SocArXiv

Preprint : August 3, 2023 URL/DOI GOES HERE

Feature Engineering for Quantitative Analysis of Cultural Evolution

Fabio Celli
Maggioli Research

Abstract

In this paper we introduce the use of time-resolved variables to represent the evolution of categorical variables through time. Traditionally, the presence or absence of categorical variables are treated as 1 or 0 for computational purposes, and then aggregated with compression techniques, potentially generating biases when the information compressed into different variables is uneven. Our tests reveal that the use of time-resolved variables can help to prevent these biases, to assure higher reliability of the results and to allow an easier explanation of the models.

Keywords: Cliodynamics, Social Evolution, Cultural Evolution, Seshat Databank, Feature Engineering.

1. Introduction and Related work

The fields of cliometrics and quantitative cultural history made big and fast advancements in the last decade. In particular the search for the drivers of social complexity has made huge strides forward from Feinman (2013) to Turchin and Gavrilets (2021). The data-driven analysis of cultural evolution by Turchin et al. (2022) revealed a strong causal role played by a combination of increasing agricultural productivity and adoption of military technologies like iron weapons and cavalry in the bronze age, while the adoption of moralizing religions is reported to have a negative causal effect on societal evolution. These developments have been possible through the availability of historical data suitable for computational analyses, like Seshat Turchin et al. (2020). The basic concept of Seshat is to provide quantitative and structured or semi-structured data about past societies, defined as political units (polities) from 35 sampling points across the globe, in a time window from roughly 10000 BC to 1900 CE and sampled with a time-step of 100 years. Seshat provides many dimensions that report the presence or absence of a cultural trait in a polity at a specific point in time. These fine-grained dimensions are treated with One-Hot Encoding (OHE), a widely used approach for expressing the presence or absence of categorical features into numerical ones represented as 1 or 0. As noted by Ul Haq et al. (2019) among others, OHE yields a high information sparseness in the transformed data, making harder the extraction of models with statistics or machine learning. A solution is to aggregate the One-Hot encoded features into coarse-grained variables with compression techniques like Principal Component Analysis (PCA), as described in Greenacre et al. (2022). However, this might generate biased PCA-transformed variables because some can be high-dimensional, containing information from many fine-grained variables, and others can be low dimensional, containing information from few fine-grained variables. For example the models in Turchin et al. (2022) showed that warfare can explain alone more than half of the evolution in hierarchical complexity ($R^2 = 0.68$) and agriculture productivity could explain alone one third of the improvement on specialization of governance ($R^2 = 0.32$), but the PCA-compressed variable for warfare contains about 40 dimensions, while agriculture productivity is low dimensional.

1.1. Our Proposal

We suggest that, in order to better capture historical dynamics, categorical variables should be expressed as numerical scales representing a development stage of a specific dimension at each level, ordered by the time of the first known appearance of that dimension in history. For example the scale of military technologies could place stone at the first level, copper at the second, bronze at the third, iron as fourth and so on. We will call *time-resolved variables* the features treated in this way. We test whether they are more predictive and more robust to bias with respect to variables treated with OHE and PCA. These are the two hypotheses to test: 1) the dimensionality of OHE-PCA compressed variables influence their predictive power; 2) time-resolved variables are able to explain more variance than the ones processed with OHE and PCA. To do so we annotated with time-resolved variables a new dataset derived from Seshat, that we will call *Chronos*, and we compare it to OHE-PCA compressed variables from Seshat in regression tasks. In what follows we describe the data in detail (Section 2); then we describe the analyses we performed and discuss the results we obtained (Section 3). Finally we include some concluding remarks and we set the agenda for future work (Section 4). Our results show that dimensionality of PCA-compressed variables affects their predictive power and might create a bias. However, given roughly the same dimensionality, the time-resolved variables are more predictive, especially with non-linear modeling.

