Point Processes and Spatiotemporal Modeling of Alphabet Spread in Ancient Italy

Presenter: Tom Francis (tdfrancis@ucla.edu) | Contributor: Muhammad Rehan (rehanmuh@g.ucla.edu)

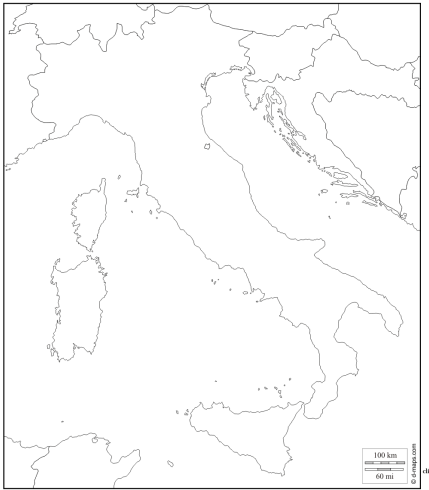
Classics and PIES
University of California, Los Angeles

DataX Stacks Xchange

May 18, 2026

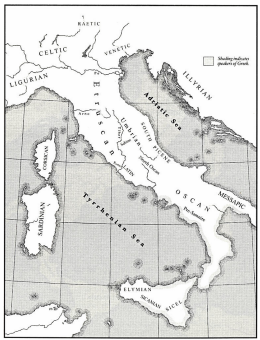
Overview

What languages are attested in the inscriptional record in 1st millennium BCE Italy?



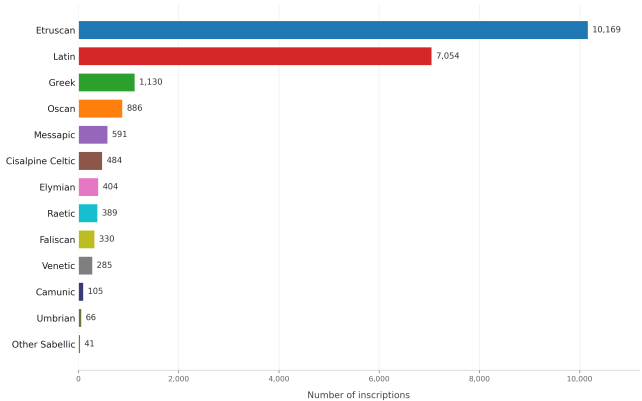
Languages attested in the inscriptional record in 1st millennium BCE Italy

A. Language map of ancient Italy

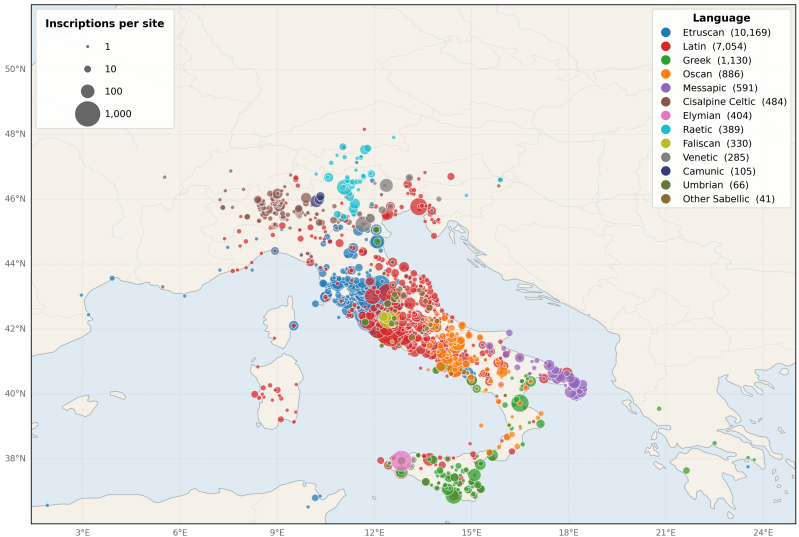


Wallace 2007:x

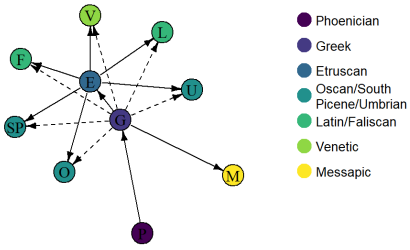
B. Language distribution in the dataset



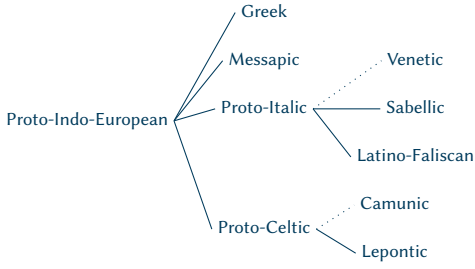
Spatial distribution of languages in the epigraphic record



Linguistic relationships vs. alphabet relationships



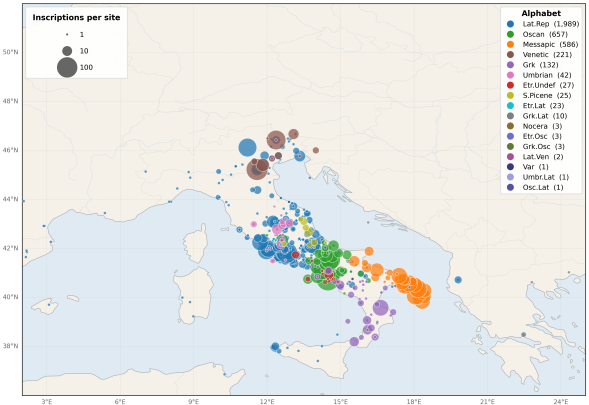
Generally accepted direction of alphabet spread in Italy



Phylogeny of Indo-European languages in Ancient Italy

The need for a new dataset

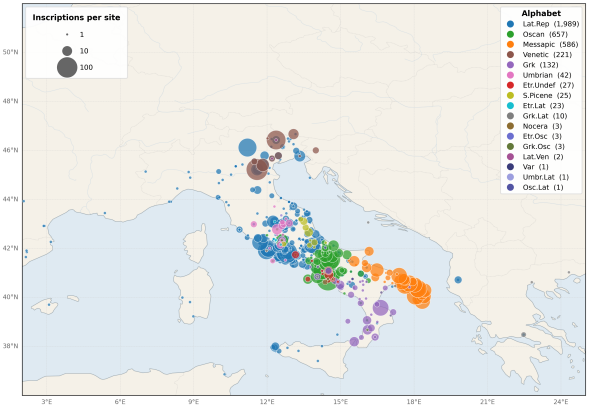
- Reuben Pitts published the *Corpus of the Epigraphy of the Italian Peninsula in the 1st Millennium* BCE (*CEIPoM*) in 2022



Alphabet heatmap from Pitts' 3,868-inscription dataset for 1st-millennium BCE Italy

The need for a new dataset

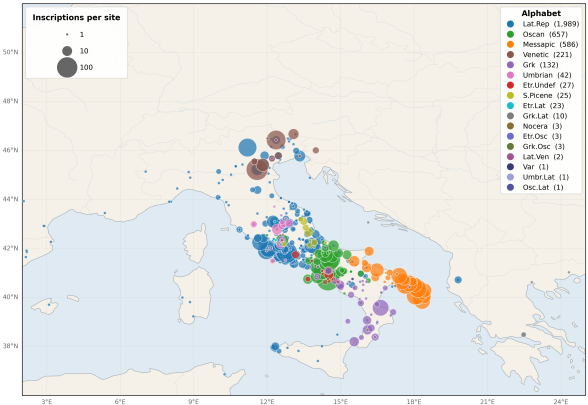
- Reuben Pitts published the *Corpus of the Epigraphy of the Italian Peninsula in the 1st Millennium* BCE (*CEIPoM*) in 2022
- Excludes inscriptions on islands, Celtic and non-Indo-European languages (e.g., Etruscan), and Latin after 100 BCE



Alphabet heatmap from Pitts' 3,868-inscription dataset for 1st-millennium BCE Italy

The need for a new dataset

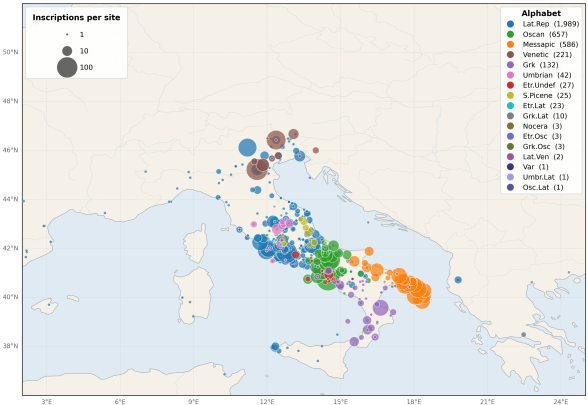
- Reuben Pitts published the *Corpus of the Epigraphy of the Italian Peninsula in the 1st Millennium* BCE (*CEIPoM*) in 2022
- Excludes inscriptions on islands, Celtic and non-Indo-European languages (e.g., Etruscan), and Latin after 100 BCE
- Emphasis on linguistic information leads to conflation of alphabets (Faliscan = Latin), etc.



Alphabet heatmap from Pitts' 3,868-inscription dataset for 1st-millennium BCE Italy

The need for a new dataset

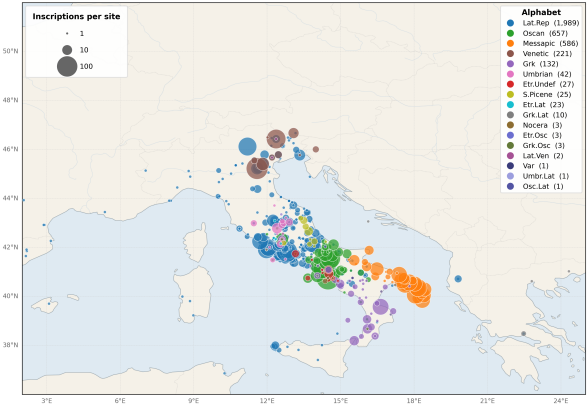
- Reuben Pitts published the *Corpus of the Epigraphy of the Italian Peninsula in the 1st Millennium* BCE (*CEIPoM*) in 2022
- Excludes inscriptions on islands, Celtic and non-Indo-European languages (e.g., Etruscan), and Latin after 100 BCE
- Emphasis on linguistic information leads to conflation of alphabets (Faliscan = Latin), etc.
- Inconsistent encoding processes: inscriptions encoded twice (e.g., CIL X 289) and with different coordinates for the same specified location (Diano, Dimale, Gabii Castiglione, etc.)



