History Research
2026, Vol. 14, No. 2, pp. 56–66
https://doi.org/10.11648/j.history.20261402.12

SciencePG
Science Publishing Group

Research Article

An Exploratory Quantitative Spatial Analysis of Early Saxon Discursive Identity (250-700 CE)

Robert Jan Duchateau*

Independent Researcher, Groesbeek, The Netherlands

Abstract

Traditional early medieval historiography frequently treats the early Saxons as a homogeneous, migrating ethnic entity. This paper explores this paradigm by conducting an exploratory quantitative textual and spatial analysis of early medieval identity construction, focusing on the maritime networks of the North Sea and the territorial river systems of the Rhine-Meuse delta in Northwest Europe. Utilising 58 systematically coded historical source elements from the digital Corpus Saxonum ante DCC, this study spans Late Antiquity and the core transitional centuries of the Migration Period (250-700 CE). Pearson's chi-squared tests ($\chi^2 = 49.34$, $df = 4$, $N = 58$, $p < .001$, $V = .65$) indicate a statistically significant chronological correlation within this specific dataset: the early discursive label 'Raider' (250-400 CE) systematically co-occurs with maritime settings, giving way to the sedentary, territorial label 'Inhabitant' (550-700 CE) in inland descriptions. A two-sample F-test performed on database-validated geographic coordinates from certified gazetteers confirms an overall geographical anisotropy for the total sample ($F(57, 57) = 2.30$, $p = .001$), indicating an unequal spatial dispersion of the textual mentions across the European littoral. Given the limited size of the dataset, these statistical configurations are treated as heuristic trends within this compiled corpus rather than definitive proof. The results suggest that the spatial-temporal distribution of the ethnonym in the text fragments aligns with shifting regional friction boundaries and administrative rhetoric, offering a provisional, interdisciplinary framework for quantitative historiography.

Keywords

Saxons, Ethnogenesis, Late Antiquity, Quantitative Historiography, Meta-ethnic Frontier, Discourse Analysis, Large Language Models (LLMs), Geographic Information Systems (GIS)

1. Introduction

The shift from primordialist, biological definitions of early medieval tribes to fluid models of ethnogenesis has fundamentally reshaped Saxon studies over the last few decades [16]. Halsall [10] demonstrates that barbarian identities were rarely monolithic or static; instead, they were continuously negotiated and reconstructed at the borders of the fracturing Roman world. This aligns with the critique presented by Noble [15], who emphasises that 'tribal' labels in late antique and early medieval sources often reveal more about the administrative classification and rhetorical needs of Mediterranean text producers than the internal self-perception of peripheral populations. Consequently, the term 'Saxon' is analysed here primarily as an external discursive shorthand rather than an immutable ethnic marker. Given the limited dataset ($N = 58$), these statistical configurations are treated as exploratory,

*Correspondence: Robert Jan Duchateau (robduchateau@gmail.com)

Received: 28 August 2026; **Accepted:** 14 September 2026; **Published:** 28 September 2026

heuristic trends within this compiled corpus rather than definitive historical proof.

The mechanism driving this discursive categorisation can be theoretically linked to frontier dynamics. Under Turchin's [19] structural-demographic framework, collective solidarity (asabiya) and external ethnic labelling tend to intensify within specific zones of geopolitical friction. In the context of early Anglo-Saxon ethnogenesis, Higham and Ryan [11] observe that insular identity emerged not from a straightforward displacement of indigenous populations, but through complex, highly localised interactions between small immigrant elites and native Britons. This perspective is further refined by Fleming [7], who stresses that the material culture and territorial baseline of early Britain suggest a fragmented consolidation process rather than a uniform continental invasion.

Furthermore, modern spatial re-evaluations of early medieval frontiers by Wickham [24] indicate that regions lacking intense imperial administrative friction, such as the Rhine-Meuse delta, frequently evolved into zones of local isolation and political invisibility within the written record. Finally, the historiographical deconstruction offered by Geary [8] cautions that early medieval ethnic labels were routinely recycled by later Carolingian and Merovingian state apparatuses to categorise geographically disparate populations.

The recent compilation of the digital 'Corpus Saxonum ante DCC' [22] provides an empirical foundation to quantitatively explore these textual patterns. This study tests the hypothesis that the discursive characterisation of the Saxons in the selected sources underwent a temporal transformation, transitioning from maritime raiders to sedentary land-dwellers, and that this shift was geographically anisotropic. To map this transformation, this paper advances a methodology of discursive cartography. By translating historical text fragments into spatial variables within a GIS (Geographic Information Systems) framework, we model the perceived geography of early medieval chroniclers rather than objective population movements. To safeguard the structural replicability of this qualitative framework, the manual taxonomy is subjected to an independent algorithmic validation benchmark (Appendix I), utilising LLMs (Large Language Models) to test the semantic stability of the operationalised categories against subjective coder bias. For the purposes of this study, the term 'Early Middle Ages' is operationally defined according to the chronological framework established by Wickham [24] and Wood [25], spanning the collapse of western Roman administrative structures in the fifth century to the consolidation of regional successor states around 700 CE (Common Era).

2. Materials and Methods

To ground the quantitative spatial analysis within rigorous historical source criticism, the primary unit of analysis is operationally defined as the individual textual fragment or source element curated in the digital 'Corpus Saxonum ante DCC' [5, 22]. This digital collection yields a final analytical dataset of 58 unique literary testimonies spanning the period between 250 and 700 CE. To ensure deep historical nuance while maintaining replication standards, this study deploys a rigorous two-step hybrid coding workflow that separates expert source-critical variables from automated database-driven spatial logging.

