From Geospatial Patterns to Ancient Signals: A Signal Based Framework for Archaeological Machine Learning

1st Dylan Foley
Dept. of Computing
Atlantic Technological University
Letterkenny, Ireland
dylangfoley@gmail.com

2nd Dr. Eoghan Furey
Dept. of Computing
Atlantic Technological University
Letterkenny, Ireland
eoghan.furey@atu.ie

***Abstract*—To enable machine learning in archaeology, data must encode both observed traces—what we detect on Earth's surface today—and their inferred past states. We propose a signal-based framework where archaeological features are treated as degraded spatiotemporal signals emitted by past human activity, and archaeology becomes a form of inverse signal reconstruction. Unlike traditional typological or period-based models, this approach models each site's history as a vector in a unified 3D spatial manifold with temporal relationships anchored to a fixed reference (J2000), enabling integration with astronomical models and signal processing techniques.**

Applying this to northwest Ireland, we test signal correlations between archaeological distributions and hypothesised territorial boundaries circa (t_0+400) (1600 CE). Using contiguity matrices, kernel density estimation, and K-nearest neighbor graph analysis, we find significant alignments between spatial signal clusters and historic boundaries across a 6,000-year span. This suggests that we can greatly increase the scope, resolution, and accuracy of our temporal pattern analyses even in noisy, incomplete data.

By framing archaeology as an inverse signal reconstruction problem, our model aligns with broader signal science paradigms, including SETI's search for technosignatures. It enables scalable, quantitative inference of long-term human territorial behavior and is extensible to other domains dealing with temporal degradation.

***Index Terms*—Archaeological signal processing, spatial manifold embedding, temporal-spatial transformation, technosignatures, machine learning, territorial analysis, inverse signal reconstruction, SETI**

I. Introduction

Archaeological data structures remain rooted in 19th-century typologies and broad temporal labels (e.g., "Bronze Age," "ringfort"), limiting their compatibility with signal science and machine learning. These frameworks obscure uncertainty, mask temporal dynamics, and blend observed and inferred data, undermining reproducibility. Archaeology needs to move beyond these paradigms to enable automated analysis and integration with remote sensing [1].

The non-repeatable nature of excavation—due to its destructive process—makes it essential to capture data with as much detail as possible at the time of recording. While

Fig. 1: Rathra, Co Roscommon, age unknown, a good example of ancient motion traces (signals) with later interference from fencelines. Current archaeological data models find it difficult to represent this temporal complexity effectively

excavation techniques are well developed, recording systems are not standardised and are incompatible with computational analysis. Integration with scientific approaches is patchy and is slowed by resistance to scientific paradigms within the discipline [2]. Therefore, archaeology still lacks a formal system to model data as dynamic, evolving systems [3].

Archaeological signals are complex ancient patterns, subject to natural decay, and masking from vegetation, animal activity, and later human activities. An example can be seen in Rathra, Co. Roscommon **??**, where a sequence of ancient and modern features overly each other. While archaeological survey techniques capture sites in detail, they do not capture temporal data except as discrete layers, creating disconnected snapshots. The challenge that must be solved to progress archaeological science, is to create a data model that can represent this temporal dimension in computable form [4].

We propose reframing archaeological inference broadly as an inverse signal reconstruction problem. Material traces (structures, artefacts) are modelled as signals of past human activity (motion patterns) —subject to environmental decay. Archaeology must reconstruct these signals to enable probabilistic inference of their time of origin and the patterns of motion that produced them.

We anchor our model to a fixed celestial reference frame (J2000) to support integration with astronomical datasets—enabling interdisciplinary integration. This includes with palaeontology, systems biology and in line with recent calls from astronomers in SETI (Search for Extra Terrestrial Intellignece), notably Douglas Vakoch, for archaeological input in the interpretation of technosignatures, signals created by technological activity [5]. Vakoch's work positions archaeology as an essential field for detecting and understanding non-human intelligence.