Alphabet heatmap from Pitts' 3,868-inscription dataset for 1st-millennium BCE Italy

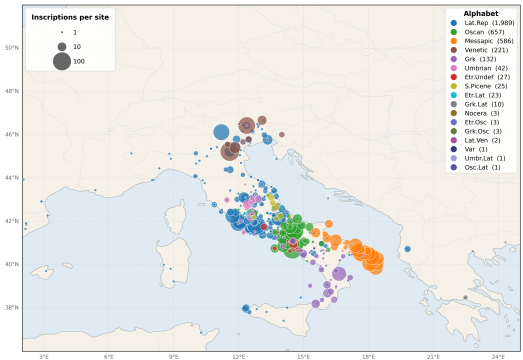
The need for a new dataset

- Reuben Pitts published the *Corpus of the Epigraphy of the Italian Peninsula in the 1st Millennium BCE* (*CEIPoM*) in 2022
- Excludes inscriptions on islands, Celtic and non-Indo-European languages (e.g., Etruscan), and Latin after 100 BCE
- Emphasis on linguistic information leads to conflation of alphabets (Faliscan = Latin), etc.
- Inconsistent encoding processes: inscriptions encoded twice (e.g., CIL X 289) and with different coordinates for the same specified location (Diano, Dimale, Gabii Castiglione, etc.)
- Notable errors, including Etruscan and Hernican inscriptions in Anagni listed as Oscan language and alphabet



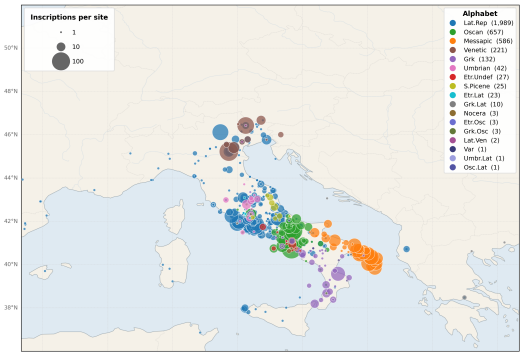
Alphabet heatmap from Pitts' 3,868-inscription dataset for 1st-millennium BCE Italy

Comparison of alphabets in the epigraphic record between databases

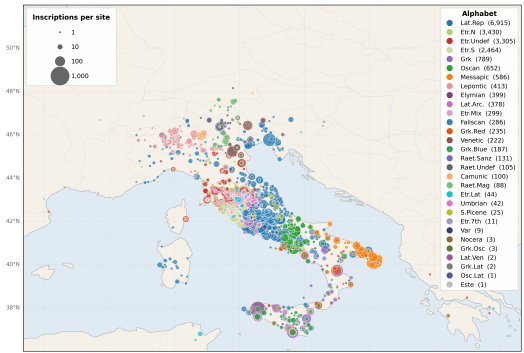


Pitts' 3,868-inscription dataset

Comparison of alphabets in the epigraphic record between databases



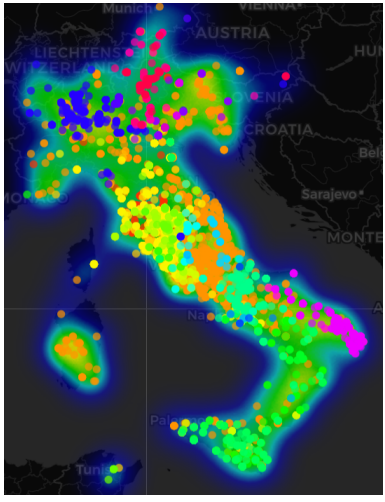
Pitts' 3,868-inscription dataset



Current 21,937 inscription dataset

Data collection for the model: variables

- Each row *i* in the corpus encodes the attestation of one *inscription event*.
- We record five attributes for every inscription:
  - `Date.Start`, `Date.End` (calendar years)
  - `Latitude`, `Longitude` (decimal degrees)
  - `Alphabet` (integer label, e.g. 1 = S.Etruscan)
  - `Region`
  - `Object.Type` (genre)
  - `Site_Id` derived from rounded (lat, lon)
  - Fields under construction for further analyses: Archaeological Context (including whether or not it exists), Inscription Type, Alphabet Direction, Character States for Individual Letters

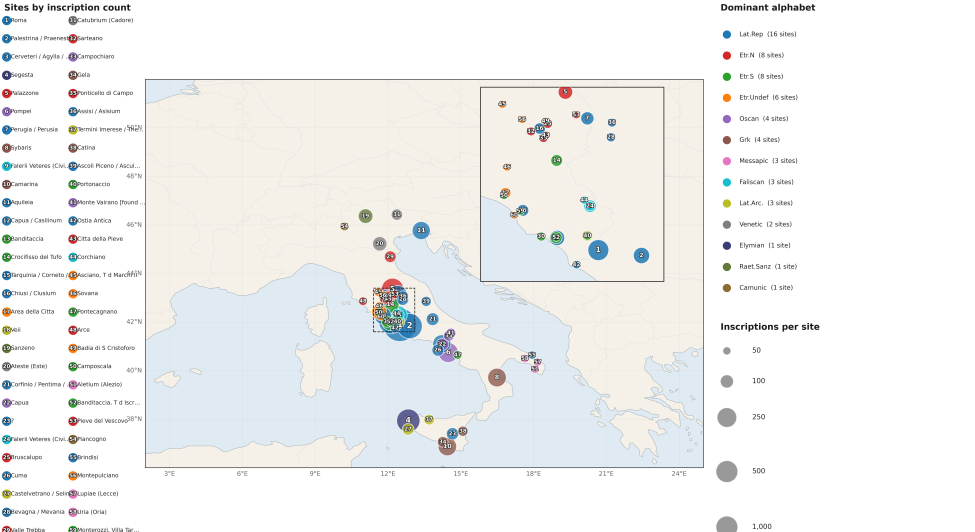


Spatial density of all inscriptions

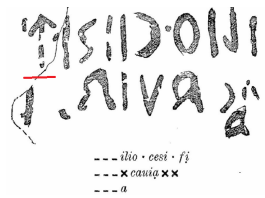
Data collection for the model: major editions and databases consulted

Language	Primary Edition(s) Used	Database(s) Used
Etruscan	Rix and Meiser 2014	N/A
Greek	Dubois 1989, 1995, 2002, 2008	I.Sicily
Latin	*CIL* I^2, III, VI, XI	*CEIPoM*, *EDCS*
Elymian	Agostiniano 1977, 2021	I.Sicily
Raetic	Volumes cited in *TIR*	*TIR*
Lepontic	Volumes cited in *Lexicon Leponticum*	*Lexicon Leponticum*
Cisalpine Gaulish	Volumes cited in *Lexicon Leponticum*	*Lexicon Leponticum*
Sabellic Languages	Crawford et al. 2011, Rix 2002	*CEIPoM*
Faliscan	Bakkum 2009	*CEIPoM*
Messapic	De Simone and Marchesini 2002	*CEIPoM*
Camunic	Tibiletti Bruno 1990, Prosdocimi 1965	N/A
Venetic	Lejeune 1974, Prosdocimi and Pellegrini 1967	*CEIPoM*

Alphabet distribution: major epigraphic centers (N ≥50)

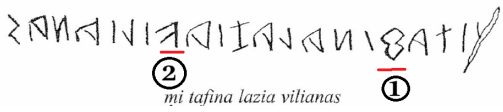


<F> in the Italic chat: distinguishing between alphabets used by different languages in the same area: Ager Faliscus



CIE 8215 (Faliscan)

CIE 8255 (Latin)



CIE 8901 (South Etruscan)

Bonus: *Imagines Italicae* Interamnia Praetuttiorum 2

(South-Picene, not in Ager Faliscus)

Distinguishing between alphabets used by the same language: Etruscan

REE 37:340.2; Bronze colander, Orvieto, 4th–3rd c. BCE; sinistroverse /ʃut^{h}ina/ ‘tomb goods’

REE 51:271.176; Attic style vase, Cerveteri?, 5th c. BCE?

REE 69:388.84; Tomb inscription, Siena, after 400 BCE

Table 14.1 Regional Spelling of the Sibilants (/s/ and /ʃ/) and the Velar Stop (/k/).

Regions	/s/	/ʃ/	/k/
Northern (Chiusi)	M	≀	ϗ
Central (Tarquinia)	≀	M	ↄ
Southern (Caere and Veii)	≀	⋛	ↄ

Wallace 2016:206; exceptions are common; Veii also has X /s/

Distinguishing between alphabets used by the same language: Raetic

	alpha	epsilon	waw	zeta	heta	theta	iota	kappa	lambda	mu
M	𐗀	𐗁	𐗂	𐗃	𐗄	𐗅	𐗆	𐗇	𐗈	𐗉
S	𐗊	𐗋	𐗌	–	𐗍	𐗎	𐗏	𐗐	𐗑	𐗒
	a	*e*	*v*	*z*	*h*	*θ*	*i*	*k*	*l*	*m*

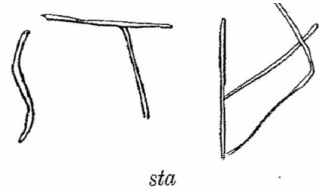
	nu	pi	san	rho	sigma	tau	/tᵘ/	upsilon	phi	chi
M	𐗓	𐗔	𐗕	𐗖	𐗗	𐗘	𐗙	𐗚	𐗛	𐗜
S	𐗓	𐗔	𐗕	𐗖	𐗗	𐗘	𐗙	𐗚	𐗛	𐗜
	n	*p*	*ś*	*r*	*s*	*t*	*ṗ*	*u*	*φ*	*χ*

Table 1. The characters of the Magrè (M) and Sanzeno (S) alphabets (standardised and sinistroverse) with transliteration letters.

Figure: Distinctions of the two primary Raetic alphabets from Salomon 2020. The sequence *-nuale* from two inscriptions; characters from *Thesaurus Inscriptionum Raeticarum*

Encoding problems: short inscriptions

Lexicon Leponticum BG-28.1: /pit/ or /lit/?



CIE 8013 (Ager Faliscus); Latin or Faliscan?

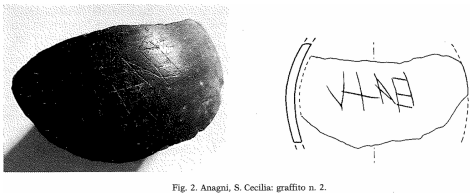


Fig. 2. Anagni, S. Cecilia: graffito n. 2.

Colonna and Gatti 1990: /hasu/ or /haθu/?

Thesaurus Inscriptionum Raeticarum SZ-04: is this an <X> or a non-linguistic design?