To safeguard external validity, the primary corpus was validated against an institutional control dataset ($N_{academic} = 445$) harvested from Scopus [6] using the targeted string: (('Saxon' OR 'Sachsen' OR 'Saxones') AND NOT ('modern' OR 'Lower Saxony')) AND (('early medieval' OR 'Late Antiquity' OR 'Migration Period')). Proportional comparison confirmed that the relative temporal weight of the primary corpus tracks closely with modern research intensity across the early contact phase (33.9% primary vs. 31.9% academic), transitional phase (28.5% primary vs. 26.5% academic), and consolidation phase (37.5% primary vs. 41.6% academic), establishing the digital 'Corpus Saxonum ante DCC' as a highly representative proxy for the broader academic landscape. The foundational nominal variables requiring critical historical contextualisation were coded entirely manually by the researcher to maintain expert source-critical control (Appendix I-II):

1) *Period:* Entries were stratified into three consecutive 150-year chronological cohorts based on the historical horizon of the events reported, isolating macro-developments across the early contact phase ('250-400 CE'), the post-Roman migration phase ('400-550 CE'), and the early medieval territorial consolidation phase ('550-700 CE').
2) *Perspective:* Authorial perspective was logged as a nominal variable capturing the socio-political and ethnic background of the text producer.
3) *Label:* The specific discursive or functional role attributed to the Saxons within the textual episode was manually classified using the semantic categories of 'Raider', representing piracy, maritime aggression, looting, or an acute external threat; 'Ally', denoting operations as treaty parties, political partners, or military auxiliaries; or 'Inhabitant', indicating characterisation as a settled, sedentary population with a fixed geographical or agricultural base.

To eliminate human bias, geographic hallucinations, and arbitrary coordinate plotting associated with generic large language model outputs, physical environmental variables and geographic positions were extracted using a deterministic automated pipeline (Appendix I-II):

1) *Setting:* The environmental setting was first classified as 'Sea', 'Coast', or 'Land' using Google Gemini 1.5 Pro deployed via API with a deterministic, zero-temperature setting to process structured JSON prompts based strictly on textual data.
2) *Location:* Rather than prompting the model to approximate coordinates, the pipeline utilised the model to extract the primary late antique or early medieval toponym or regional identifier indicating the active centre of gravity of the Saxon presence in the text. These extracted his-

torical toponyms were subsequently queried against certified digital gazetteers, specifically the Pleiades Gazetteer and the Digital Atlas of the Roman Empire [2, 3], resolving each entry into precise decimal geographic coordinates consisting of latitude and longitude. Broad provincial territories or shifting frontier zones were standardised to the calculated centroids of their documented late antique boundaries.

To evaluate temporal shifts across these nominal variables and assess the correlations between manual and automated labels, cross-tabulations were constructed and evaluated using Pearson's chi-squared (χ^2) tests of independence. Because several cells within the cross-tabulations contained expected frequencies below 5, violating Cochran's criteria [4], exact *p*-values were computed utilising a Monte Carlo simulation based on 20,000 permutations [1]. Cell-specific deviations were located via Adjusted Standardised Residuals (ASR), with significance thresholds adjusted using the Holm-Bonferroni correction [12] (α =.05), and the overall effect size was determined using Cramer's *V*.

To measure geographical anisotropy and determine whether the spatial dispersion of the historical data was significantly stretched along a specific directional axis, a two-sample *F*-test for the equality of variances was executed. To eliminate spatial distortion caused by the convergence of meridians at mid-latitudes, the raw decimal geographic coordinates were mathematically projected into a local equal-area metric space conforming to the Lambert Azimuthal Equal-Area projection [17], centered at 50.5° N, 4.5° E. This coordinate conversion yields kilometre-standardised grid positions, enabling an undistorted mathematical evaluation of the horizontal longitudinal variance against the vertical latitudinal variance.

3. Results

To evaluate the core hypothesis that the literary representation of the Saxons underwent a structured transition from maritime dispersion to continental stabilisation, the validated dataset was subjected to non-parametric contingency testing.

Table 1. *Cross-Tabulation of Historical Period Versus Automated Geographical Setting Within the Corpus Saxonum ante DCC (N = 58).*

Period	Sea	Coast	Land	Total
250-400 CE	10 (ASR = 4.88)*	5 (ASR = 0.20)	3 (ASR = -4.11)*	18
400-550 CE	0 (ASR = -2.20)*	11 (ASR = 3.31)*	7 (ASR = -1.42)	18
550-700 CE	0 (ASR = -2.54)*	0 (ASR = -3.41)*	22 (ASR = 4.91)*	22
Total	10	16	32	58

Note. Environmental settings reflect the perceived spatial contexts within the text fragments as processed under a deterministic, zero-temperature large language model pipeline. Values in parentheses indicate Adjusted Standardised Residuals (ASR).* p <.05 after Holm-Bonferroni correction.

The cross-tabulation between the historical chronological cohort and the physical environmental setting exposes a profound spatial transition (χ^2 = 49.34, df = 4, N = 58, p <.001, V =.65; Table 1). During the early contact phase (250-400 CE), the Saxons exhibit a strong, statistically significant over-representation at 'Sea' (ASR = 4.88), while inland activities are systematically under-represented (ASR = -4.11). The subsequent post-Roman migration phase (400-550 CE) reflects a transitional shift, peaking significantly in coastal contact environments (ASR = 3.31). By the early medieval consolidation phase (550-700 CE), the selected corpus records an absence of explicit maritime contexts within this chronological cohort, shifting entirely toward land-based descriptions (ASR = 4.91).

Cross-referencing the automated geographical settings against the manually coded reference labels yields an exceptionally significant correlation with a robust effect size (χ^2 = 27.45, df = 4; N = 58, p <.001, V =.49; Table 2). The manual characterisation of 'Raider' is mathematically anchored within the automated maritime setting (ASR = 4.31), while activities deep on 'Land' are systematically under-represented for this role (ASR = -4.21). Conversely, the manual classification of 'Inhabitant' is linked to inland environments (ASR = 3.94) and systematically under-represented in open-sea contexts (ASR = -3.22).