More prosaically, the challenge of interpreting degraded signals from intelligent activity is fundamental to understanding our own evolutionary past. Machine learning is increasingly used in palaeo-anthropological research on human evolution, but not always successfully. Studies spanning millions of years with extremely ambiguous signals will greatly benefit from ML techniques, but issues with data have led to calls for interdisciplinary studies with ML expertise included. [6]

II. Approach and Data Development

A. Temporal Representation in Current Models

Time remains a core challenge in both archaeology and computer science. Archaeological models reduce temporal complexity to epochs, while ML frameworks encode time abstractly as indexes or sinusoids, disconnected from physical space. Relational databases struggle with intervals and uncertainty, and deep learning models assume temporal variable independence—impossible in archaeological data [7].

Deep Learning applications on Irish ringfort data achieved high accuracy on clean data but poor performance on partial/noisy data, with high false positive rates (roundabouts identified as sites) due to missing temporal dimensions. False negatives occur from signal interference, vegetation, or truncation [9]. GIS temporal representation creates splitting effects on temporal data, making continuous change modelling difficult [10].

Spatial autocorrelation challenges geographic and temporal data by breaking ML's independent variable assumptions [11]. Both spatial and temporal data present fundamental ML challenges through autocorrelation combined with non-stationarity, irregular sampling, and high-dimensional feature spaces violating independence assumptions. Recent advances (Time2Vec, Neural ODEs, spatiotemporal Transformers, diffusion models) improve time series handling but lack physical grounding [12]. Most treat time as abstract sequence labels, not trajectories, and cannot handle uncertain, missing, or varying time series data [13]. Our approach contrasts by anchoring time as spatially-referenced vectors, enabling coherent geometric reasoning over spatio-temporal fields.

Conceptual Framework

We define a 3D spatial manifold $\mathcal{M}$ where archaeological entities are embedded as spatiotemporal signals. Time as not treated as a separate dimension, instead we transform temporal information into spatial relationships by representing each object's history as a motion vector extending from its present surface location on Earth.

Archaeological signals are thus represented as vectors originating from positions distant in the past and terminating at their present locations on the Earth's surface. This approach is analogous to astronomy, where we observe signals that originated at sources in the distant past. By anchoring the model to a fixed celestial reference frame (J2000), we establish a consistent spatial orientation that abstracts away Earth's orbital motion, treating the planet as a static reference point within the embedded space.

This temporal-to-spatial transformation allows time relationships to be expressed as geometric relationships, enabling the application of spatial reasoning and physical transformations to archaeological data while avoiding the complexities of discrete time slices and temporal databases.

Core Principles:

1) **Temporal Elimination**: Traditional archaeological epochs ("Bronze Age") and computational timestamps are replaced by spatial vector relationships, avoiding the limitations of both cultural periodization and discrete temporal databases.
2) **Signal Degradation**: Archaeological entities detected on Earth's surface are treated as degraded signals of past human activity, requiring inverse modeling to infer original states.
3) **Typological Neutrality**: Cultural labels ("ringfort", "barrow") are replaced by measurable feature vectors ϕ_i, promoting objective classification.

Formal Definition: Each archaeological signal s_i is defined by spatial coordinates $x_i \in \Omega$ (Earth's surface), temporal depth τ_i (referenced to J2000), and observable features ϕ_i. The signal field $F : \Omega \to \mathcal{P}(S)$ maps spatial locations to detected signals, where $S = \{s_i\}_{i=1}^{N}$.

III. Implementation and Data Preprocessing

A. Temporal Unification

We collapsed typological and categorical period labels into scalar temporal estimates referenced to the J2000 epoch (t_0). This converts legacy periodisations into a continuous, signal-compatible temporal field. Similar to how time-frequency transforms in signal processing convert temporal signals to spatial/frequency representations for analysis. We embed archaeological temporal relationships as spatial coordinates to enable geometric reasoning over historical processes.

B. Signal Representation

Each archaeological observation is stripped of its typological label and encoded as a signal defined by location, geometry, scale, and degradation indicators. Human-readable names are preserved only for cross-referencing. Probabilistic interpretations (e.g., function or cultural affiliation) are kept separate from the observed signal data.