Research questions

- Broadest level: linguistic (e.g., Indo-European vs. non-Indo-European, Celtic vs. Italic, Latin vs. Sabellic) and geographic (e.g., Latium vs. Magna Graecia vs. Sicily vs. Cisalpine Gaul) categories have been adopted as conventional ways to divide the material of ancient Italy (in part to keep print volumes shorter): **to what degree do these divisions inhibit our ability to understand alphabet spread in ancient Italy?**

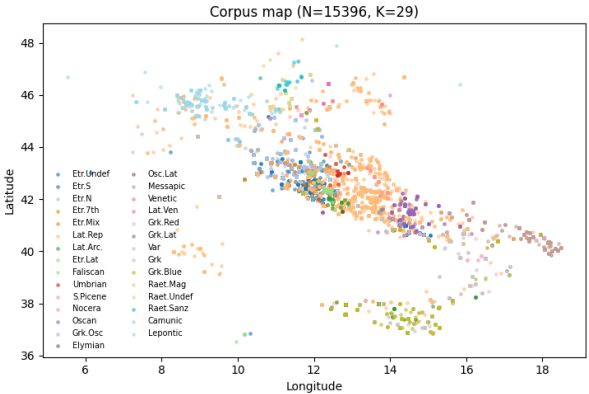
Research questions

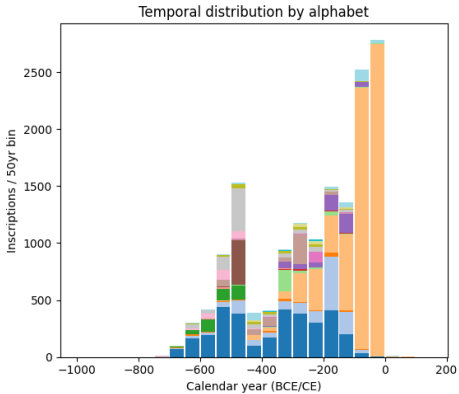
- Broadest level: linguistic (e.g., Indo-European vs. non-Indo-European, Celtic vs. Italic, Latin vs. Sabellic) and geographic (e.g., Latium vs. Magna Graecia vs. Sicily vs. Cisalpine Gaul) categories have been adopted as conventional ways to divide the material of ancient Italy (in part to keep print volumes shorter): **to what degree do these divisions inhibit our ability to understand alphabet spread in ancient Italy?**
- **Latin** is often invoked to explain alphabet distribution changes in the final 2 centuries BCE (e.g., De Simone and Marchesini 2002; Wallace 2007; Wallace 2008): **is there a monolithic response to Roman conquest, or are narratives of (alphabetic) resistance lost in this narrative?**

Research questions

- Broadest level: linguistic (e.g., Indo-European vs. non-Indo-European, Celtic vs. Italic, Latin vs. Sabellic) and geographic (e.g., Latium vs. Magna Graecia vs. Sicily vs. Cisalpine Gaul) categories have been adopted as conventional ways to divide the material of ancient Italy (in part to keep print volumes shorter): **to what degree do these divisions inhibit our ability to understand alphabet spread in ancient Italy?**
- **Latin** is often invoked to explain alphabet distribution changes in the final 2 centuries BCE (e.g., De Simone and Marchesini 2002; Wallace 2007; Wallace 2008): **is there a monolithic response to Roman conquest, or are narratives of (alphabetic) resistance lost in this narrative?**
- When moving outside the Greco-Roman lens of classics, **what new insights into diffusion and community interactions in 1st millennium BCE can be inferred from the inscriptional record?**

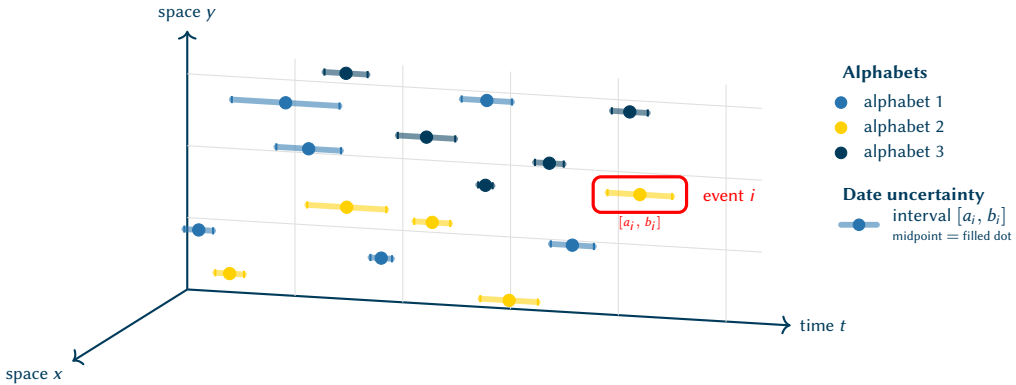
Corpus data for modeling





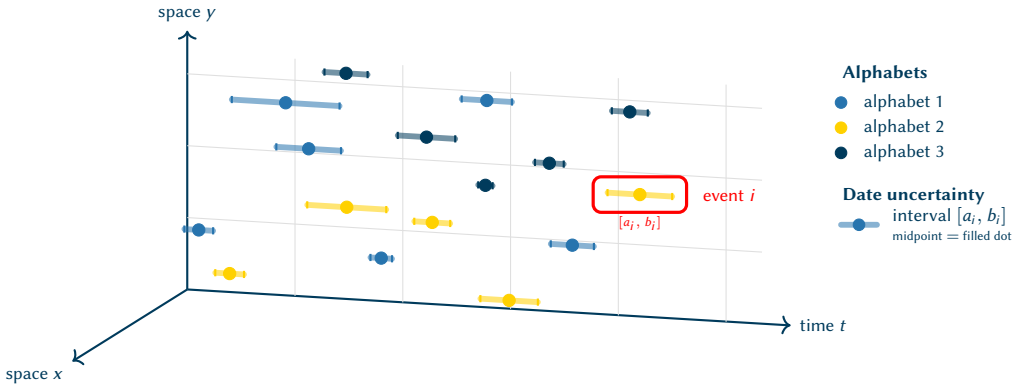
Marked Point Process modeling of alphabet dynamics

Each inscription is a *point in space-time, marked with an alphabet.* The date is rarely known exactly — it's an *interval* $[a_i, b_i]$.



Marked Point Process modeling of alphabet dynamics

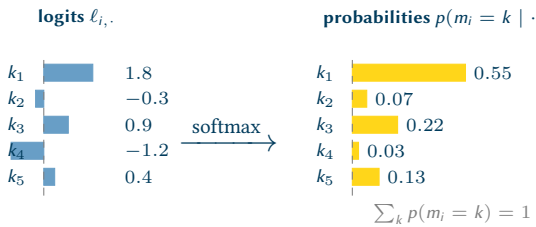
Each inscription is a *point in space-time, marked with an alphabet.* The date is rarely known exactly — it's an *interval* $[a_i, b_i]$.



- For each event i: **given location, time interval, and the alphabets of nearby earlier events, what's the probability i is in alphabet k?**

A Probability Vector via Softmax

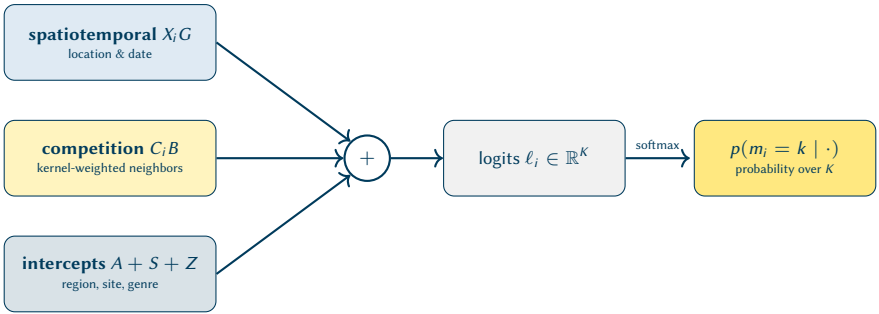
- For event i, we calculate the log-odds K of the inscription being written in a given alphabet.
- We turn these into probabilities using a softmax function.



$$p(m_i = k) = \frac{\exp(\ell_{i,k})}{\sum_{j=1}^{K} \exp(\ell_{i,j})}.$$

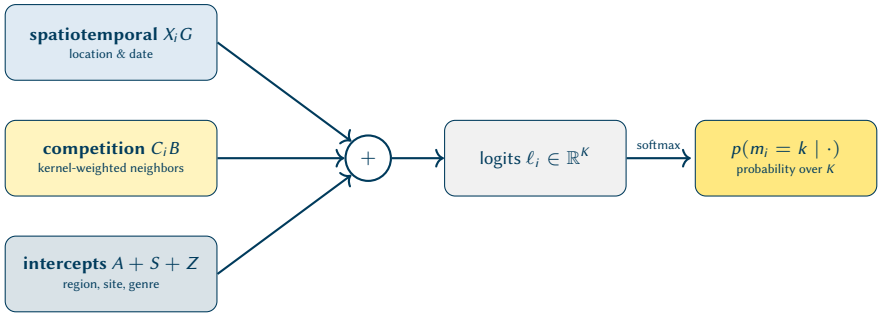
The modeling problem is essentially: **how do we build the logit vector ℓ_i for each event?**

Visualizing the model



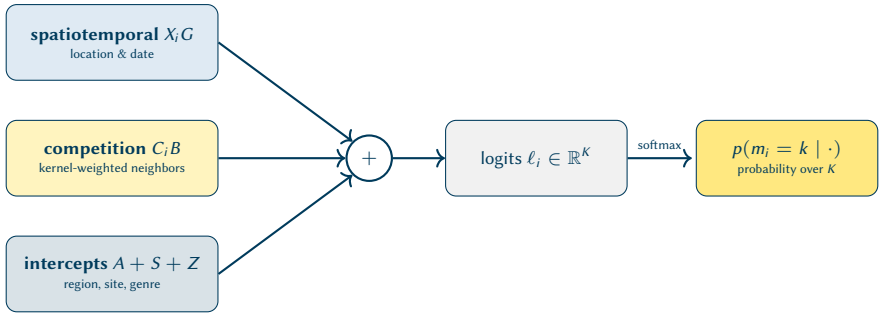
1. Each model component captures a different aspect of alphabet choice and controls for some variance that might have been attributed to competition:

Visualizing the model



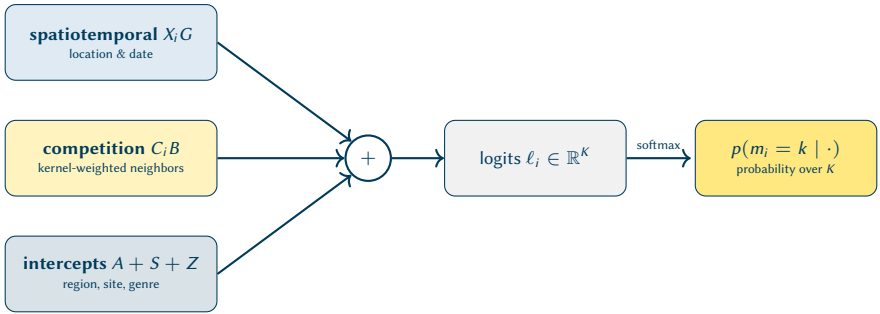
1. Each model component captures a different aspect of alphabet choice and controls for some variance that might have been attributed to competition:
   - **Spatiotemporal basis** X_iG: how much does this *place and time* favor each alphabet?