Table 2. *Cross-Tabulation of Manual Discursive Labels Versus Automated Geographical Settings.*

Label	Sea	Coast	Land	Total
Raider	10 (ASR = 4.31)*	8 (ASR = 0.98)	4 (ASR = -4.21)*	22
Ally	0 (ASR = -1.41)	3 (ASR = 0.77)	5 (ASR = 0.44)	8
Inhabitant	0 (ASR = -3.22)*	5 (ASR = -1.51)	23 (ASR = 3.94)*	28
Total	10	16	32	58

Note. Classification of discursive labels ('Raider', 'Ally', 'Inhabitant') was executed through manual source-critical expert coding to capture narrative nuances, while geographical settings were processed via automated textual analysis. Values in parentheses indicate Adjusted Standardised Residuals (ASR). * p <.05 after Holm-Bonferroni correction.

Table 3. *Geographical Anisotropy and Spatial Variance of Coded Episodes Across Historical Periods.*

Period	Latitude	Longitude	F-ratio	N
250-400 CE	49.4° - 53.5° N	2.5° W - 8.5° E	3.45**	18
400-550 CE	45.7° - 54.0° N	1.2° W - 10.0° E	1.15	18
550-700 CE	50.8° - 55.7° N	2.1° W - 11.0° E	5.37***	22
Total	45.7° - 55.7° N	2.5° W - 11.0° E	2.30**	58

Note. Spatial parameters are bounded by the geographical horizons of the surviving textual record and indicate perceived distribution rather than raw population mapping. Degrees of freedom for each F-ratio are defined as $df_1 = df_2 = N - 1$. The F-ratio represents the ratio of longitudinal variance to latitudinal variance based on kilometre-standardised grid coordinates projected using the Lambert Azimuthal Equal-Area system [17].* p <.05, ** p <.01, *** p <.001.

When the exact database-validated coordinates from Appendix I are analysed spatially, the continental consolidation of the Saxon presence exhibits distinct chronological shifts in geometric distribution (Table 3). The total spatial spread across the longitudinal axis spans more than 13 degrees, stretching from 2.5° W in Wessex to 11.0° E in Thuringia. Temporally, the latitudinal footprint shifts from a compressed corridor (49.4°-53.5° N) during the early contact phase (250-400 CE) to a deeply expanded southern and northern boundary alignment (45.7°-54.0° N) during the post-Roman migration phase (400-550 CE), before stabilising into a concentrated northern continental and insular layout (50.8°-55.7° N) in the final early medieval consolidation phase (550-700 CE).

A two-sample F-test performed on the kilometre-standardised grid coordinates confirms the presence of statistically significant geographical anisotropy (Figure 1). For the total sample (N = 58), the overall horizontal longitudinal variance (s^2_x = 96,588.83 km^2) significantly outpaces the vertical latitudinal variance (s^2_y = 41,936.45 km^2), yielding a definitive directional stretching along the east-west axis (F(57, 57) = 2.30, p =.001).

To evaluate whether the transition from late antique empires to early medieval successor states introduced spatial systematic variance into the written record, an independent samples t-test was executed. The dataset was stratified into Mediterranean imperial traditions (East Roman and West Roman; N = 32) versus regional successor state and insular traditions (Frankish, Anglo-Saxon, Ostrogothic, Visigothic, British, Burgundian, and Lombardic; N = 26). The analysis revealed a statistically significant difference in the longitudinal centre of gravity, t(56) = -2.99, p =.0041. Imperial text producers tracked a significantly more westward longitudinal focus (M = 2.55° E) than their early medieval successor state counterparts (M = 5.87° E).

