book/chicago/H/bo4149962.html (accessed 12 December 2023)

Foucault M (1976) *Les mots et les choses: une archéologie des sciences humaines*. Bibliothèque des sciences humaines. Paris: Gallimard

Gascoigne B (1986) How to Identify Prints: A Complete Guide to Manual and Mechanical Processes from Woodcut to Ink Jet. Thames and Hudson, London

Geels FW (2006) The hygienic transition from cesspools to sewer systems (1840–1930): The dynamics of regime transformation. Res Policy 35(7):1069–1082

Goode JP (1927) The Map as a Record of Progress in Geography. Ann Am Assoc Geogr 17(1):1–14

Hardie A and McEnery T (eds) (2011) Corpus-based studies of synchronic and diachronic variation. In: *Corpus Linguistics: Method, Theory and Practice*. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press, pp. 94–121. Available at: https://www.cambridge.org/core/books/corpus-linguistics/corpusbased-studies-of-synchronic-and-diachronic-variation/D0712A1B1BCC60BF8743DBFF13B82B47 (accessed 11 May 2023)

Harley JB (1989) Deconstructing the map. Cartographica: Int J Geogr Infor and Geovisualization 26(2):1–20

Hausdorff F (1914) *Grundzüge der Mengenlehre*. Leipzig: Leipzig Viet. Available at: http://archive.org/details/grundzgedermen00hausuoft (accessed 2 June 2022)

Heitzler M, Hurni L (2020) Cartographic reconstruction of building footprints from historical maps: A study on the Swiss Siegfried map. Trans GIS 24(2):442–461

Helsel DR, Frans LM (2006) Regional Kendall Test for Trend. Environ Sci Technol 40(13):4066–4073

Hennet A-J-U (1811) *Recueil Méthodique Des Lois, Décrets, Règlemens, Instructions et Décisions Sur Le Cadastre de La France*. Imprimerie impériale, Paris

Herold H (2015) *An Evolutionary Approach to Adaptive Image Analysis for Retrieving and Long-term Monitoring Historical Land Use from Spatiotemporally Heterogeneous Map Sources*. PhD thesis. Technische Universität Dresden, Dresden, Germany. Available at: https://tud.qucosa.de/landing-page/?tx_dlf [id]=https%3A%2F%2Ftud.qucosa.de%2Fapi%2Fqucosa%253A28248%2Fmets (accessed 11 June 2020)

Herrault P-A, Sheeren D, Fauvel M, et al. (2013) Automatic Extraction of Forests from Historical Maps Based on Unsupervised Classification in the CIELab Color Space. In: Vandenbroucke D, Bucher B, and Crompvoets J (eds) *Geographic Information Science at the Heart of Europe*. Lecture Notes in Geoinformation and Cartography. Cham: Springer International Publishing, pp. 95–112. Available at: https://doi.org/10.1007/978-3-319-00615-4_6 (accessed 31 May 2022)

Hiippala T, Bateman JA (2022) Semiotically-grounded distant viewing of diagrams: insights from two multimodal corpora. Digit Scholarsh Hum 37(2):405–425

Hilpert M and Gries STH (2016) Quantitative approaches to diachronic corpus linguistics. In: Kytö M and Pahta P (eds) *The Cambridge Handbook of English Historical Linguistics*. Cambridge Handbooks in Language and Linguistics. Cambridge: Cambridge University Press, pp. 36–53. Available at: https://www.cambridge.org/core/books/cambridge-handbook-of-english-historical-linguistics/quantitative-approaches-to-diachronic-corpus-linguistics/C61092A8B47C5731955DD48D70C9C35F (accessed 11 May 2023)

Hofmann C, Tesson S (2012) Artistes de la carte: De la Renaissance au XXIe siècle, l'explorateur, le stratège, le géographe. Autrement, Paris

Hosseini K, McDonough K, van Strien D et al (2021) Maps of a Nation? The Digitized Ordnance Survey for New Historical Research. J Vic Cult 26(2):284–299

Hosseini K, Wilson DCS, Beelen K, et al. (2022) MapReader: a computer vision pipeline for the semantic exploration of maps at scale. In: *Proceedings of the 6th ACM SIGSPATIAL International Workshop on Geospatial Humanities*, New York, NY, USA, 11 November 2022, pp. 8–19. GeoHumanities '22. Association for Computing Machinery. Available at: https://doi.org/10.1145/3557919.3565812 (accessed 22 November 2022)

Hurni L (2008) Cartographic mountain relief presentation: 150 Years of Tradition and Progress at ETH Zurich. In: *Mountain mapping and visualisation: proceedings of the 6th ICA mountain cartography workshop*, 2008, pp. 11–15. Available at: http://mountaincartography.org/publications/papers/papers_lenk_08/hurni.pdf