Visualizing the model



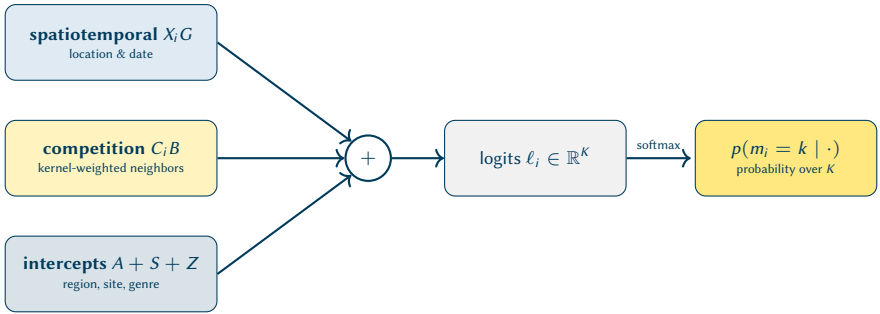
1. Each model component captures a different aspect of alphabet choice and controls for some variance that might have been attributed to competition:
   - **Spatiotemporal basis** $X_i G$: how much does this *place and time* favor each alphabet?
   - **Competition** $C_i B$: how much do the *nearby earlier alphabets* push toward each one?

Visualizing the model



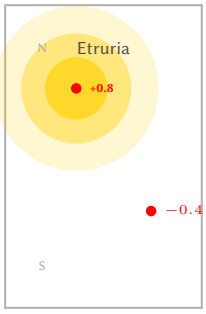
1. Each model component captures a different aspect of alphabet choice and controls for some variance that might have been attributed to competition:
   - **Spatiotemporal basis** $X_i G$: how much does this *place and time* favor each alphabet?
   - **Competition** $C_i B$: how much do the *nearby earlier alphabets* push toward each one?
   - **Intercepts**: what is the *baseline* for this region/site/genre?

Visualizing the model

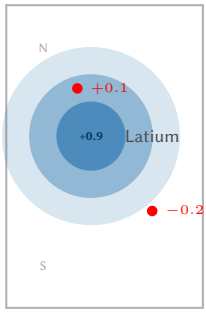


1. Each model component captures a different aspect of alphabet choice and controls for some variance that might have been attributed to competition:
   - **Spatiotemporal basis** $X_i G$: how much does this *place and time* favor each alphabet?
   - **Competition** $C_i B$: how much do the *nearby earlier alphabets* push toward each one?
   - **Intercepts**: what is the *baseline* for this region/site/genre?
2. We interpret the competition matrix B as cross-alphabet contact after accounting for the variation explained by location, time, region, site, and object type (genre).

The spatiotemporal effect $X_i G$ — "Where and when is each alphabet localized?"



Etruscan spatiotemporal basis
5th-c. BCE



Latin spatiotemporal
1st-c. BCE

Key

favored zone ($X_i G > 0$)

neutral ($X_i G \approx 0$)

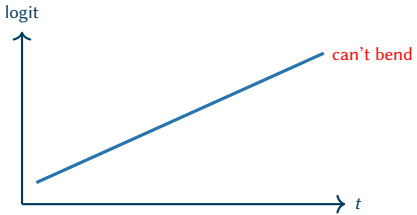
one event i

+0.8 spatiotemporal basis logit at i

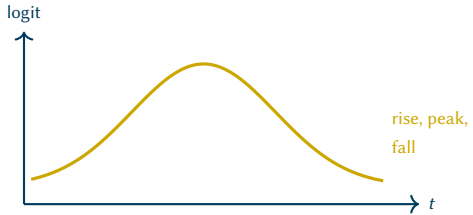
Interpretation: Given only event i's location and date, how much does this place-and-time *favor* each alphabet? The model learns one such surface per alphabet, parameterized as a low-order polynomial in (x, y, t).

Modeling time as a cubic polynomial

Alphabet usage isn't linear in time. Most alphabets in this corpus *rise, peak, and decline.* A linear time feature alone can't produce a "bendy" trend.



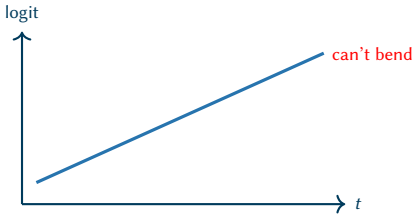
linear only: $G_{z_t} z_t$



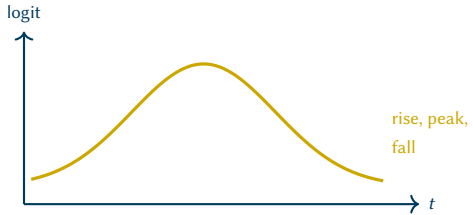
cubic: $G_{z_t} z_t + G_{z_t^2} z_t^2 + G_{z_t^3} z_t^3$

Modeling time as a cubic polynomial

Alphabet usage isn't linear in time. Most alphabets in this corpus *rise, peak, and decline.* A linear time feature alone can't produce a "bendy" trend.



linear only: $G_{z_t} z_t$

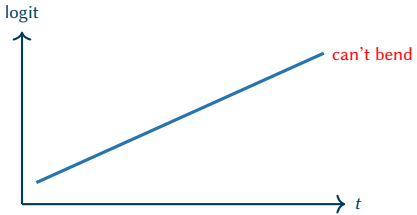


cubic: $G_{z_t} z_t + G_{z_t^2} z_t^2 + G_{z_t^3} z_t^3$

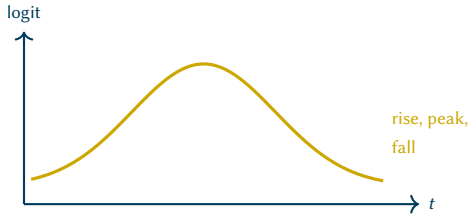
- **Latin** is rare early, grows through the Republic, and dominates late.

Modeling time as a cubic polynomial

Alphabet usage isn't linear in time. Most alphabets in this corpus *rise, peak, and decline*. A linear time feature alone can't produce a "bendy" trend.



linear only: $G_{z_t} z_t$

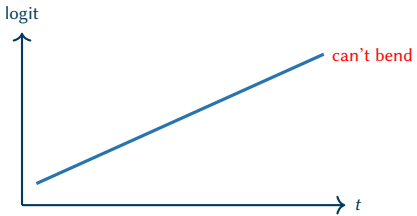


cubic: $G_{z_t} z_t + G_{z_t^2} z_t^2 + G_{z_t^3} z_t^3$

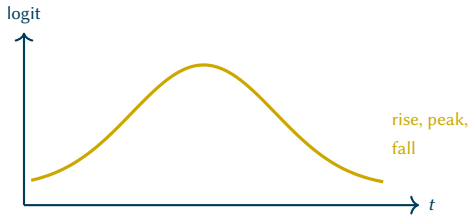
- **Latin** is rare early, grows through the Republic, and dominates late.
- **Etruscan** grows, peaks around the 4th–3rd c. BCE, then declines. A cubic polynomial in z_t can capture both.

Modeling time as a cubic polynomial

Alphabet usage isn't linear in time. Most alphabets in this corpus *rise, peak, and decline.* A linear time feature alone can't produce a "bendy" trend.



linear only: $G_{z_t} z_t$

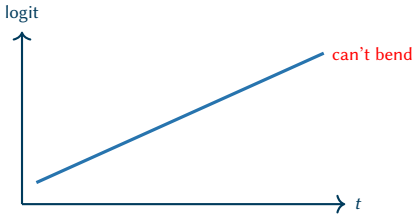


cubic: $G_{z_t} z_t + G_{z_t^2} z_t^2 + G_{z_t^3} z_t^3$

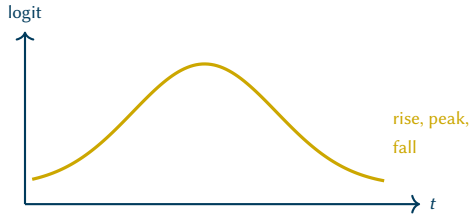
- **Latin** is rare early, grows through the Republic, and dominates late.
- **Etruscan** grows, peaks around the 4^{th}–3^{rd} c. BCE, then declines. A cubic polynomial in z_t can capture both.
- That's why the spatiotemporal basis includes z_t, z_t^2, z_t^3, plus the east/west interaction terms $z_t \cdot \mathbf{1}[\text{west}]$ and $z_t \cdot \mathbf{1}[\text{east}]$ for trajectories that differ on the two sides of the peninsula.

Modeling time as a cubic polynomial

Alphabet usage isn't linear in time. Most alphabets in this corpus *rise, peak, and decline.* A linear time feature alone can't produce a "bendy" trend.



linear only: $G_{z_t} z_t$

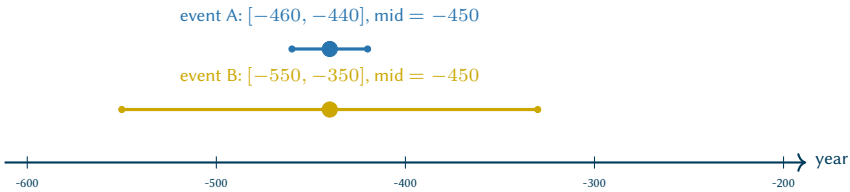


cubic: $G_{z_t} z_t + G_{z_t^2} z_t^2 + G_{z_t^3} z_t^3$

- **Latin** is rare early, grows through the Republic, and dominates late.
- **Etruscan** grows, peaks around the 4^{th}–3^{rd} c. BCE, then declines. A cubic polynomial in z_t can capture both.
- That's why the spatiotemporal basis includes z_t, z_t^2, z_t^3, plus the east/west interaction terms $z_t \cdot \mathbf{1}[\text{west}]$ and $z_t \cdot \mathbf{1}[\text{east}]$ for trajectories that differ on the two sides of the peninsula.
- How do we compute z_t^n when the date is an *interval*, not a point?

Date Uncertainty: The Problem

Every inscription comes with a date *range* $[a_i, b_i]$, not a point. Using just the midpoint discards information about how confident we are.

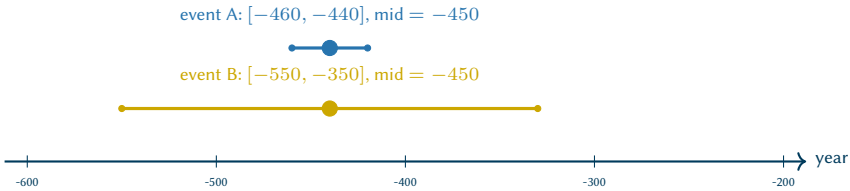


Both events have a midpoint of -450.

- Midpoint-only treatment says they are *identical* on the time axis.

Date Uncertainty: The Problem

Every inscription comes with a date *range* $[a_i, b_i]$, not a point. Using just the midpoint discards information about how confident we are.

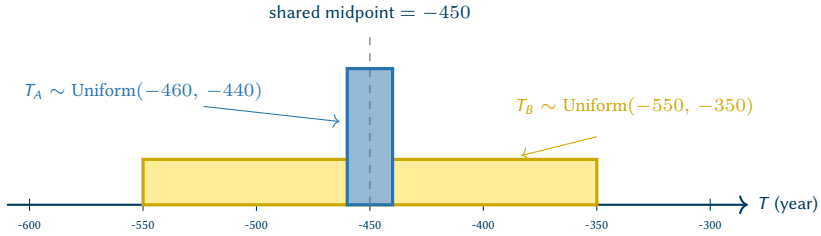


Both events have a midpoint of -450.