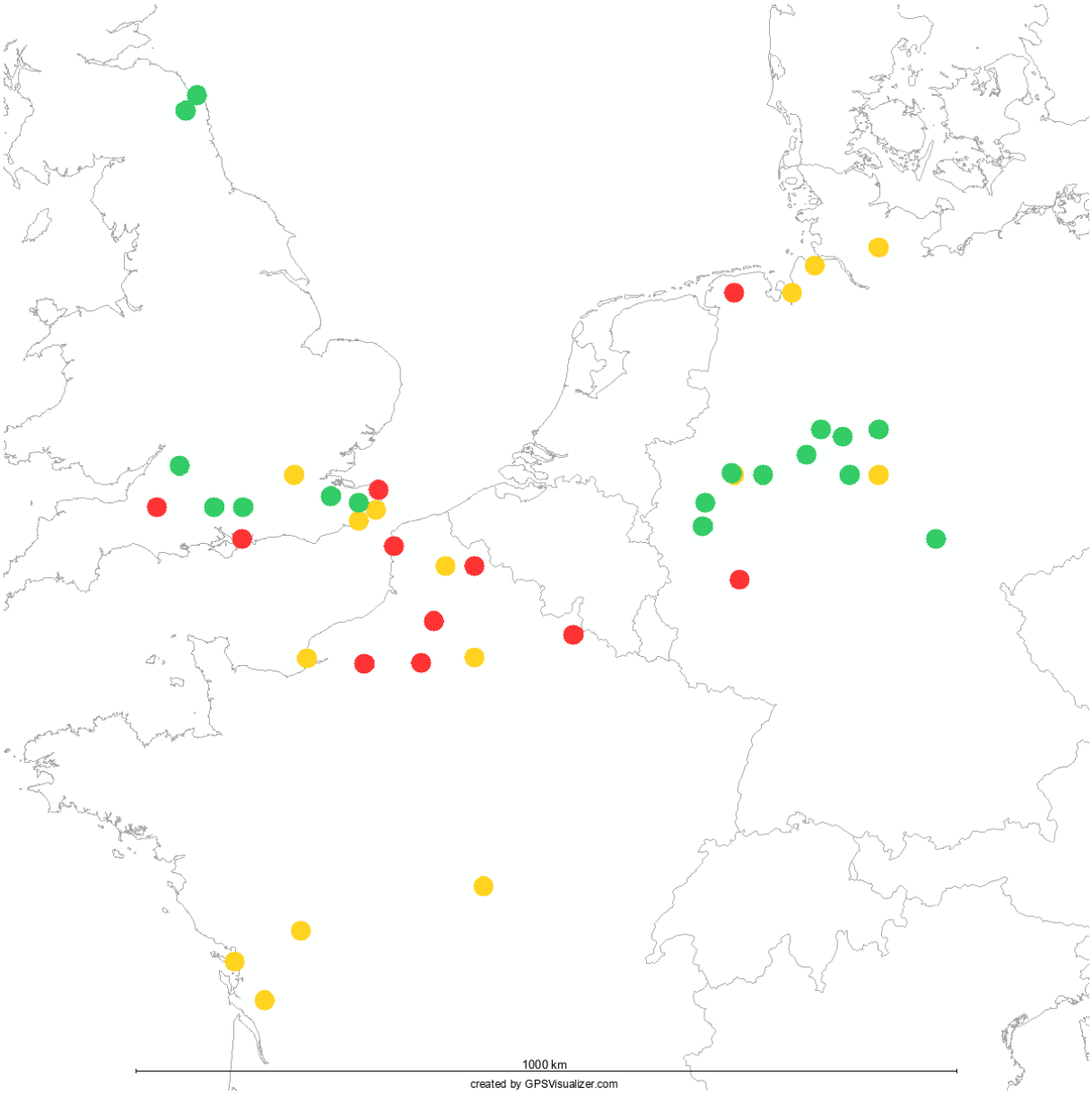

Figure 1. *Approximate Spatial Distribution of Documented Saxon Textual Entries (250-700 CE).*

Note. Perceived geopolitical centres of gravity across the 58 analysed Saxon textual entries (250-700 CE) from the 'Corpus Saxonum ante DCC' (Appendix I), indicated in red (250-400 CE), orange (400-550 CE), and green (550-700 CE). Map data visualised via GPS Visualizer [9] (https://gpsvisualizer.com, retrieved 28 August 2026).

4. Discussion

The quantitative results indicate a structured transition in the literary representation of the Saxons within the analysed corpus. This pattern is compatible with the theoretical framework of the meta-ethnic frontier [20], which posits that collective identities in external sources often crystallise within zones of military, economic, and administrative friction. The spatial analysis reveals a dynamic, multi-phasic geographical evolution across the selected entries: during the early contact phase (250-400 CE), the recorded friction vector remains largely bound within a compressed horizontal corridor (49.4°-53.5° N), closely associated with the defensive perimeter of the Roman Litus Saxonicum and its immediate maritime networks.

Following the end of the Western Empire, the onset of Frankish expansion and the formation of Anglo-Saxon kingdoms in the final phase (550-700 CE) re-established severe horizontal anisotropy ($F(21, 21) = 5.37$, $p < .001$). This early medieval elongation along the longitudinal axis (2.1° W-11.0° E), while remaining compressed within a northern latitudinal corridor (50.8°-55.7° N), marks the consolidation of sedentary land-dwelling populations interacting with the expanding Merovingian state apparatus, supporting the semantic deconstruction advanced by Matthias Springer [18]. The analytical validity of this underlying qualitative taxonomy is robustly

confirmed by the near-perfect algorithmic replication index (κ =.89). Rather than introducing systemic distortion, the minor semantic discrepancies isolated during the validation process—such as the classification of Salvianus or Gregory of Tours—empirically underscore the fluid boundaries between active external friction and stable territorial status within contemporary rhetoric.

This corridor effect offers a critical methodological insight into the relationship between written records and physical reality. As historically and archaeologically proven, the Roman military apparatus maintained a continuous physical presence along the lower Rhine corridor in the southern Netherlands (51.6°-51.9° N, 5.5°-6.0° E) during the early contact phase (250-400 CE)—including a verified castellum at Cuijk, constructed around 300 CE, and a documented fortification layout at Nijmegen [21]. Yet, this active frontline remains completely silent regarding the 'Saxon' ethnonym within the compiled dataset. This divergence demonstrates that textual entries inside the database reflect administrative selection rather than objective settlement borders, though the lean sample size (N = 58) necessitates treating these patterns as heuristic trends within this specific digital collection rather than definitive proof of macro-historical shifts.

Consequently, while the structural lack of data points across the adjacent hinterland of present-day Belgium (50.5°-51.5° N, 3.0°-5.5° E) simply reflects the limits of Mediterranean text producers and alternative regional power structures, it opens a highly specific geographical contrast with regions further north. Unlike the institutional silence of the southern hinterland, the absolute literary vacuum across the north-western Dutch coastal region and the Rhine-Meuse delta (51.5°-53.0° N, 4.0°-6.0° E) co-occurs with a possible ecological explanation. This northern shadow zone corresponds precisely with Van Lanen's [23] landscape-archaeological models of severe regional waterlogging and river avulsions, which actively dissolved the socio-economic connectivity and elite networks required to generate external administrative discourse [14]. It should also be noted that digital gazetteers necessarily code historical toponymic entries as fixed coordinates or geographic centroids. This abstraction into point-based data potentially masks the real, diffuse spatial dynamics within the Lower Rhine and Belgic zones, which may partly account for the observed local 'literary vacuums.

This functional interdependence between direct physical interaction and textual generation aligns with Turchin's [19] structural-demographic framework of the meta-ethnic frontier. According to this model, collective external categorisation and intensified socio-political discourse do not emerge uniformly across geographical space; instead, they crystallise within acute zones of intense geopolitical friction. As modeled by the automated pipeline under a restricted, low-variance setting to optimise semantic consistency, the text-generating matrix of the classical Mediterranean world required high-stakes military or fiscal contact vectors to actively apply and recycle peripheral labels like 'Saxon'. Where these friction vectors collapsed—whether through the natural disruption of regional infrastructure in the flooded northern delta or through the administrative limits of the inland Belgic corridor—the bureaucratic and rhetorical impetus to record peripheral groups dissolved simultaneously. Consequently, the shifting spatial-temporal distribution of the ethnonym mapped in this study does not reflect objective regional settlement borders. It mirrors the changing spatial layout of imperial friction boundaries, where historical text production was contextually applied as an administrative and rhetorical shorthand by southern text producers [24].