2. Data

We extracted 188 polities from Seshat and produced two datasets with the same target variables and different independent variables. The first one, that we will refer to as Seshat-pca, contains the following variables:

agriculture technology (agritech): the PCA of fertilizer, cropping, irrigation;

military technology (militech): the PCA of copper, bronze, iron, steel, gunpowder siege artillery;

warfare tools and tactics (warfare): the PCA of small vessels, plate armor, laminar armor, settlement in defensive position, earth ramparts, moats, scaled armor, stone walls, horses, camels, spears, wood barks, leather cloth, shields, helmets, breastplates, limb protection, donkeys, composite bows, battle axes, daggers, swords, modern fortifications, bows, javelins, slings, crossbows, handled firearms, complex fortifications, fortified camps, chain mail tension siege engines, war clubs, elephants, ditch, pole arms, specialized military vessels, stone walls, atlatl;

morality in religion (religion): the PCA of moral enforcement in this life, moralizing enforcement is agentic, moral religion is adopted by elites, moralizing is certain, broad moralizing norms, moral concern is primary, rulers are gods;

economy management (economy): the PCA of tokens, articles, precious metals, indigenous coins, foreign coins, paper currency;

information management (infomedia): the PCA of lists and tables, calendars, sacred text, religious literature, scientific literature, fiction, philosophy, practical literature, history;

writing system (alphabeth): the PCA of phonetic writing, non-phonetic writing, script, mnemonic devices, non-written records;

communication systems: the PCA of couriers, postal stations, general postal service;

infrastructure levels (infrastructure): the PCA of roads, mines and quarries, ports, canals, drinking water supply, irrigation and production systems, bridges, markets, food storage sites;

political system (politics): the PCA of constraints on executive by non-government, constraints on executive by government, legal impeachment.

We annotated the other dataset, that we will call Chronos, with the following time-resolved variables:

Type of agriculture (agritech) encodes the stage of agriculture development through history. At the base level there is *spontaneous* cropping, practiced by nomad and semi-nomad hunter-gatherers societies. At the first level there is the *swidden/slash-and-burn* agriculture, practiced for example by populations of archaic mesoamerica described by Kennett (2012). and possibly started around 6000 BC or earlier. The following level is *fallow* agriculture, a farming technique in which arable land is left without sowing for one or more vegetative cycles. According to Adams et al. (1974), fallow agriculture was practiced in ancient Mesopotamia possibly since 5000 BC.At the third level there are *two fields/crop rotation*, a technique that consists in planting spring crops for livestock in place of grains for human consumption in order to restore or to maintain productive soils. According to Needham and Bray (1984) it has been fully adopted since the Western Zhou empire around 1000 BC.At the fourth level there are *nitrogen-fix/fertilizers*, that start with the use of plants that contribute to nitrogen fixation and evolve into chemical fertilizers invented around 1800. According to Nene (2012), after the collapse of Mauryan Empire, around 180 BC, in the northern half of the Indian subcontinent the farmers discovered that plants need mineral elements such as nitrogen and phosphorus for optimal growth and started to use legumes or soybean as nitrogen fixers.

Technology for military purposes (militech) encodes the level of military technology of a polity through time and is related to the military use of metals. At the first level there are common raw materials used until the neolithic, such as *stone/wood/clay*; At the second level there is *copper*, used since 5000 BC. At the third level there is *bronze*, that started the bronze age around 3000 BC. At the fourth level there is *iron* that started the iron age around 1400 BC. The fifth level is *steel* that appeared in different varieties since 400 BC. The sixth level is the one of *gunpowder*, whose formula appeared in China around 800 but was not used for military purposes until 1044, as described in the military compendium of Wujing Zongyao and reported by Andrade (2017).

Warfare tactics (warfare) encodes the level of military strategy of a polity. At the first level there are *hunters with projectiles*, like bows, javelins, atlatls and slings. At the beginning of the Holocene, in the 11000 BC, these hunting and warfare methods were already well developed according to McEwen et al. (1991). At the second level there are *armies and fortifications*. Possibly this kind of warfare begun around 4500 BC at Yumuktepe (south coast of Turkey) where has been found one of the earliest fortifications in human history Caneva and Sevin (2004), for sure this warfare method was present in bronze age mesopotamia around 3000 BC when the first professional soldiers were found, as reported in Gabriel and Metz (1991). At the third level there are *armies with animals and chariots*, that appear around 1700 BC in the Hittite army and then in the Egyptian army, as reported by Healy (1992). The fourth level is the one of *armies with naval and siege forces*, that appeared by 1100 BC during the bronze age collapse when maritime warfare increased dramatically as discussed by Emanuel (2016). At the fifth level there are *armies with industrial forces*, that started with the American civil war in 1860s, where machine guns and ammunition supply through railroad played a decisive role, and rapidly evolved into the industrial production of tanks and war aircrafts.