- Midpoint-only treatment says they are *identical* on the time axis.
- But event B carries far more uncertainty about *when* the text was actually inscribed.

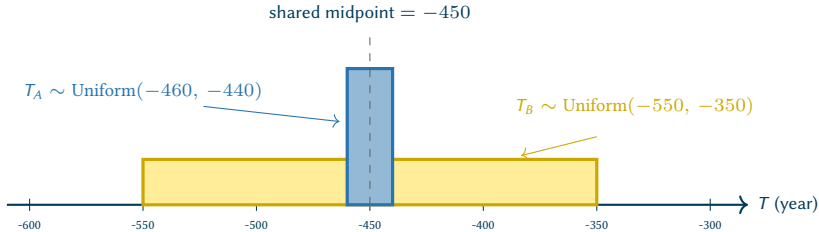
Treating the True Date as a Random Variable

- We don't know the true date T_i. We know that's it's probably somewhere in $[a_i, b_i]$.
- We assume that there is a uniform chance of the date being anywhere in this interval.



Treating the True Date as a Random Variable

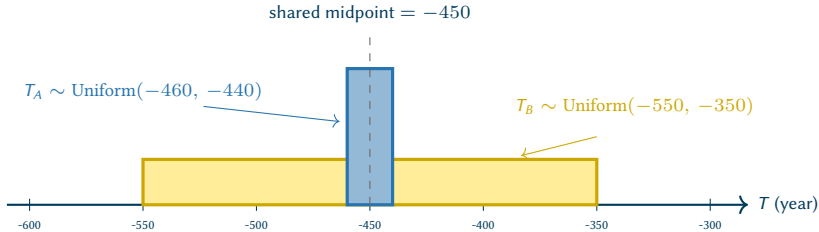
- We don't know the true date T_i. We know that's it's probably somewhere in $[a_i, b_i]$.
- We assume that there is a uniform chance of the date being anywhere in this interval.



- Both distributions are centered at -450. But event A's mass is concentrated; event B's mass is spread.

Treating the True Date as a Random Variable

- We don't know the true date T_i. We know that's it's probably somewhere in $[a_i, b_i]$.
- We assume that there is a uniform chance of the date being anywhere in this interval.



- Both distributions are centered at -450. But event A's mass is concentrated; event B's mass is spread.
- Instead of plugging in the midpoint and *then* raising it to the n-th power for the z_t^n polynomial, we ask: what is the *expected* value of T^n under each distribution?

Using moments to define the cubic date terms

For $T \sim \text{Uniform}(a, b)$, the first three raw moments have closed forms:

$$\mathbb{E}[T] = \frac{a+b}{2}, \quad \mathbb{E}[T^2] = \frac{a^2+ab+b^2}{3}, \quad \mathbb{E}[T^3] = \frac{a^3+a^2b+ab^2+b^3}{4}$$

Using moments to define the cubic date terms

For $T \sim \text{Uniform}(a, b)$, the first three raw moments have closed forms:

$$\mathbb{E}[T] = \frac{a+b}{2}, \quad \mathbb{E}[T^2] = \frac{a^2+ab+b^2}{3}, \quad \mathbb{E}[T^3] = \frac{a^3+a^2b+ab^2+b^3}{4}$$

For event A (low uncertainty) and event B (high uncertainty), $\mathbb{E}[T]$ is the same. But $\mathbb{E}[T^2]$ and $\mathbb{E}[T^3]$ are not.

	$\mathbb{E}[T]$	$\mathbb{E}[T^2]$	$\mathbb{E}[T^3]$
Event A: $[-460, -440]$	-450	**202,533**	**−91,170,000**
Event B: $[-550, -350]$	-450	**205,833**	**−95,625,000**
midpoint-only treatment	-450	202,500	−91,125,000

Using moments to define the cubic date terms

For $T \sim \text{Uniform}(a, b)$, the first three raw moments have closed forms:

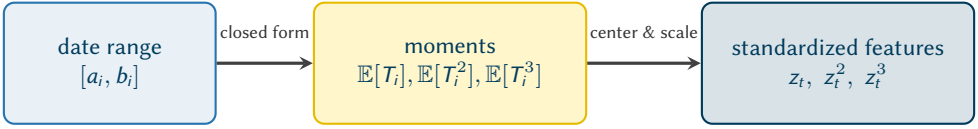
$$\mathbb{E}[T] = \frac{a+b}{2}, \quad \mathbb{E}[T^2] = \frac{a^2+ab+b^2}{3}, \quad \mathbb{E}[T^3] = \frac{a^3+a^2b+ab^2+b^3}{4}$$

For event A (low uncertainty) and event B (high uncertainty), $\mathbb{E}[T]$ is the same. But $\mathbb{E}[T^2]$ and $\mathbb{E}[T^3]$ are not.

	$\mathbb{E}[T]$	$\mathbb{E}[T^2]$	$\mathbb{E}[T^3]$
Event A: $[-460, -440]$	-450	**202,533**	**−91,170,000**
Event B: $[-550, -350]$	-450	**205,833**	**−95,625,000**
midpoint-only treatment	-450	202,500	−91,125,000

The difference $\mathbb{E}[T^2] - (\mathbb{E}[T])^2 = \text{Var}(T) = \frac{(b-a)^2}{12}$ is the uncertainty signal. Wider date range ⇒ larger gap ⇒ different z_t^2 feature value.

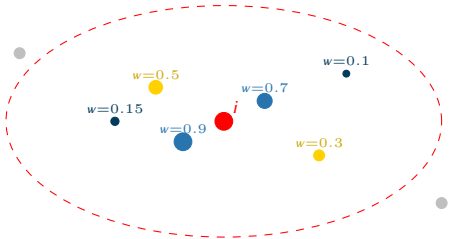
Plugging the Moments into the Design Matrix



- Using moments allows us to model the difference between an inscription that has lower date uncertainty than one with a higher date uncertainty.

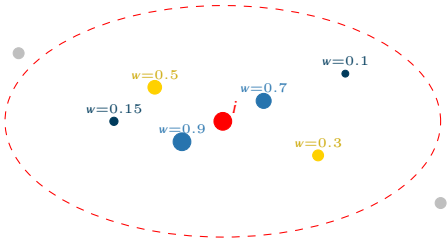
The Competition C_iB — What alphabets do neighboring inscriptions use: Part I

First we build C_i: a probability vector saying *what fraction of event i's recent local neighborhood used each alphabet.*



The Competition C_iB — What alphabets do neighboring inscriptions use: Part I

First we build C_i: a probability vector saying *what fraction of event i's recent local neighborhood used each alphabet.*

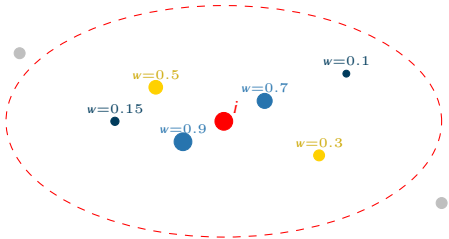


Each earlier neighbor j gets a weight $w_{ij} = \underbrace{\exp(-\beta\,(t_i - t_j))}_{\text{exponential in time}} \cdot \underbrace{\exp\!\left(-\tfrac{1}{2}|0\mathbf{r}_i - \mathbf{r}_j|0^2/\sigma^2\right)}_{\text{Gaussian in space}}$.

- Closer in time *and* space ⇒ larger weight.

The Competition C_iB — What alphabets do neighboring inscriptions use: Part I

First we build C_i: a probability vector saying *what fraction of event i's recent local neighborhood used each alphabet.*

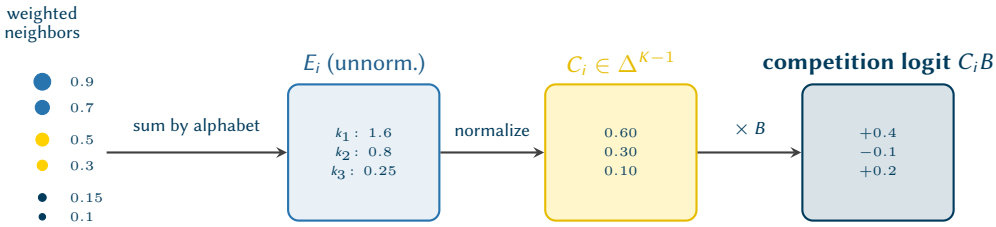


Each earlier neighbor j gets a weight $w_{ij} = \underbrace{\exp(-\beta\,(t_i - t_j))}_{\text{exponential in time}} \cdot \underbrace{\exp\!\left(-\tfrac{1}{2}|0\mathbf{r}_i - \mathbf{r}_j|0^2/\sigma^2\right)}_{\text{Gaussian in space}}$.

- Closer in time *and* space $\Rightarrow$ larger weight.
- The kernel parameters (β, σ) are themselves learned from the data.

The Competition C_iB — What alphabets do neighboring inscriptions use: Part II

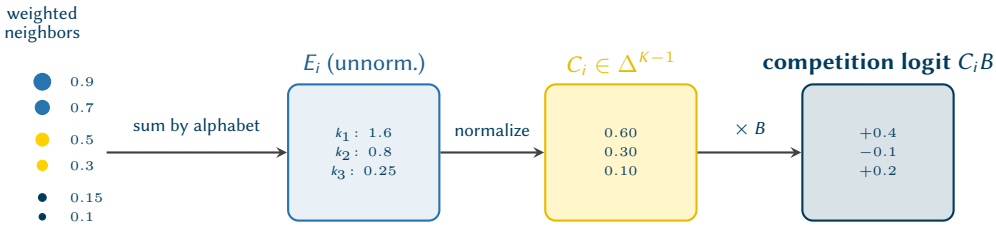
Sum the weights by alphabet, normalize $\Rightarrow C_i$. Then C_iB scores each alphabet by how favorably its mix of neighbors influences it.



- C_i: What alphabets are around me, kernel-weighted?

The Competition C_iB — What alphabets do neighboring inscriptions use: Part II

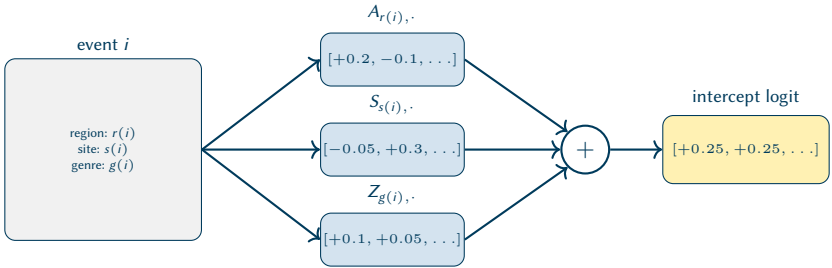
Sum the weights by alphabet, normalize $\Rightarrow C_i$. Then C_iB scores each alphabet by how favorably its mix of neighbors influences it.