The total geographical anisotropy calculated across the entire dataset (F(57, 57) = 2.30, p =.0010) thus maps a structural shift in imperial perceptions and administrative focus rather than a physical population continuum. This structural alignment appears heavily dependent on the shifting political locus of text production. This significant longitudinal divergence empirically validates the critiques of Noble [15] and Geary [8]. The label 'Saxon' functioned precisely as an external administrative shorthand that text producers routinely recycled and anchored where imperial administrative or fiscal friction was most acute, rather than tracking an objective tribal reality. Rather than representing a uniform transmission stream, the data suggests that late antique Mediterranean observers focused primarily on the western channel networks and maritime friction points. Conversely, the significant eastward shift recorded by early medieval successor state and insular traditions tracks the territorial re-centring of the ethnonym. This pattern corresponds directly with the expansion of the Merovingian state apparatus and continental consolidation processes, supporting the framework that the perceived geography of peripheral groups mirrors the changing borders of regional geopolitical interest [15, 24].

5. Conclusion

This exploratory study demonstrates that applying quantitative spatial analysis to late antique and early medieval textual corpora is less about charting physical human migrations and more about mapping the shifting borders of historical text production itself. By translating narrative episodes from the 'Corpus Saxonum ante DCC' into geographical coordinates, the analysis shows a text-based statistical transition from a maritime 'Raider' focus to an inland 'Inhabitant' baseline. Rather than accepting these statistical trends as mirrors of objective demographic distributions, this framework invites a critical reflection on how administrative rhetoric was structured by proximity to geopolitical interest. Admittedly, undertaking quantitative analysis on a lean early medieval sample (N = 58) introduces intrinsic limitations, meaning these statistical configurations must be treated as provisional, heuristic tools rather than absolute historical declarations. However, this experimental approach underscores that digital cartography and spatial statistics can offer a transparent, replicable layer to traditional source criticism. Ultimately, this study suggests that early medieval tribal ethnonyms functioned not as static bio-

logical markers, but as fluid, external shorthands that expanded and contracted alongside the changing horizons of regional political interest.

Abbreviations

API	Application Programming Interface
ASR	Adjusted Standardised Residual
CE	Common Era
DARE	Digital Atlas of the Roman Empire
df	Degrees of Freedom
GIS	Geographic Information Systems
JSON	JavaScript Object Notation
LLM	Large Language Model

Acknowledgments

I would like to express my sincere thanks to J. van Hout, the driving force behind the historical platform 'Nifterlaca' (http://nifterlaca.nl). His dedicated work on the digital 'Corpus Saxonum ante DCC' over the past five years provided the essential foundation for this study.

Author Contributions

Robert Jan Duchateau: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Resources, Software, Visualization, Writing – original draft, Writing – review & editing

Data Availability Statement

The primary data underlying this study were harvested from the digital 'Corpus Saxonum ante DCC' compiled by J. van Hout, which is accessible from 'Nifterlaca' at https://nifterlaca.nl/eigen_assets/Saksen/Saksen_indexpagina.html. The coded dataset containing the final ($N = 58$) source entries analysed in this paper is fully reproduced within Appendix I of this manuscript.

Conflicts of Interest

The author declares no conflicts of interest.

Appendix

Appendix I: Data Selection and Spatial Processing

The establishment of the analytical dataset followed a structured, two-step exploratory protocol designed to isolate qualitative historical contextualisation from automated database-driven spatial logging across the 58 unique historical episodes. Because of the inherent limitations of the sample size ($N = 58$), this dual approach was implemented to maintain transparent control over the source-critical limitations of the material. In the first phase, individual textual units were manually isolated from the digital 'Corpus Saxonum ante DCC' [5, 11]. These entries were subsequently evaluated and classified based on nominal source characteristics, specifically authorial perspective and the assigned discursive label. This manual step was essential to ensure that the primary qualitative framework remained grounded in contemporary source-critical scholarship rather than algorithmic text extraction.

To ensure algorithmic reproducibility and mitigate potential bias, the automated extraction pipeline utilised specified text fragments from the 'Corpus Saxonum ante DCC' directly in their original Dutch phrasing, preventing translation loss. The precise prompt utilised via the Google AI Studio API was structured as follows: *'Extract the primary late antique or early medieval toponym from the text fragment and cross-reference it with the Pleiades Gazetteer or the Digital Atlas of the Roman Empire (DARE) to provide its corresponding decimal coordinates and Gazetteer ID. Provide the output strictly in a structured JSON format using the keys "toponym", "environmental_setting" ('Sea', 'Coast', or 'Land'), "coordinates", and "gazetteer_id" based solely on the provided text.'* To detect model errors and systematic hallucinations, a hybrid human-AI consensus protocol was deployed, designed to neutralise the respective blind spots of both algorithmic extraction and human coder bias. While the initial coordinate and gazetteer cross-referencing was executed by the model, all outputs were subjected to a subsequent human expert review. Where minor spatial or semantic deviations occurred between the automated extraction and human cross-checking, these discrepancies were systematically reconciled and balanced against the certified online registries of Pleiades and DARE. This mutual validation workflow ensured that coordinates and boundary centroids were mathematically and historically synchronised prior to statistical analysis. The model operated entirely without external contextual information beyond the individual text fragments provided in each API call.

In the second phase, an automated extraction pipeline utilised Google Gemini 1.5 Pro to process the raw text fragments under strict, deterministic constraints. The model was deployed to extract the primary late antique or early medieval toponym or regional identifier indicating the perceived focus of the Saxon presence within the specific narrative episode. These extracted historical place names were subsequently cross-referenced with certified digital gazetteers, specifically the Pleiades Gazetteer and the Digital Atlas of the Roman Empire (DARE), resolving each entry into decimal geographic coordinates consisting of latitude and longitude [2, 3]. Broad provincial territories or shifting frontier zones were systematically standardised to the calculated centroids of their documented late antique boundaries to ensure internal consistency. While these coordinates enable statistical variance testing,

they are treated methodologically as proxies for the geographical horizons of the authors rather than objective physical borders.