Imhof E (1982) Cartographic Relief Presentation. In: *Cartographic Relief Presentation*. Berlin, New York: De Gruyter. Available at: https://www.degruyter.com/document/doi/10.1515/9783110844016/html?lang=en (accessed 12 December 2023)

Jakobson R (1973) *Main Trends in the Science of Language (Routledge Revivals)*. 1st ed. London: Routledge. Available at: https://doi.org/10.4324/9781315831947

Jomard EF, Férussac d'Audebard A-É-J-P-J-F, Jacotin P et al. (1826) Examen de la question de savoir si la Lithographie peut-être appliquées avec avantage à la publication des Cartes Géographiques tant sous le rapport du mérite de l'exécution que sous le rapport de l'économie. Bull de la Société de géogr 5:693–708

Keil J, Edler D, Kuchinke L et al. (2020) Effects of visual map complexity on the attentional processing of landmarks. PLOS ONE 15(3):e0229575

Kendall MG (1938) A New Measure of Rank Correlation. Biometrika 30(1·2):81–93

Klingemann M (2020) Quasimondo/RasterFairy. Python. Available at: https://github.com/Quasimondo/RasterFairy (accessed 13 June 2020)

Korzybski A (1994) *Science and Sanity: An Introduction to Non-Aristotelian Systems and General Semantics*. 5th edition. Brooklyn, N.Y: Institute of General Semantics

Lengyel B, Bokányi E, Di Clemente R et al. (2020) The role of geography in the complex diffusion of innovations. Sci Rep 10(1):15065

Leyk S, Boesch R (2009) Colors of the past: color image segmentation in historical topographic maps based on homogeneity. GeoInformatica 14(1):1

Liu G, Li J, Nie P (2022) Tracking the history of urban expansion in Guangzhou (China) during 1665–2017: Evidence from historical maps and remote sensing images. Land Use Policy 112:105773

Liu T, Miao Q, Tian K et al. (2016) SCTMS: Superpixel based color topographic map segmentation method. J Vis Commun Image Represent 35:78–90

Maaten Lvander, Hinton G (2008) Visualizing Data using t-SNE. J Mach Learn Res 9(Nov):2579–2605

Maule J, Skelton AE, Franklin A (2023) The Development of Color Perception and Cognition. Annu Rev Psychol 74(1):87–111

Mazier F, Broström A, Bragée P et al. (2015) Two hundred years of land-use change in the South Swedish Uplands: comparison of historical map-based estimates with a pollen-based reconstruction using the landscape reconstruction algorithm. Veget Hist Archaeobot 24(5):555–570

Merian M (1638) Tigurum Zürÿch. Frankfurt am Main. Available at: https://doi.org/10.3931/e-rara-69352

MMSegmentation: OpenMMLab Semantic Segmentation Toolbox and Benchmark (2022) OpenMMLab. Available at: https://github.com/open-mmlab/mmsegmentation

Moretti F (2013) Distant Reading. Verso, London; New York

Nominatim (n.d.). Available at: https://nominatim.org/ (accessed 2 June 2022)

Nöth W (1995) Handbook of Semiotics. 1st ed. Advances in Semiotics. Bloomington & Indianapolis: Indiana University Press

Ojala T, Pietikainen M, Maenpaa T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell 24(7):971–987

Olson K (2010) Maps for a New Kind of Tourist: The First Guides Michelin France (1900–1913). Imago Mundi 62(2):205–220

Otsu N (1979) A Threshold Selection Method from Gray-Level Histograms. IEEE Trans Syst Man Cybern 9(1):62–66

Pearson AW (2015) Relief Depiction. The History of Cartography, Volume 6: Cartography in the Twentieth Century Monmonier M (ed.). The History of Cartography. Chicago, IL: University of Chicago Press. Available at: https://press.uchicago.edu/ucp/books/book/chicago/H/bo4149962.html (accessed 12 December 2023)

Petitpierre R (2020) Neural networks for semantic segmentation of historical city maps: Cross-cultural performance and the impact of figurative diversity. Master's thesis. Ecole Polytechnique Fédérale de Lausanne

Petitpierre R (2023) Mapping Memes in the Napoleonic Cadastre: Expanding Frontiers in Memetics. In: Digital Humanities 2023: Book of Abstracts, Graz, Austria, 2023, p. 3. Zenodo

Petitpierre R, Kaplan F and di Lenardo I (2021) Generic Semantic Segmentation of Historical Maps. In: CHR 2021: Computational Humanities Research Conference, Amsterdam, The Netherlands, 17 November 2021. CEUR Workshop Proceedings. Available at: http://ceur-ws.org/Vol-2989/long_paper27.pdf

Razali NM, Wah YB (2011) Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. J Stat Model Anal 2(1):21–33