- C_i: What alphabets are around me, kernel-weighted?
- $B_{k \to j}$: Given a neighbor wrote in k, how much does that push me toward writing in j?

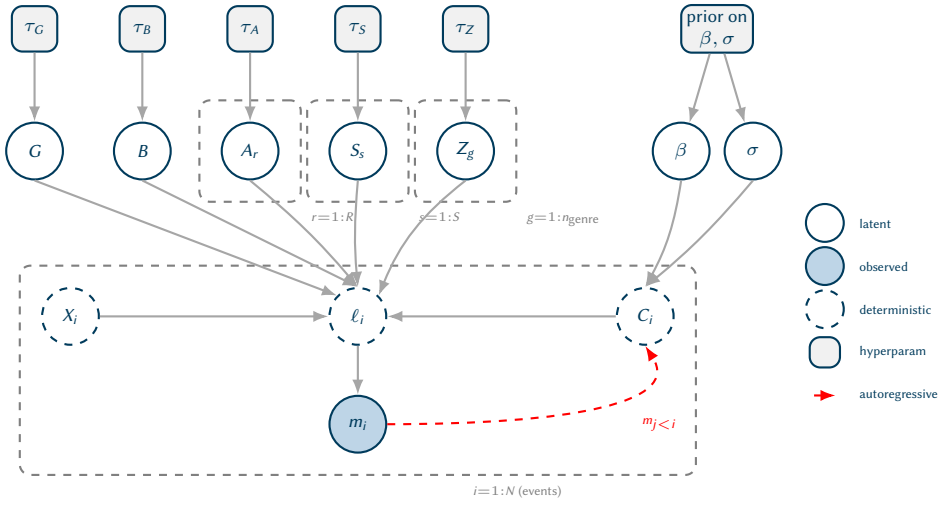
Hierarchical intercepts: Defining baseline propensities by inscription attributes

Three additional parameters that absorb everything not captured by location, time, or the neighborhood.



- $A_{r,\cdot}$ — a K-vector per region (Etruria, Latium, Apulia, ...): captures region-wide baselines.
- $S_{s,\cdot}$ — a K-vector per site: captures single-find-spot idiosyncrasies (one tomb, one sanctuary).
- $Z_{g,\cdot}$ — a K-vector per genre (funerary, votive, ownership-mark, ...): captures genre-specific script preferences.

Graphical representation of the model



Inferring model parameters with Stochastic Variational Inference

The model has $\sim 10^3$–10^4 latent parameters (depending on the random effect structure).

Inference Algorithm: SVI

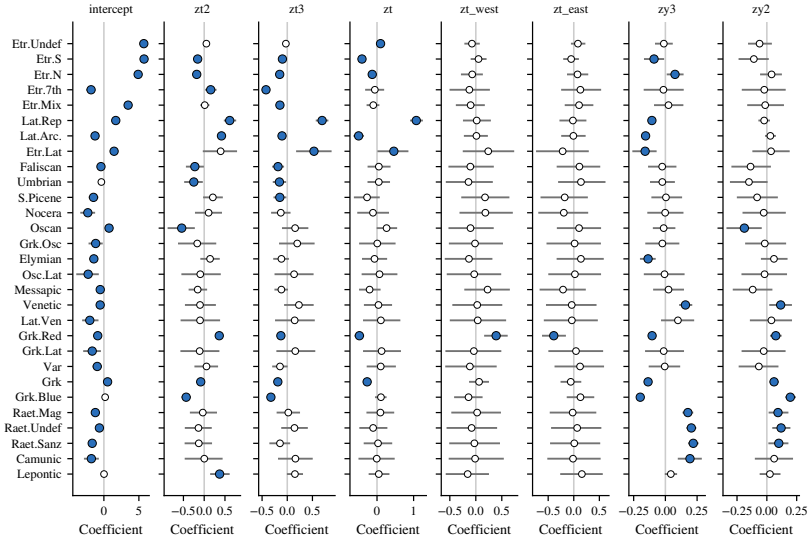
- **Guide:** AutoLowRankMultivariateNormal, rank 32.
- **Optimizer:** AdamW with warmup-cosine LR (5×10^{-3} peak).
- **Process:** $10{,}000$ steps | 4 particles per step | clipped gradients.

Motivating the use of SVI

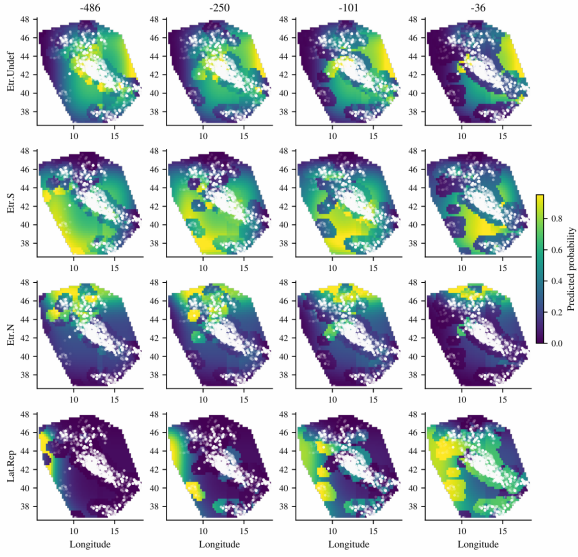
The geometry of B + the sparse edge graph + the kernel parameters makes Monte Carlo sampling methods slow at this scale.

$\Rightarrow$ **SVI gives us competitive posterior intervals**, at a fraction of the cost of sampling.

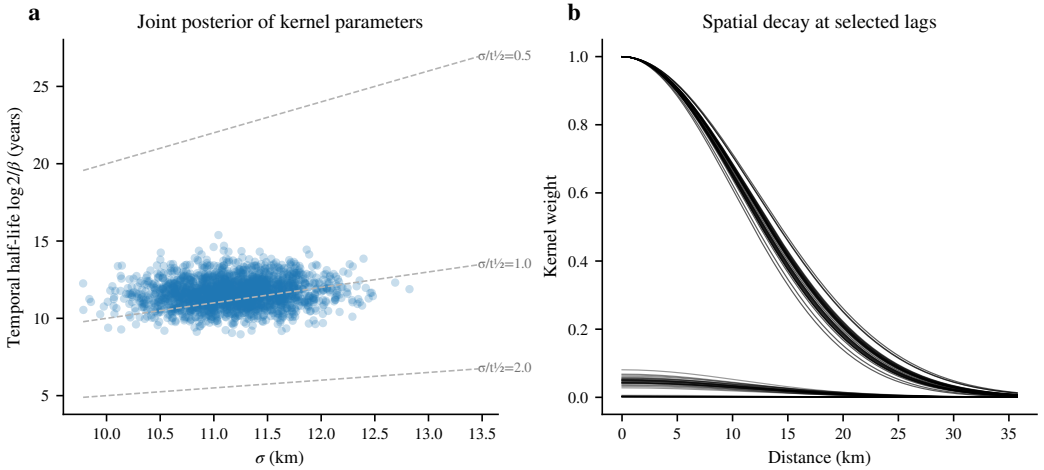
Spatiotemporal coefficents



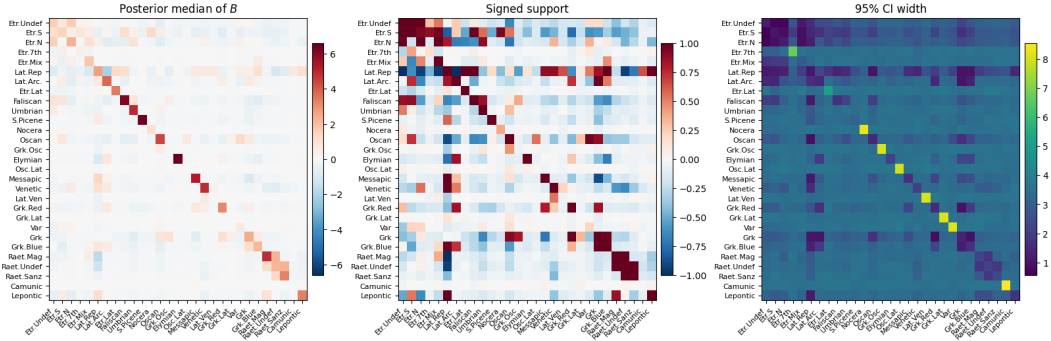
Spatiotemporal probability fields for selected alphabets



Kernel parameters

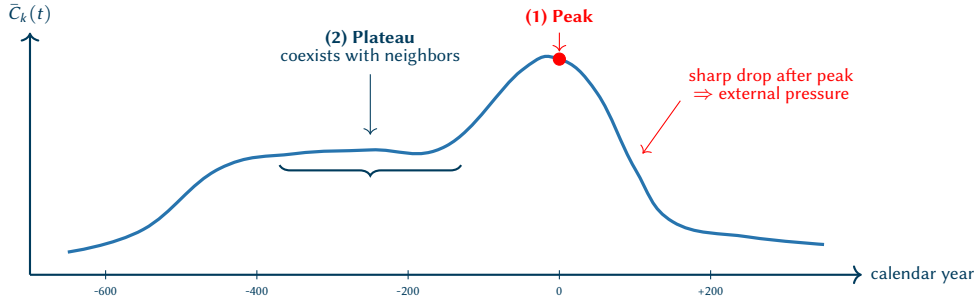


Competition effect posteriors



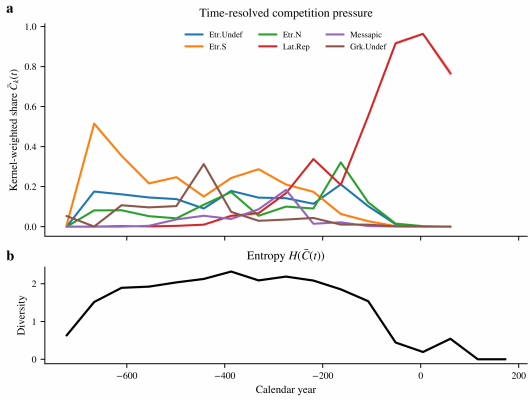
Change in competition pressures

Each curve shows $\bar{C}_k(t)$ for one alphabet k. Three things to look for:

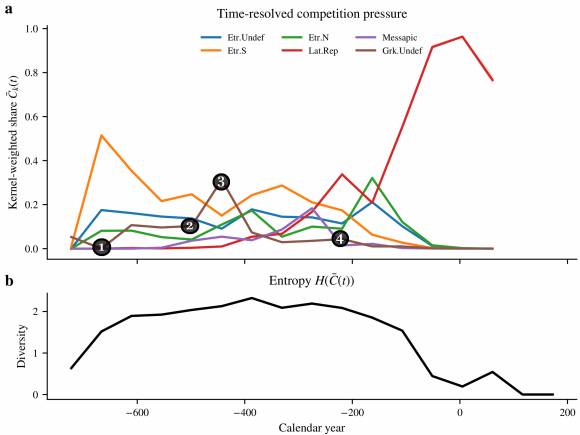


Interpretation: $\bar{C}_k(t)$ is a kernel-weighted *share* of the neighborhood, which is robust to overall inscription rates and recovery bias.