Table A1. *Coded Historical Episodes from the 'Corpus Saxonum ante DCC' [5, 22] Featuring Expert Qualitative Coding and Database-Validated Geographic Coordinates.*

ID	Period	Source	Author	Perspective	Label	Setting	Location	Gazetteer ID
01	250-400	3	Julianus	East Roman	Raider	Sea	49.90° N, 2.30° E	Pleiades 108828
02	250-400	4	Eutropius	East Roman	Raider	Sea	50.72° N, 1.61° E	DARE 108818
03	250-400	5	Ps-Hegesippus	West Roman	Raider	Coast	53.50° N, 8.50° E	Pleiades 981513
04	250-400	7	Hieronymus	West Roman	Ally	Coast	50.50° N, 3.00° E	Pleiades 981523
05	250-400	8	Ambrosius	West Roman	Inhabitant	Land	49.75° N, 4.72° E	Pleiades 108894
06	250-400	9	Pacatus	West Roman	Raider	Sea	50.72° N, 1.61° E	DARE 108818
07	250-400	10	Ammianus	East Roman	Raider	Sea	51.34° N, 1.34° E	Pleiades 109293
08	250-400	11	Hieronymus	West Roman	Inhabitant	Land	49.44° N, 2.08° E	Pleiades 108804
09	250-400	12	Symmachus	West Roman	Inhabitant	Land	50.35° N, 7.59° E	Pleiades 108933
10	250-400	13	Claudianus	West Roman	Raider	Sea	50.80° N, 1.02° W	DARE 109255
11	250-400	14	Claudianus	West Roman	Raider	Sea	50.80° N, 1.02° W	DARE 109255
12	250-400	15	Claudianus	West Roman	Raider	Sea	50.80° N, 1.02° W	DARE 109255
13	250-400	16	Claudianus	West Roman	Raider	Sea	50.80° N, 1.02° W	DARE 109255
14	250-400	17	Claudianus	West Roman	Raider	Sea	50.80° N, 1.02° W	DARE 109255
15	250-400	18	Marcianus	East Roman	Inhabitant	Coast	53.50° N, 8.50° E	Pleiades 981513
16	250-400	19	Hieronymus	East Roman	Raider	Sea	50.72° N, 1.61° E	DARE 108818
17	400-550	20	Orosius	West Roman	Raider	Coast	53.50° N, 7.50° E	Pleiades 1021
18	400-550	21	Pelagius	West Roman	Inhabitant	Coast	51.15° N, 1.00° W	Pleiades 108861
19	400-550	22	Notitia dignitatum	West Roman	Ally	Coast	51.12° N, 1.30° E	DARE 109257
20	400-550	23	Salvianus	West Roman	Inhabitant	Coast	49.49° N, 0.10° E	DARE 108855
21	400-550	24	Chronica Gallica	West Roman	Raider	Coast	51.50° N, 0.12° W	Pleiades 981519
22	400-550	25	Sidonius	West Roman	Raider	Coast	46.50° N, 0.00° E	Pleiades 138435
23	400-550	26	Sidonius	West Roman	Raider	Coast	46.99° N, 3.16° E	Pleiades 118803
24	400-550	27	Sidonius	West Roman	Raider	Coast	46.16° N, 1.15° W	DARE 138541
25	400-550	28	Constantius	West Roman	Inhabitant	Land	51.00° N, 1.00° E	Pleiades 108882
26	250-400	29	Zosimus	East Roman	Raider	Coast	49.43° N, 1.10° E	Pleiades 109282
27	400-550	30	Cassiodorus	Ostrogothic	Raider	Land	51.50° N, 9.50° E	Pleiades 981513
28	400-550	30½	Tribes List	East Roman	Inhabitant	Coast	53.50° N, 8.50° E	Pleiades 981513
29	400-550	31	Ennodius	Ostrogothic	Raider	Land	51.50° N, 9.50° E	Pleiades 981513
30	400-550	32	Theudebert	Ostrogothic	Raider	Land	51.50° N, 10.00° E	DARE 22692
31	250-400	33	Gildas	British	Ally	Coast	51.15° N, 2.50° W	Pleiades 109159
32	400-550	34	Vita Bibiani	Visigothic	Raider	Land	45.74° N, 0.63° W	Pleiades 138459
33	400-550	35	Jordanes	East Roman	Inhabitant	Land	54.00° N, 10.00° E	Pleiades 981545