Ristow WW (1975) Lithography and maps, 1796-1850. In: Woodward D (ed.) Five Centuries of Map Printing. 1st ed. The Kenneth Nebenzahl Jr Lectures in the History of Cartography at the Newberry Library. Chicago; London: The University of Chicago Press, pp. 77–112

Robinson AH (1975) Mapmaking and Mapprinting: The Evolution of a Working Relationship. In: Woodward D (ed.) Five Centuries of Map Printing. 1st ed. The Kenneth Nebenzahl Jr Lectures in the History of Cartography at the Newberry Library. Chicago; London: The University of Chicago Press, pp. 1–23

Rossi P (2005) La carte géologique de la France, historique. Géochronique 96(1):14–15

Royston P (1995) Remark AS R94: A Remark on Algorithm AS 181: The W-test for Normality. J R Stat Soc Ser C (Appl Stat) 44(4):547–551

Salvatore S, Guitton P (2004) Contour line recognition from scanned topographic maps. J WSCG 12(1–3):419–426

Santamaria U, Bailey AM (1984) A Note on Braudel's Structure as Duration. Hist Theory 23(1):78–83

Scharr H (2000) Optimale Operatoren in der Digitalen Bildverarbeitung. PhD thesis. Universität Heidelberg, Heidelberg, Germany. Available at: https://doi.org/10.11588/heidok.00000962

Schulten S (2012) Mapping the Nation: History and Cartography in Nineteenth-Century America. Illustrated edition. The University of Chicago Press

Senefelder A (1818) Vollständiges Lehrbuch der Steindruckerey: enthaltend eine richtige und deutliche Anweisung zu den verschiedenen Manipulations-Arten derselben in allen ihren Zweigen und Manieren, belegt mit den nöthigen Musterblättern: nebst einer vorangehenden ausführlichen Geschichte dieser Kunst von ihrem Entstehen bis auf gegenwärtige Zeit. F.H. Leipold

Solgi RM (2020) geneticalgorithm. Available at: https://github.com/rmsolgi/geneticalgorithm

Soulier and Berdez (1827) Instructions Données Par Le Département Des Finances Pour La Levée Des Plans et l'établissement Du Cadastre, Ensuite de La Décision du Conseil d'Etat Du 6 Décembre 1826. Frères Blanchard. Lausanne

spaCy: Industrial-strength NLP (2022) Spacy. Available at: https://spacy.io/ (accessed 19 June 2023)

Tajima S, Komine K (2015) Saliency-based color accessibility. IEEE Trans Image Process 24(3):1115–1126

Uhl JH, Leyk S, Chiang Y-Y et al. (2018) Map Archive Mining: Visual-Analytical Approaches to Explore Large Historical Map Collections. ISPRS Int J Geo-Inf 7(4):148

Uhl JH, Leyk S, Chiang Y et al (2020) Automated Extraction of Human Settlement Patterns From Historical Topographic Map Series Using Weakly Supervised Convolutional Neural Networks. IEEE Access 8:6978–6996

Uhl JH, Leyk S, Chiang Y-Y et al (2022) Towards the automated large-scale reconstruction of past road networks from historical maps. Comput Environ Urban Syst 94:101794

Vivier N (2007) 1814-1870, les débats pour définir le cadastre. In: Cent Millions de Parcelles En France: 1807, Un Cadastre Pour l'Empire. Editions Publi-Topex. Paris, pp. 63–77

Winther RG (2020) When Maps Become the World. First Edition. Chicago; London: University of Chicago Press

Woodward D (1996) Maps as Prints in the Italian Renaissance: Makers, Distributors & Consumers. The Panizzi lectures 1995. British Library, London

Yuan Y, Chen X and Wang J (2020) Object-Contextual Representations for Semantic Segmentation. In: Computer Vision – ECCV 2020 (eds A Vedaldi, H Bischof, T Brox, et al.), Cham, 2020, pp. 173–190. Lecture Notes in Computer Science. Springer International Publishing

Zhang ZY, Suen CY (1984) A fast parallel algorithm for thinning digital patterns. Commun ACM 27(3):236–239

Author contributions
RP: data collection, methodology, implementation, results analysis and discussion, writing original draft, revising draft. JU: joint supervision, methodology, revising draft. IL: joint supervision, revising draft, funding acquisition. FK: joint supervision, methodology, revising draft.

Competing interests
The authors declare no competing interests.

Ethical approval
This article does not contain any studies with human participants, living organisms, or biological material, performed by any of the authors.

Informed consent
This article does not contain any studies with human participants, living organisms, or biological material, performed by any of the authors.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1057/s41599-024-02840-w.

Correspondence and requests for materials should be addressed to Remi Petitpierre.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.