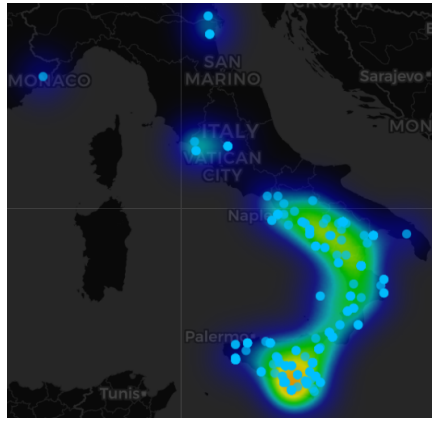
Interpreting Alphabet Competition Over Time: Historical Events and Data Bias



Interpreting Alphabet Competition Over Time: Greek Alphabet

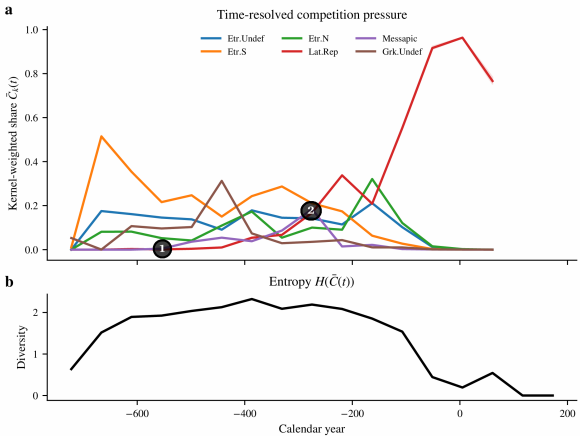


(a) Time-resolved competition pressure

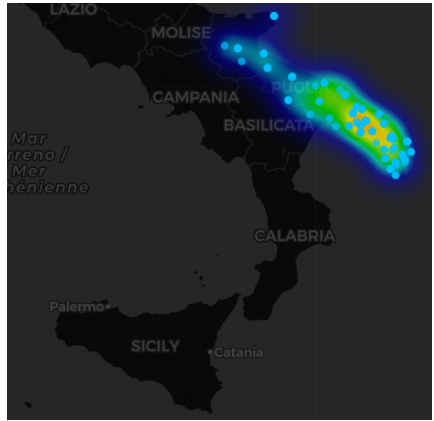


(b) Heatmap of inscriptions in the Greek alphabet

Interpreting Alphabet Competition Over Time: Messapic Alphabet

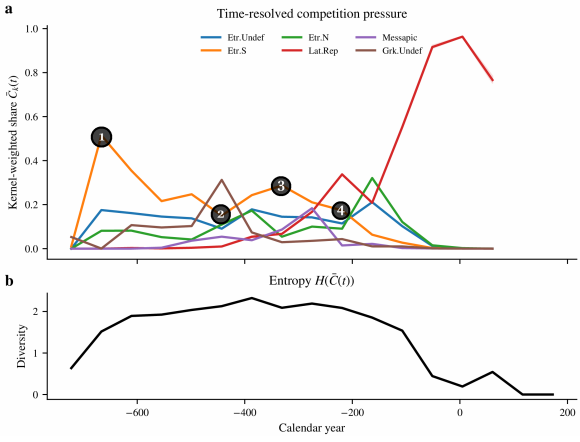


(a) Time-resolved competition pressure

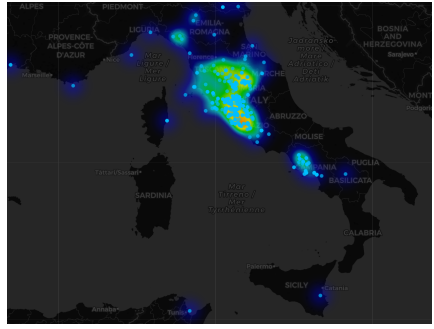


(b) Heatmap of inscriptions in the Messapic alphabet

Interpreting Alphabet Competition Over Time: Southern Etruscan Alphabet

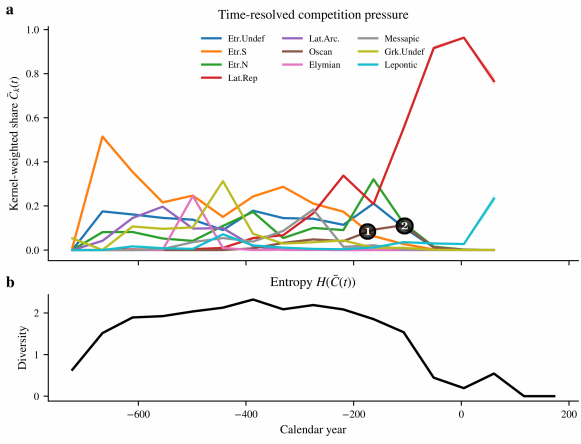


(a) Time-resolved competition pressure

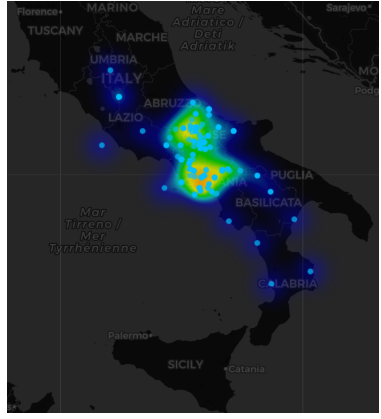


(b) Heatmap of inscriptions in the Southern Etruscan alphabet

Interpreting Alphabet Competition Over Time: Southern Etruscan Alphabet and Oscan (not just Latin!)

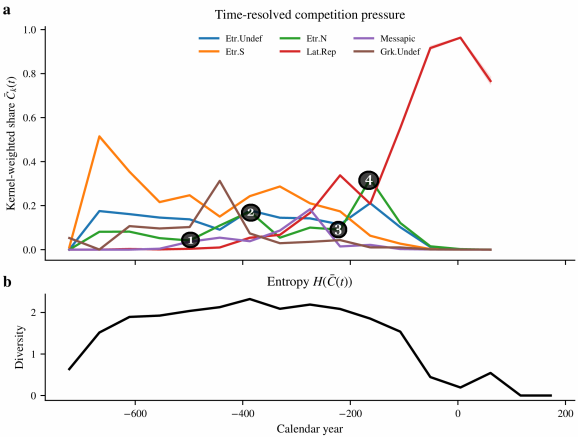


(a) Time-resolved competition pressure

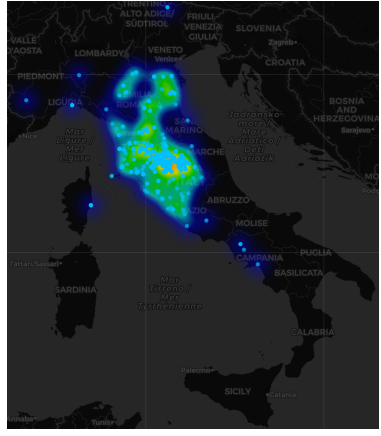


(b) Heatmap of inscriptions in the Oscan alphabet

Interpreting Alphabet Competition Over Time: Northern Etruscan Alphabet

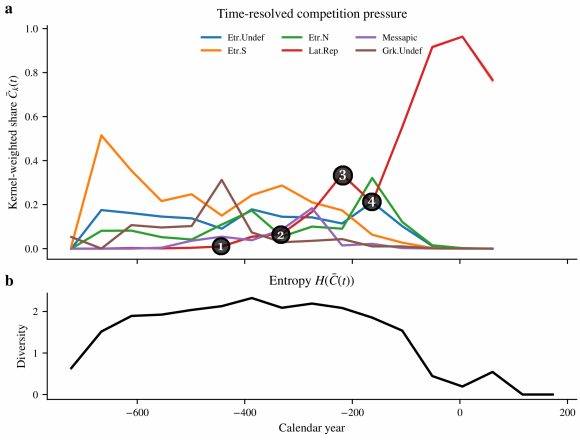


(a) Time-resolved competition pressure

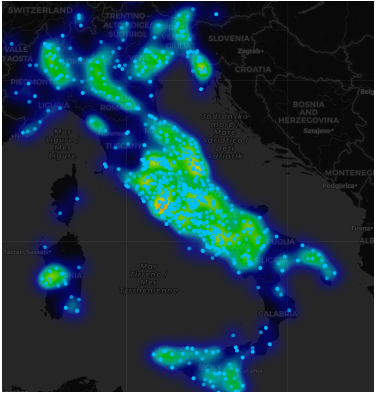


(b) Heatmap of inscriptions in the Northern Etruscan alphabet

Interpreting Alphabet Competition Over Time: Latin Alphabet



(a) Time-resolved competition pressure



(b) Heatmap of inscriptions in the Latin alphabet

Model performance

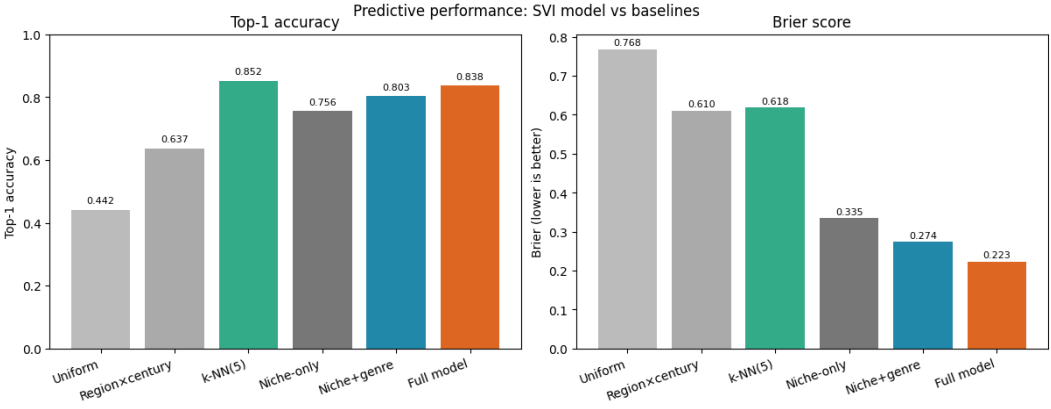
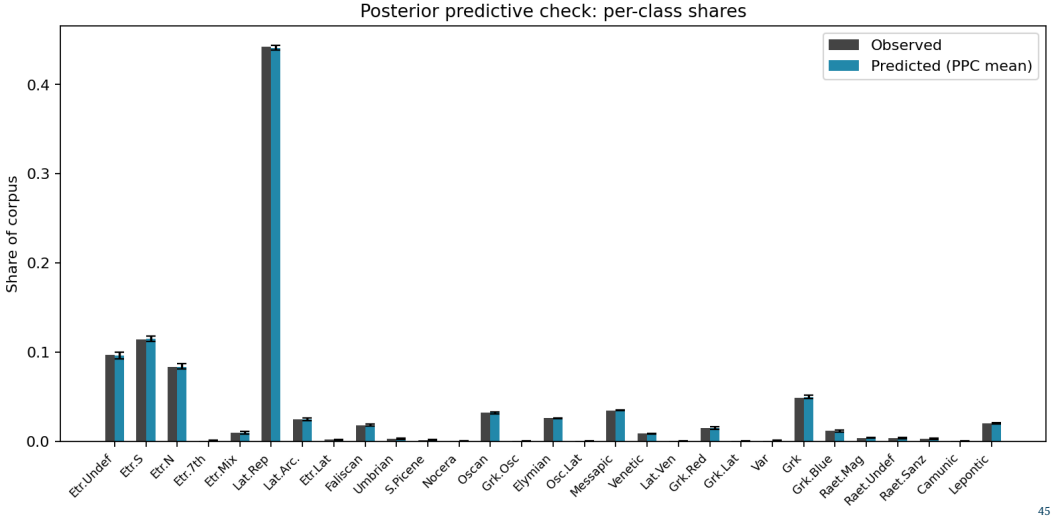


Figure: Prediction accuracy and error

Posterior predictive checks



Conclusions

- Roman conquest does explain *some* changes in alphabet competition, but colonization efforts and migration by several other actors (including Punics, Greeks, Syracusans, Oscans, Etruscans, etc.) are also essential for understanding alphabet dynamics in 1st millennium BCE.