Religious system (religion) puts along a scale the types of religions, not just their moral characteristics like in Seshat. At the first level we have the *cult of the dead and spirits*, with evidence of human burials since the Paleolithic, as reported in Harrold (1980). At the second level there is the *cult of ancestors/home/totem*, begun around 8500 BC with the secondary mortuary practices of the Natufian culture that, , according to Kuijt (1996), was a way to demonstrate a lineage was a powerful means of social integration of nomad tribes. At the third level there is the *cult of fertility*, characterized by a Mother Goddess, a Sky Father and figurines of fertile women. By 4800 BC the cult of Mother Earth and Father Sky was established in Ubaid with the holy wedding between the king and the goddess Inanna, as described by Harrison (2013) and Schneider (2011). The fourth stage is *polytheism*. According to Rosalie (2002) each region of Egypt originally had its own patron deity but, as communities were conquered or absorbed each other, the god of the defeated area was incorporated into the winner's tradition. By 3100 BC this process resulted in polytheism and a collective mythology that put the basis for incorporating new cultures into large empires such as the Greek and Roman ones. At the fifth level there is *monotheism* that, according to Albertz (1994), started with the monotheistic worship of Yahweh in ancient Israel at the time between prophets Elijah and Hosea, around 830 BC. and surely after the captivity in Babylon after 600 BC. The sixth level is the one of *philosophies*, that started with the diffusion of Jainism around 600 BC in India and followed within a few centuries by Buddhism, Confucianism and Taoism. Also the Western tradition developed its own philosophical schools. It started around the same time with Thales of Miletus, but only centuries later some of them developed into spiritual practices, for example Stoicism, Platonism and Epicureanism.

Economic level (economy), encodes the most advanced economic type of a polity. At the base level there are *subsistence and exogamy*, the minimal necessary form of economic exchange in small communities. According to Dow et al. (2016) exogamy rate is higher when community sizes is small, but also when productivity differs across communities. At the first level there are *precious goods and livestock*. Despite archaeological and ethnographical evidence about bartered or token-based monetary transactions are inconclusive, Khalaily and Valla (2013) detected an early obsidian trade route from Anatolia to the site of Eynan-Ain Mallaha in Israel around 10200 BC. The later developments of this goods exchange system can be found in the form of barter in the palatial economy of Mesopotamia and Egypt (food credit in exchange of labor) as reported by Svizzero and Tisdell (2019), as well as in the periodic livestock markets of chalcolithic described by Arbuckle (2012). The second level is the one of *metals and weights*. According to Ialongo et al. (2021) the introduction of weighing technology around 3000 BC between Mesopotamia and Egypt, and later widely adopted in Western Eurasia, allowed merchants to rely on an objective reference to quantify economic value of goods. Moreover, using cosine quantogram analysis and Monte Carlo simulations, Ialongo and Lago (2021) showed that metal fragments comply with weight systems and conclude that metal fragments were likely used as money in the Atlantic Bronze age Europe. The third level is the one of *coinage*, started in Lydia around 600 BC when king Alyattes minted coins made of electrum. At the fourth level we have *paper money*, first introduced in 1024 and derived from coin deposit receipts given by trusted merchants. At the fifth level there is the *stock market*, that was created in Amsterdam in 1611 when the Dutch East India Company, in need to raise capitals, decided to sell stock and pay dividends of the shares to investors.