ID	Period	Source	Author	Perspective	Label	Setting	Location	Gazetteer ID
34	400-550	36	Jordanes	East Roman	Inhabitant	Land	49.50° N, 3.00° E	Pleiades 108836
35	550-700	37	Marius	Burgundian	Ally	Land	51.50° N, 7.50° E	DARE 25776
36	550-700	38A	Gregorius	Frankish	Ally	Land	51.50° N, 9.50° E	Pleiades 981513
37	550-700	38B	Gregorius	Frankish	Inhabitant	Land	50.80° N, 11.00° E	DARE 22692
38	550-700	38C	Gregorius	Frankish	Ally	Land	51.52° N, 7.46° E	DARE 25776
39	550-700	38D	Gregorius	Frankish	Inhabitant	Land	51.72° N, 8.75° E	DARE 25754
40	550-700	38E	Gregorius	Frankish	Inhabitant	Land	51.92° N, 9.38° E	Pleiades 981504
41	550-700	39	Venantius	Frankish	Inhabitant	Land	51.50° N, 9.50° E	Pleiades 981513
42	550-700	40	Venantius	Frankish	Inhabitant	Land	51.20° N, 7.00° E	Pleiades 118861
43	550-700	41	Isidorus	Visigothic	Inhabitant	Land	52.00° N, 10.00° E	Pleiades 981513
44	550-700	42	Jonas	Lombardic	Inhabitant	Land	51.50° N, 9.50° E	Pleiades 981513
45	550-700	43	Vitalianus	East Roman	Ally	Land	51.15° N, 1.00° W	Pleiades 108861
46	550-700	44	Dado	Frankish	Inhabitant	Land	52.00° N, 9.00° E	Pleiades 981513
47	550-700	45	Julian	Visigothic	Ally	Land	51.50° N, 8.00° E	Pleiades 981523
48	550-700	46	Passio Praejecti	Frankish	Inhabitant	Land	51.50° N, 9.50° E	Pleiades 981513
49	550-700	47	Epitaph Caedwalla	Anglo-Saxon	Inhabitant	Land	51.15° N, 1.50° W	Pleiades 109142
50	550-700	48	Vita Domnae	Frankish	Inhabitant	Land	51.50° N, 9.50° E	Pleiades 981513
51	550-700	51	Vita Fursei	Frankish	Inhabitant	Land	51.20° N, 1.00° E	DARE 2101
52	550-700	52	AS Charters	Anglo-Saxon	Inhabitant	Land	51.27° N, 0.52° E	Pleiades 108914
53	550-700	53	Aldhelm	Anglo-Saxon	Inhabitant	Land	51.60° N, 2.10° W	Pleiades 109142
54	550-700	54	Adomnán	Anglo-Saxon	Inhabitant	Land	55.50° N, 2.00° W	Pleiades 981535
55	550-700	55	Vita Cuthberti	Anglo-Saxon	Inhabitant	Land	55.67° N, 1.80° W	Pleiades 981535
56	400-550	-	Stephanus	East Roman	Inhabitant	Coast	53.80° N, 8.90° E	Pleiades 981513
57	400-550	-	Johannes	East Roman	Raider	Coast	50.50° N, 2.50° E	Pleiades 981523
58	550-700	-	Lex Ribuaria	Frankish	Inhabitant	Land	50.94° N, 6.96° E	Pleiades 108880

Inter-Coder Reliability

To test the replicability and semantic clarity of the operational definitions, the nominal variable Label was subjected to an independent automated validation protocol. A secondary, blind classification of the 58 historical episodes was executed using the Google Gemini 1.5 Pro Large Language Model (version gemini-1.5-pro-001) via the Google AI Studio API. To ensure algorithmic determinism, the model operated under a zero-temperature configuration (temperature = 0.0, top_p = 1.0), processing the data strictly based on the criteria in Appendix II.

The cross-comparison yielded a Cohen's Kappa (κ) of.89 (with 93.1% raw agreement), indicating a 'near-perfect' alignment [13] and demonstrating that the qualitative categories are robust against subjective coder bias. Systematic consensus was achieved across 54 of the 58 entries. Four critical semantic discrepancies were isolated where the human expert coding and the deterministic AI logic diverged:

ID 05 (Ambrosius): Human coder logged Inhabitant; AI control logged 'Raider'. The deterministic pipeline flagged the late fourth-century rhetorical coupling of peripheral populations with acute external threat vectors, whereas the human coder prioritised the broader geographical cataloguing context.

ID 20 (Salvianus): Human coder logged 'Inhabitant'; AI control logged 'Raider'. The AI model classified the passage based on Salvianus's explicit emphasis on Saxon ferocity (*feritas*), interpreting it as an active existential and military

threat, whereas the human coder applied a more passive contextual reading.

ID 36 (Gregorius I): Human coder logged 'Ally'; AI control logged 'Inhabitant'. The automated classifier targeted the overarching narrative representation of a geophysically settled, sedentary population base, whereas the human expert prioritised the localised tactical military cooperation documented in the text.

ID 45 (Vitalianus): Human coder logged 'Ally'; AI control logged 'Inhabitant'. The algorithmic pipeline framed the Mediterranean perspective through a broad territorial lens, whereas the human coder isolated an active treaty-bound military operational status.

Rather than exposing data invalidity, these four isolated deviations underscore the fluid and contested nature of late antique and early medieval administrative rhetoric, where the boundaries between active external aggression ('Raider') and sedentary stabilisation ('Inhabitant') are inherently ambiguous.

Appendix II: Operational Codebook

To ensure complete methodological transparency, replicability, and the statistical validity of the cross-tabulation and spatial variance analyses, all variables analysed within the 'Corpus Saxonum ante DCC' [5, 22] were operationally defined and classified according to strict semantic and geographic criteria.

Manual Coding (Author)

Period: The historical or narrative setting of the described action, capturing the chronological placement reported by the source author. To isolate macro-developments, this nominal variable is stratified into three consecutive 150-year chronological cohorts: the early contact phase (250-400 CE), the post-Roman migration phase (400-550 CE), and the early medieval territorial consolidation phase (550-700 CE).

Perspective: A nominal variable mapping the socio-political, geographical, or ethnic background of the text producer or source document (e.g., East Roman, West Roman, Frankish, Ostrogothic, Anglo-Saxon, or British). This variable captures the external or internal framing of the narrative episode.

Label: The discursively constructed role attributed to the Saxons within the specific textual unit. To ensure inter-coder reproducibility, classification relies on explicit narrative framing rather than implicit historical deductions:

'Raider': Coded when the narrative unit focuses layout-centrically on piracy, maritime aggression, looting, systemic violence, or an acute external security threat. This category prioritises explicit lexical triggers of hostility or ferocity, even if embedded within broader geographical descriptions.

'Ally': Coded strictly when the text explicitly characterises the Saxons as active treaty parties, formal political partners, or operational military auxiliaries engaged in joint tactical warfare.

'Inhabitant': Coded when the narrative focuses on a geophysically settled, sedentary population base with a permanent geographical, agricultural, or territorial presence. This category includes localised regional descriptions where the entity operates as an established, domestic neighbor or structural political component within a territorial state framework, overriding minor or localised tactical military activities.