Conclusions

- Roman conquest does explain *some* changes in alphabet competition, but colonization efforts and migration by several other actors (including Punics, Greeks, Syracusans, Oscans, Etruscans, etc.) are also essential for understanding alphabet dynamics in 1st millennium BCE.
- Classics as a discipline privileges the Greco-Roman literary record, essentially setting the clock for studying ancient Italy to the third century BCE.

Conclusions

- Roman conquest does explain *some* changes in alphabet competition, but colonization efforts and migration by several other actors (including Punics, Greeks, Syracusans, Oscans, Etruscans, etc.) are also essential for understanding alphabet dynamics in 1st millennium BCE.
- Classics as a discipline privileges the Greco-Roman literary record, essentially setting the clock for studying ancient Italy to the third century BCE.
- Indo-European linguistics privileges the Indo-European languages at the expense of other languages in direct contact with these languages (e.g., Rix 2002; Crawford et al. 2011).

Conclusions

- Roman conquest does explain *some* changes in alphabet competition, but colonization efforts and migration by several other actors (including Punics, Greeks, Syracusans, Oscans, Etruscans, etc.) are also essential for understanding alphabet dynamics in 1st millennium BCE.
- Classics as a discipline privileges the Greco-Roman literary record, essentially setting the clock for studying ancient Italy to the third century BCE.
- Indo-European linguistics privileges the Indo-European languages at the expense of other languages in direct contact with these languages (e.g., Rix 2002; Crawford et al. 2011).
- Both disciplines largely exclude studies of Etruscan, the largest corpus of epigraphy and a primary driver of epigraphic culture in ancient Italy.

Conclusions

- Roman conquest does explain *some* changes in alphabet competition, but colonization efforts and migration by several other actors (including Punics, Greeks, Syracusans, Oscans, Etruscans, etc.) are also essential for understanding alphabet dynamics in 1st millennium BCE.
- Classics as a discipline privileges the Greco-Roman literary record, essentially setting the clock for studying ancient Italy to the third century BCE.
- Indo-European linguistics privileges the Indo-European languages at the expense of other languages in direct contact with these languages (e.g., Rix 2002; Crawford et al. 2011).
- Both disciplines largely exclude studies of Etruscan, the largest corpus of epigraphy and a primary driver of epigraphic culture in ancient Italy.
- The privileging of literary texts perpetuates ancient Greco-Roman narratives and silences the voices of the other groups of language speakers in Italy whose voice is recorded in the vibrant multi-lingual inscriptional record.

Future work

- **Date-uncertainty in the kernel.** In contrast to the spatiotemporal effects, the competition kernel uses midpoint t. We could instead use an n-node Gauss–Legendre quadrature over $\text{Uniform}(a_i, b_i)$ or some other prior distribution.

Future work

- **Date-uncertainty in the kernel.** In contrast to the spatiotemporal effects, the competition kernel uses midpoint t. We could instead use an n-node Gauss–Legendre quadrature over $\text{Uniform}(a_i, b_i)$ or some other prior distribution.
- **Periodized $B^{(p)}$.** The competition parameters might be different across time periods. This could be modeled using smooth random-walk priors or smoothing over spatiotemporal GMRFs.

Future work

- **Date-uncertainty in the kernel.** In contrast to the spatiotemporal effects, the competition kernel uses midpoint t. We could instead use an n-node Gauss–Legendre quadrature over $\mathrm{Uniform}(a_i, b_i)$ or some other prior distribution.
- **Periodized $B^{(p)}$.** The competition parameters might be different across time periods. This could be modeled using smooth random-walk priors or smoothing over spatiotemporal GMRFs.
- **Genre as a covariate vs. as a multiplier.** Right now Z_g is an additive log-odds shift. Testing whether object types, inscription types, and alphabet direction interact with B (i.e., different genres permit different kinds of script borrowing) would be a desideratum.

Future work

- **Date-uncertainty in the kernel.** In contrast to the spatiotemporal effects, the competition kernel uses midpoint t. We could instead use an n-node Gauss–Legendre quadrature over $\text{Uniform}(a_i, b_i)$ or some other prior distribution.
- **Periodized $B^{(p)}$.** The competition parameters might be different across time periods. This could be modeled using smooth random-walk priors or smoothing over spatiotemporal GMRFs.
- **Genre as a covariate vs. as a multiplier.** Right now Z_g is an additive log-odds shift. Testing whether object types, inscription types, and alphabet direction interact with B (i.e., different genres permit different kinds of script borrowing) would be a desideratum.
- **Archaeological context**. When controlling for objects with no findspot context or only secondary context (Marlowe 2013), are there significant changes in the modelling results? Can data with secure findspot contexts be used as training data to confirm/revise hypotheses of provenance for inscriptions with no findspot data?

Future work

- **Date-uncertainty in the kernel.** In contrast to the spatiotemporal effects, the competition kernel uses midpoint t. We could instead use an n-node Gauss–Legendre quadrature over $\text{Uniform}(a_i, b_i)$ or some other prior distribution.
- **Periodized $B^{(p)}$.** The competition parameters might be different across time periods. This could be modeled using smooth random-walk priors or smoothing over spatiotemporal GMRFs.
- **Genre as a covariate vs. as a multiplier.** Right now Z_g is an additive log-odds shift. Testing whether object types, inscription types, and alphabet direction interact with B (i.e., different genres permit different kinds of script borrowing) would be a desideratum.
- **Archaeological context**. When controlling for objects with no findspot context or only secondary context (Marlowe 2013), are there significant changes in the modelling results? Can data with secure findspot contexts be used as training data to confirm/revise hypotheses of provenance for inscriptions with no findspot data?
- **North Africa and Phoenician/Punic**. Languages from Italy end up in the North African inscriptional record: what interactions can be observed there?

Thank you for listening!

Comments? Questions?

References I

Agostiniano, Luciano (1977). *Iscrizioni anelleniche di Sicilia I. Le iscrizioni elime*. Firenze: L. S. Olschki.

— (2021). *Iscrizioni anelleniche di Sicilia I. Le iscrizioni elime. Appendice 1978-2020*. Roma: L'Erma di Bretschneider.

Bakkum, Gabriël Cornelis Leonides Maria, ed. (2009). *The Latin dialect of the Ager Faliscus: 150 years of scholarship*. Amsterdam: Amsterdam University Press. 2 pp. ISBN: 978-90-5629-562-2 978-90-485-0875-4.

Crawford, Michael Hewson et al. (2011). *Imagines Italicae: a corpus of Italic inscriptions*. 3 vols. Bulletin of the Institute of Classical Studies supplement 110. London: Institute of Classical Studies, School of Advanced Study, University of London. ISBN: 978-1-905670-30-7 978-1-905670-35-2 978-1-905670-36-9.

De Simone, Carlo and Simona Marchesini (2002). *Monumenta Linguae Messapicae*. Wiesbaden: L. Reichert. ISBN: 978-3-89500-228-1.

References II

Dubois, Laurent (1989). *Inscriptions grecques dialectales de Sicile. 1.* Collection de l'Ecole Française de Rome 119. Num Pages: 303. Rome: École Française. ISBN: 978-2-7283-0185-0.

— (1995). *Inscriptions grècques dialectales de Grande Grèce. 1: Colonies eubéennes. Colonies ioniennes. Emporia.* École Pratique des Hautes Études, IVe Section, Sciences Historiques et Philologiques III, Hautes études du monde gréco-romain 21. Num Pages: 202. Genève: Droz. ISBN: 978-2-600-00085-7.

— (2002). *Inscriptions grecques dialectales de Grande Grèce. 2 Colonies achéennes.* Hautes études du monde gréco-romain 30. Genève: Droz. ISBN: 978-2-600-00690-3.

— (2008). *Inscriptions grecques dialectales de Sicile. 2.* Hautes études du monde gréco-romain 40. Genève: Droz. ISBN: 978-2-600-01340-6.

Lejeune, Michel (1974). *Manuel de la langue Vénète.* Indogermanische Bibliothek 1. Reihe. Heidelberg: C. Winter. ISBN: 978-3-533-02353-1.

References III

- Marlowe, Elizabeth (2013). *Shaky Ground: Context, Connoisseurship and the History of Roman Art*. Duckworth Debates in Archaeology. London: Bloomsbury Publishing. 1 p. ISBN: 978-0-7156-4064-7 978-1-4725-0210-0 978-1-4725-0209-4.
- Prosdocimi, Aldo Luigi (1965). "Per una edizione delle iscrizioni della Val Camonica". In: *Studi Etruschi* 33, pp. 575–599.
- Prosdocimi, Aldo Luigi and G. B. Pellegrini (1967). *La lingua venetica I: Le iscrizioni*. Fir: Istituto di glottologia dell'Università di Padova–Circolo Linguistico Fiorentino.
- Rix, Helmut (2002). *Sabellische Texte: die Texte des Oskischen, Umbrischen und Südpikenischen*. Handbuch der italischen Dialekte / begr. von Emil Vetter 5. Heidelberg: Winter. 227 pp. ISBN: 978-3-8253-0853-7.
- Rix, Helmut and Gerhard Meiser (2014). *Etruskische Texte: editio minor*. Neu Bearbeit. Studien zur historisch-vergleichenden Sprachwissenschaft Bd. 4. Hamburg: Baar. ISBN: 978-3-935536-71-4.

References IV

Tibiletti Bruno, Maria Grazia (1990). “Nuove iscrizioni camune”. In: *Quaderni Camuni* 49-50, pp. 29–171.

Wallace, Rex E. (2007). *The Sabellic languages of ancient Italy*. Languages of the world / Materials 371. München: LINCOM Europa. 72 pp. ISBN: 978-3-89586-990-7.

— (2008). *Zikh Rasna: a manual of the Etruscan language and inscriptions*. Ann Arbor, Mich.: Beech Stave Press. 271 pp. ISBN: 978-0-9747927-4-3.