Information management (infomedia) encodes the level of contents that a culture can mediate and diffuse for information management. The basic stage of this scale is *no information record* that includes all the pre-pottery neolithic polities of hunter-gatherers for which we have evidence of material culture, like arrowheads, but not a mediated content. The first step in the scale is *tradition/symbolism*, that appeared at Göbekli Tepe by 9000 BC. The appearence of symbolism is debated: according to Mithen et al. (2022) it has to do with ideology. The second level is *administration*, with mediated lists, laws or calendars. Although the first seal stamps appear in Ubaid by 5400 BC, their use for administrative purposes is attested since around 4000 BC when the earliest-known cuneiform documents were produced at the sacred temple precinct in the city-state of Uruk, as described by Woods et al. (2010). The third level of content mediation is the one of *religious/philosophical/scientific texts*. According to Willard (2008) instructional literature, letters, monument dedication prayers and even medicine essays are known to exist in ancient Egypt at the introduction of hieroglyphs around 3000 BC. At The fourth level we find the *literature/fiction* contents that, according to Tigay (2002), appears around 2100 BC with the epic of Gilgamesh. At the fifth level of mediated content are the *news/opinions*, that appear in Germany in 1605 with the first newspaper, as reported in Weber (2006), and from there in all Europe within few decades.

Writing system (alphabeth) encodes the evolution of the writing system. At the base level there is no writing but just mnemonic devices. The first stage is the *logographic* writing, appeared with Mesopotamian pictographs and Egyptian hieroglyphs around 3500 BC. At the second stage there is the *syllabic* alphabeth, started with cuneiform around 2500 BC, as reported by van Soldt (2010). The third stage is the *phonetic* alphabeth, whose development started in Ugarit around 1500 BC and completed in Phoenicia by 1100 BC.

Communication systems (communication) encodes the evolution of long-distance communication. At the first level there are *couriers* that, according to Dee et al. (2013), appear after 3500 BC as Egypt begins to develop the first form of territorial administration. At the second level there are *postal stations* for long-distance couriers, introduced around 2000 BC by king Shulgi of the third dynasty of Ur, as reported by Postgate (2017). The third level is a *centralized postal service*, that, although debated, possibly appeared for the first time after 200 BC in western Han China, as reported by Lien (2015). The fourth level is *electric narrowcasting*, introduced with the telegraph around 1800.

Infrastructure Level (infrastructure) encodes the level of control over infrastructures of a polity. The first level is the one of *routes and quarries* that, according to Aurenche et al. (2001), was already present by 10500 BC with an obsidian trade route from Anatolia to the Natufian site of Eynan-Ain Mallaha in Israel. At the second level there are *common or special buildings*. As suggested by Mazurowski et al. (2009), in Tell Qaramel, one of the oldest settlements of the holocene, after 10000 BC appear 3 circular towers, the oldest known on earth, that possibly served as shrine or common house. At the third level there are *irrigation and production systems*. Although the exact date of the first irrigation system is not known, It developed on large scale in Mesopotamia during Ubaid period 2, around 4800 BC, as reported by Kurt (1996). The polities based mainly on animal husbandry instead developed productions systems for wool and other animal-derived materials, as reported by Makarewicz et al. (2017) about central Anatolia sites around 4000 BC. The fourth level of infrastructure is *urbanization and markets*, that appeared around 3100 BC in Dynastic Egypt where, according to Altenmüller (2001), the goods exchanged were primarily consumables (beer, bread, dried meat/fish, vegetables and fruits) with non-consumables like household artifacts. At the fifth level there are *portual systems* for goods supply. According to Knapp (1993) this infrastructural development started around 1500 BC with the first attempts towards thalassocracy of the Minoan, Caananite (Phoenician) and Egyptian polities, that finally, coupled with iron weapons an chariot warfare, yield to the bronze age collapse. The sixth and latest level includes *railways and telecommunications*, begun around 1850s and is the main infrastructural level today.

Political system (politics) encodes the development of the limits that are imposed to the rulers. At the base level there is a *sole ruler*, and no limits to the ruler's power. According to Boehm (2012) this condition is the norm for the primates with social dominance hierarchies such as bonobos, chimpanzees and humans as well. However, Turchin (2016) argues that early human communities with projectile weapons were able to put a limit to the alpha male power and develop egalitarian societies, where decisions were made in a collective *assembly*. This is our first level and the presence of common houses like the one in Tell Qaramel can be considered an evidence of the presence of assemblies by 10000 BC or even before. At the second level we have *representatives* of the population, that are needed in large societies to put a constraint on the decisions taken by aristocratic assemblies and are the foundation of democracies. According to Smith (2012) the first representatives of the population were introduced in the Republican Rome with the Tribunes of the Plebs around 470 BC. The third level is the on of *legal impeachment*, that allows a parliament to legally remove the powers from a ruler without physically killing him. It is difficult to determine exactly when it started, as around 1700, the people of the Ashante empire of Ghana had a longstanding tradition of destooling a king that has been cruel or ineffective. Possibly it has been a part of their cultural and political practices for centuries, as described in Obeng (1996).