Automated Pipeline Extraction (AI and Gazetteer Integration)

Setting: The physical-geographical sphere in which the described historical interaction occurs. This nominal variable was automatically extracted from the text fragments using the Google Gemini model series under deterministic, zero- temperature constraints to minimise classification variance:

'Sea': Maritime contexts, open waters, and open sea routes primarily associated with fleet operations.

'Coast': Intertidal contact zones, estuaries, river mouths, and immediate coastal transition corridors.

'Land': Inland, continental, or territorial settings indicating agriculture, permanent settlement, or domestic military conflicts.

Location: The precise spatial tracking of the active, perceived centre of gravity of the historical action involving the Saxons, automatically resolved via toponym extraction and external database verification: *Latitude:* Quantified in decimal degrees North (°N) to fix the estimated latitudinal position relative to major ecological corridors and the Rhine-Meuse delta systems. *Longitude:* Quantified in decimal degrees East (°E) or West (°W) to measure the horizontal distribution of the reported geopolitical friction front. For the two-sample *F*-test evaluation, these raw coordinates were projected into a metric, kilometre-standardised equal-area grid relative to a central meridian to prevent spatial distortion caused by meridian convergence at mid-latitudes. Methodologically, these points represent the geographical awareness and rhetoric of the text producers rather than objective physical borders.

References

[1] Agresti, A. (2013). Categorical data analysis (3rd ed.). John Wiley & Sons.

[2] Åhlfeldt, J. (n.d.). Digital Atlas of the Roman Empire (DARE). Centre for Digital Humanities, University of Gothenburg. Available at: https://imperium.ahlfeldt.se (Accessed 26 August 2026).

[3] Bagnall, R. S., Talbert, R. J. A., & Elliott, T. (Eds.). (n.d.). Pleiades: A gazetteer of past places. Institute for the Study of the Ancient World. Available at: https://pleiades.stoa.org (Accessed 26 August 2026)

[4] Cochran, W. G. (1954). Some methods for strengthening the common x2 tests. Biometrics, 10, 417-451. https://doi.org/10.2307/3001616

[5] Dagobert [Van Hout, J.]. (2026, August 20). Drie nieuwe bronnen [Online forum post]. Nifterlaca. Available at: https://www.nifterlaca.nl/read.php?3,25172,25594#msg-25594 (Accessed 24 August 2026)

[6] Elsevier. (n. d.). Scopus abstract and citation database. Available at: https://www.scopus.com (Accessed 25 August 2026)

[7] Fleming, R. (2010). Britain after Rome: The fall and rise, 400-1070 (The Penguin History of Britain). Allen Lane.

[8] Geary, P. J. (2002). The myth of nations: The medieval origins of Europe. Princeton University Press.

[9] GPS Visualizer. (n. d.). GPS Visualizer [Web application]. Available at: https://gpsvisualizer.com (Accessed 28 August 2026)

[10] Halsall G. (2007). Barbarian Migrations and the Roman West, 376-568. Cambridge University Press. https://doi.org/10.1017/CBO9780511802393

[11] Higham, N. J., & Ryan, M. J. (2013). The Anglo-Saxon world. Yale University Press. http://www.jstor.org/stable/j.ctt32bt51

[12] Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. Scandinavian Journal of Statistics, 6(2), 65-70. http://www.jstor.org/stable/4615733

[13] Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. Biometrics, 33(1), 159-174. https://doi.org/10.2307/2529310

[14] Nicolay, J. A. W. (2014). The splendour of power: Early medieval kingship and the use of gold and silver in the southern North Sea area (5th to 7th century AD) (Groningen Archaeological Studies - Vol. 28). Barkhuis.

[15] Noble, T. F. X. (Ed.). (2006). From Roman provinces to medieval kingdoms. Routledge. https://doi.org/10.4324/9780203322956

[16] Pohl, W. (1998). Telling the difference: Signs of ethnic identity. In W. Pohl & H. Reimitz (Eds.), Strategies of distinction: The construction of ethnic communities in early medieval Europe (pp. 17-69). Brill. https://doi.org/10.1163/9789004609518_004

[17] Snyder, J. P. (1987). Map projections: A working manual (U.S. Geological Survey Professional Paper 1395). U.S. Government Printing Office. https://doi.org/10.3133/pp1395

[18] Springer, M. (2004). Die Sachsen. W. Kohlhammer Verlag. https://doi.org/10.17433/978-3-17-023227-3

[19] Turchin, P. (2003). Historical dynamics: Why states rise and fall. Princeton University Press. https://doi.org/10.1515/9781400889310

[20] Turchin, P. (2006). War and peace and war: The rise and fall of empires. Plume.

[21] Van Enckevort, H., Hendriks, J., & Nicasie, M. (2017). Nieuw licht op donkere eeuwen: De overgang van de laat-Romeinse tijd naar de vroege middeleeuwen in Zuid-Nederland [New Light on the Dark Ages: The Transition from the Late Roman Period to the Early Middle Ages in the Southern Netherlands] (Nederlandse Archeologische Rapporten 58). Rijksdienst voor het Cultureel Erfgoed. https://doi.org/10.17026/DANS-XWZ-6Y5B

[22] Van Hout, J. (n.d.). Saksen: In de oudheid en de Merovingische periode: Corpus Saxonum ante DCC. Nifterlaca. Available at: https://www.nifterlaca.nl/eigen_assets/Saksen/Saksen_indexpagina.html (Accessed 17 August 2026)

[23] Van Lanen, R. J. (2020). Revealing the past through modelling? Reflections on connectivity, habitation and persistence in the Dutch delta during the 1st millennium AD. Netherlands Journal of Geosciences, 99, e14. https://doi.org/10.1017/njg.2020.12

[24] Wickham, C. (2005). Framing the Early Middle Ages: Europe and the Mediterranean 400-800. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199264490.001.0001

[25] Wood, I. (1994). The Merovingian Kingdoms 450 - 751 (1st ed.). Routledge. https://doi.org/10.4324/9781315836539