Finally, we have the same target variables for both datasets:

Social scale (SCALE), is the PCA of the log-transformed polity population, polity territory and the population of the largest settlement.

Hierarchical complexity (HIER), is the PCA of the raw count of levels in administrative, military and settlement hierarchies.

Specialization of governance (GOV), is the PCA of the following 11 One-Hot encoded variables: professional officers, soldiers, priests, lawyers, full-time bureaucrats, specialized buildings for government, examination system, merit promotion, formal legal code, full-time judges, courts.

3. Experiments and Discussion

In order to test hypothesis 1, that the number of fine-grained variables compressed with PCA influences its predictive power, we predicted the target variables with a high dimensional PCA (warfare) and low-dimensional PCA (militech) in Sesaht. We use linear regression with 80% training and 20% test split as evaluation settings. Both variables are from the same macro category and the results, reported in Table 1, show that warfare, the high-dimensional PCA-reduced variable, has always the best performance. This result clearly confirms our hypothesis and is in line with the R^2 reported by Turchin et al. (2022).

dataset	**variables**	**algorithm**	**target**	R^2
seshat-pca-warfare	38	linear regression	GOV	**0.638**
seshat-pca-warfare	38	linear regression	SCALE	**0.695**
seshat-pca-warfare	38	linear regression	HIER	**0.645**
seshat-pca-militech	5	linear regression	GOV	0.515
seshat-pca-militech	5	linear regression	SCALE	0.522
seshat-pca-militech	5	linear regression	HIER	0.455

Table 1: Regression of target variables with a high dimensional PCA (warfare) and low-dimensional PCA (militech) in Sesaht. The best results are marked in bold

To test hypothesis 2, namely that time-resolved variables are able to explain more variance than the ones processed with OHE and PCA, we performed a regression task with 80% training and 20% test split as evaluation settings. comparing Seshat-pca and Chronos. We also compared the performance with linear modeling (linear regression) and non-linear modeling (random forest regression).

dataset	**algorithm**	**target**	R^2
chronos	linear regression	GOV	0.771
chronos	linear regression	SCALE	0.780
chronos	linear regression	HIER	**0.815**
chronos	random forest	GOV	0.709
chronos	random forest	SCALE	**0.861**
chronos	random forest	HIER	**0.877**
chronos	avg	all	**0.800**
seshat-pca	linear regression	GOV	**0.783**
seshat-pca	linear regression	SCALE	**0.809**
seshat-pca	linear regression	HIER	0.776
seshat-pca	random forest	GOV	**0.789**
seshat-pca	random forest	SCALE	0.860
seshat-pca	random forest	HIER	0.800
seshat-pca	avg	all	0.777

Table 2: Regression of target variables with all features in Chronos and Seshat-pca. The best results are marked in bold.

feature	**dimensions/levels**	**SCALE**	**GOV**	**HIER**
chronos-warfare	5	**0.855****	0.780**	0.808**
chronos-infomedia	5	0.773**	0.817**	0.820**
chronos-infrastructure	6	0.806**	0.794**	**0.818****
chronos-communication	4	0.777**	0.754**	**0.828****
chronos-economy	5	**0.829****	**0.761****	**0.805****
chronos-militech	6	**0.756****	**0.700****	**0.740****
chronos-religion	6	**0.766****	**0.785****	**0.753****
chronos-alphabeth	3	0.610**	0.532**	0.659**
chronos-agritech	4	**0.331****	**0.362****	**0.331****
chronos-politics	4	0.007**	0.016**	0.001**
seshat-pca-warfare	38	0.813**	**0.852****	**0.820****
seshat-pca-infomedia	9	**0.826****	**0.846****	**0.831****
seshat-pca-infrastructure	9	**0.853****	**0.814****	0.804**
seshat-pca-communication	3	**0.824****	**0.827****	0.742**
seshat-pca-economy	6	0.737**	0.732**	0.697**
seshat-pca-militech	5	0.665**	0.691**	0.710**
seshat-pca-religion	7	0.727**	0.689**	0.711**
seshat-pca-alphabeth	5	**0.758****	**0.782****	**0.783****
seshat-pca-agritech	3	0.138*	0.211*	0.214
seshat-pca-politics	3	**0.302***	**0.247**	**0.161****

Table 3: Spearman correlations between the features in the two datasets and the target variables. *=p-value <0.005; **=p-value <0.001. The best correlations are marked in bold.

The results, reported in Table 2 show that the time-resolved variables of the Chronos dataset, by average, perform best, but only with non-linear modeling and for the target variables that are not computed from OHE, namely HIER and SCALE. To better understand this behavior we performed a Spearman correlation analysis between all the features and the target classes in Chronos and Seshat-pca. The results, reported in Table 3, show at least three interesting phenomena: 1) the p-values of the time-resolved variables in the Chronos dataset are always below 0.001, unlike in Seshat-pca; 2) the best correlations are equally distributed among the features of the two datasets, although the levels of the time-resolved variables in Chronos are, by average, less than the dimensions in the PCA-compressed variables in Seshat. 3) the p-values indicate that there are weak correlations only with low dimensional PCA-compressed variables, namely *seshat-pca-agritech* and *seshat-pca-politics*.

It is interesting to note that, given roughly the same size of the dimensions and levels, the time-resolved variables tend to have higher correlation strength to the target variables, like in the case of *chronos-economy*, *chronos-militech*, *chronos-religion* and *chronos-agritech*. An exception is *chronos-communication*, that has a lower correlation strength with respect to *seshat-pca-communication*. This means that the evolutive scales of time-resolved variables fit the evolution of social complexity summarizing the information better than OHE and PCA. We suggest that this might have a positive impact on model explanation. Crucially, the semantic relations between levels of the time-resolved variables can be defined easily, for example the ability to work copper enables the ability to work bronze, that is an alloy of copper and tin; and the ability to manipulate symbols as seal stamps paves the way to the use of cuneiform writings for administrative information management.

The fact that p-values are always below 0.001 with the time-resolved variables in Chronos, but not with the low-dimensional variables in Seshat, indicates that time-resolved variables have a higher reliability with respect to the variables in Seshat. This is due to the design of the time-resolved variables, that requires to define the scope of the phenomenon we want to observe and align its evolutive dimensions along a time-based scale. Crucially, the *chronos-politics* variable has a very low correlation strength, but the p-value is steadily below 0.001, while seshat-pca-politics has higher correlation strength but p-value higher than 0.005. This might indicate that a time-resolved variable is not the best way to describe politics in history, because political systems have a high variability and high social scale or government specialization can be reached by sole rulers as well as parliaments.

In general, One-Hot encoded PCA-compressed variables showed a better performance in the prediction of the specialization of government, that is a target variable obtained from OHE and PCA of 11 dimensions. Among these 11 dimensions there are some regarding law, professions and merit manegement. The concept of government specialization is defined in a broad way and it is possible that time-resolved variables, that are designed to describe more narrow concepts, are not very effective in this case.

4. Conclusion and Future Work

We introduced time-resolved variables for the computational analysis of cultural evolution and compared their performance to OHE PCA variables in a regression task. We suggest that the use of time-resolved variables can help to prevent biases in the definition of the observed dimensions, to assure higher reliability of the results and to allow an easier explanation of the models. We suggest that having roughly the same dimensionality in all the variables used for comparisons should be adopted as a best practice in order to prevent potential biases.

For the future we plan to expand the Chronos dataset from 11000 BC to 2020 with a time step of 10 years. We plan to annotate it with time-resolved variables, when possible. We also plan to explore further the reasons why the politics dimension has a low predictive power if expressed as a time-resolved variable. In general, we are interested to understand under which conditions it is convenient to use time-resolved variables and also finding other effective ways to manage information for the computational analysis of cultural evolution.

References

Adams, R. M. et al. (1974). Historic patterns of mesopotamian irrigation agriculture. *Irrigation's impact on society*, pages 1–6.

Albertz, R. (1994). *A History of Israelite Religion in the Old Testament Period: From the beginnings to the end of the monarchy*, volume 1. Westminster John Knox Press.

Altenmüller, H. (2001). Old kingdom: Fifth dynasty. *The Oxford Encyclopedia of Ancient Egypt*, 2:601.

Andrade, T. (2017). *The gunpowder age: China, military innovation, and the rise of the West in world history*. Princeton University Press.

Arbuckle, B. S. (2012). Animals and inequality in chalcolithic central anatolia. *Journal of Anthropological Archaeology*, 31(3):302–313.

Aurenche, O., Galet, P., Régagnon-Caroline, E., and Évin, J. (2001). Proto-neolithic and neolithic cultures in the middle east—the birth of agriculture, livestock raising, and ceramics: A calibrated 14c chronology 12,500-5500 cal bc. *Radiocarbon*, 43(3):1191–1202.

Boehm, C. (2012). Ancestral hierarchy and conflict. *Science*, 336(6083):844–847.

Caneva, I. and Sevin, V. (2004). *Mersin-Yumuktepe a Reappraisal*, volume 12. Congedo.

Dee, M., Wengrow, D., Shortland, A., Stevenson, A., Brock, F., Girdland Flink, L., and Bronk Ramsey, C. (2013). An absolute chronology for early egypt using radiocarbon dating and bayesian statistical modelling. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 469(2159):20130395.

Dow, G. K., Reed, C. G., and Woodcock, S. (2016). The economics of exogamous marriage in small-scale societies. *Economic Inquiry*, 54(4):1805–1823.

Emanuel, J. P. (2016). Maritime worlds collide: agents of transference and the metastasis of seaborne threats at the end of the bronze age. *Palestine Exploration Quarterly*, 148(4):265–280.

Feinman, G. M. (2013). The emergence of social complexity. *Cooperation and collective action: Archaeological perspectives*, pages 35–56.

Gabriel, R. A. and Metz, K. S. (1991). *From Sumer to Rome: The Military capabilities of ancient armies*. Number 108. ABC-CLIO.

Greenacre, M., Groenen, P. J., Hastie, T., d'Enza, A. I., Markos, A., and Tuzhilina, E. (2022). Principal component analysis. *Nature Reviews Methods Primers*, 2(1):100.

Harrison, S. (2013). *Mother Earth, Father Sky*. Open Road Media.

Harrold, F. B. (1980). A comparative analysis of eurasian palaeolithic burials. *World Archaeology*, 12(2):195–211.

Healy, M. (1992). *Armies of the Pharaohs*. Osprey.

Ialongo, N., Hermann, R., and Rahmstorf, L. (2021). Bronze age weight systems as a measure of market integration in western eurasia. *Proceedings of the National Academy of Sciences*, 118(27):e2105873118.

Ialongo, N. and Lago, G. (2021). A small change revolution. weight systems and the emergence of the first pan-european money. *Journal of Archaeological Science*, 129:105379.

Kennett, D. J. (2012). Archaic-period foragers and farmers in mesoamerica. *The Oxford Handbook of Mesoamerican Archaeology*, pages 141–150.

Khalaily, H. and Valla, F. R. (2013). Obsidian in natufian context: The case of eynan (ain mallaha), israel. *The Natufian Foragers in the Levant. Terminal Pleistocene Social Changes in Western Asia*, pages 193–202.

Knapp, A. B. (1993). Thalassocracies in bronze age eastern mediterranean trade: making and breaking a myth. *World Archaeology*, 24(3):332–347.

Kuijt, I. (1996). Negotiating equality through ritual: a consideration of late natufian and prepottery neolithic a period mortuary practices. *Journal of Anthropological Archaeology*, 15(4):313–336.

Kurt, A. (1996). Ancient near east v1.

Lien, Y. E. (2015). Reconstructing the postal relay system of the han period. In *A history of Chinese letters and epistolary culture*, pages 15–52. Brill.

Makarewicz, C. A., Arbuckle, B. S., and Öztan, A. (2017). Vertical transhumance of sheep and goats identified by intra-tooth sequential carbon (δ13c) and oxygen (δ18o) isotopic analyses: evidence from chalcolithic köşk höyük, central turkey. *Journal of Archaeological Science*, 86:68–80.

Mazurowski, R. F., Michczyńska, D. J., Pazdur, A., and Piotrowska, N. (2009). Chronology of the early pre-pottery neolithic settlement tell qaramel, northern syria, in the light of radiocarbon dating. *Radiocarbon*, 51(2):771–781.

McEwen, E., Miller, R. L., and Bergman, C. A. (1991). Early bow design and construction. *Scientific American*, 264(6):76–83.

Mithen, S., Richardson, A., and Finlayson, B. (2022). The flow of ideas: Shared symbolism between wf16 in the south and göbekli tepe in the north during neolithic emergence in south-west asia. *Antiquity*.

Needham, J. and Bray, F. (1984). *Science and Civilisation in China, Part 2, Agriculture*, volume 6. Cambridge University Press.

Nene, Y. (2012). Significant milestones in evolution of agriculture in the world. *Asian Agricultural History*, 16:219–35.

Obeng, J. P. (1996). *Asante Catholicism: Religious and cultural reproduction among the Akan of Ghana*, volume 15. Brill.

Postgate, N. (2017). *Early Mesopotamia: society and economy at the dawn of history*. Taylor & Francis.

Rosalie, D. (2002). Religion and magic in ancient egypt.

Schneider, T. J. (2011). *An Introduction to Ancient Mesopotamian Religion*. Wm. B. Eerdmans Publishing.

Smith, C. (2012). The origins of the tribunate of the plebs. *Antichthon*, 46:101–125.

Svizzero, S. and Tisdell, C. (2019). Barter and the origin of money and some insights from the ancient palatial economies of mesopotamia and egypt.

Tigay, J. H. (2002). *The evolution of the Gilgamesh epic*. Bolchazy-Carducci Publishers.

Turchin, P. (2016). *Ultrasociety: How 10,000 years of war made humans the greatest cooperators on earth.* Beresta Books Chaplin, CT.

Turchin, P. and Gavrilets, S. (2021). Tempo and mode in cultural macroevolution. *Evolutionary Psychology*, 19(4).

Turchin, P., Whitehouse, H., François, P., Hoyer, D., Alves, A., Baines, J., Baker, D., Bartokiak, M., Bates, J., Bennet, J., et al. (2020). An introduction to seshat: Global history databank. *Journal of Cognitive Historiography*, 5:115–123.

Turchin, P., Whitehouse, H., Gavrilets, S., Hoyer, D., François, P., Bennett, J. S., Feeney, K. C., Peregrine, P., Feinman, G., Korotayev, A., et al. (2022). Disentangling the evolutionary drivers of social complexity: A comprehensive test of hypotheses. *Science Advances*, 8(25):eabn3517.

Ul Haq, I., Gondal, I., Vamplew, P., and Brown, S. (2019). Categorical features transformation with compact one-hot encoder for fraud detection in distributed environment. In *Data Mining: 16th Australasian Conference, AusDM 2018, Bahrurst, NSW, Australia, November 28–30, 2018, Revised Selected Papers 16*, pages 69–80. Springer.

van Soldt, W. H. (2010). The interaction of syllabic and alphabetic cuneiform writing in ugarit. *The Idea of Writing: Play and Complexity. Leiden*, pages 345–355.

Weber, J. (2006). Strassburg, 1605: The origins of the newspaper in europe. *German history*, 24(3):387–412.

Willard, R. H. (2008). *Weights and Measures in Egypt*, pages 2244–2251. Springer Netherlands, Dordrecht.

Woods, C. et al. (2010). The earliest mesopotamian writing. *Visible language: Inventions of writing in the ancient Middle East and beyond*, pages 33–50.

Affiliation:

Fabio Celli
Maggioli Research
Santarcangelo, Italy
E-mail: fabio.celli@maggioli.it
URL: https://scholar.google.it/citations?hl=en&user=ByZnr0sAAAAJ

SocArXiv Website https://socopen.org/
SocArXiv Preprints https://osf.io/preprints/socarxiv
Preprint *Submitted:* August 3, 2023
URL/DOI GOES HERE *Accepted:* August 3, 2023