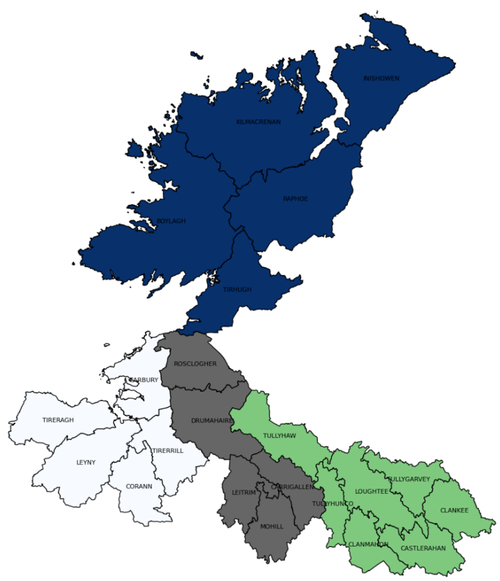

Fig. 2: Map showing the Gaelic confederated territories as they were at around the year 1600 CE ($t_0 + 400$). The territorial boundaries were reconstructed using historical maps.

C. Territorial Inference from Signal Field

Using known territorial divisions from circa 1600 CE ($t_0 + 400$) in northwest Ireland, we applied spatial analysis techniques to assess whether signal distributions—regardless of their temporal depth—retain structural alignment with these boundaries. This included KDE, border-proximity buffers, z-score analysis, and spatial clustering algorithms.

D. Data Sources and Processing

We used Ireland's Record of Monuments and Places (RMP), which includes over 150,000 entries across spatial point data. Monument types lacking attribute diversity or unique identifiers were excluded. Ordnance Survey barony polygons served as proxies for Gaelic territorial boundaries at ($t_0 + 400$), 1600 CE.

- Geometry cleaning ensured topological validity and removal of overlaps.
- Monuments with rough temporal estimates were assigned values using typological proxies, then converted into scalar temporal coordinates relative to t_0.
- Feature columns were added: buffer distance to nearest border, soil type (from national soils datasets), and clustering indices.
- Contiguity matrices (Queen adjacency and distance-based) were constructed for evaluating inter-territorial connections.

This preprocessing phase ensured a spatially and temporally structured dataset suitable for signal analysis, pattern detection, and downstream ML applications.

1) Spatial Analysis Techniques: To test the framework without relying on typological classification or period labels, we analyzed signal distributions relative to reconstructed territorial boundaries from ($t_0 + 400$).

A buffer of 1,500 meters was applied around each monument to evaluate proximity to borders—representing transitional zones rather than fixed lines. A signal was considered to intersect a boundary if it fell within this buffer.

We calculated the proportion of signals near borders by type and assigned z-scores to assess statistical significance against random spatial distributions. These values were then joined to the main signal dataset.

Temporal estimates were approximated using existing typological labels as rough proxies and converted into scalar values in years before present (BP). A uniform uncertainty margin was applied to each signal to reflect dating imprecision.

To visualize long-term patterns, signal distributions were plotted across the full 6,000-year temporal range. Two plots produced are reproduced here:

- A bubble map highlighting signal concentrations exceeding one standard deviation from expected distributions.
- A contour map visualizing spatiotemporal clustering near boundaries, indicating persistent or changing activity zones over time.

This approach allowed us to model spatial relationships independently of typology and test whether archaeological signals align with known political geographies across millennia.

2) Territory validation: To test the relationship between spatial connectivity and administrative importance, we performed correlation and distributional analyses. Historical sources indicate that Irish territories were grouped into federations. Each federation had a central ruling dynasty and a chief king based in an administrative core territory [14].

To test whether such centres could be inferred from archaeological polygon data, we used a spatial weights contiguity matrix to model inter-territorial adjacency. Territories with six or more contiguous neighbors were classified as highly connected and evaluated as potential administrative centers—those that matched to administrative centres known from records were given a binary indicator [15].

Kernel Density Estimation (KDE) was used to model the intensity of monument distributions, while K-Nearest Neighbors (KNN) graphs helped examine proximity-based relationships and centrality metrics. [11]

3) Ringfort validation: A subset of signals categorised as ringforts was extracted and a set of features was added, including substrate types, diameters, border proximity z-scores and clustering measurements using PySAL. Global Moran's I and Local Indicators of Spatial Association (LISA) were used to detect clustering behaviors and identify statistically significant spatial patterns and anomalies.

A Random Forest classifier was trained to predict spatial clustering categories of ringforts, using diameter as the primary feature.

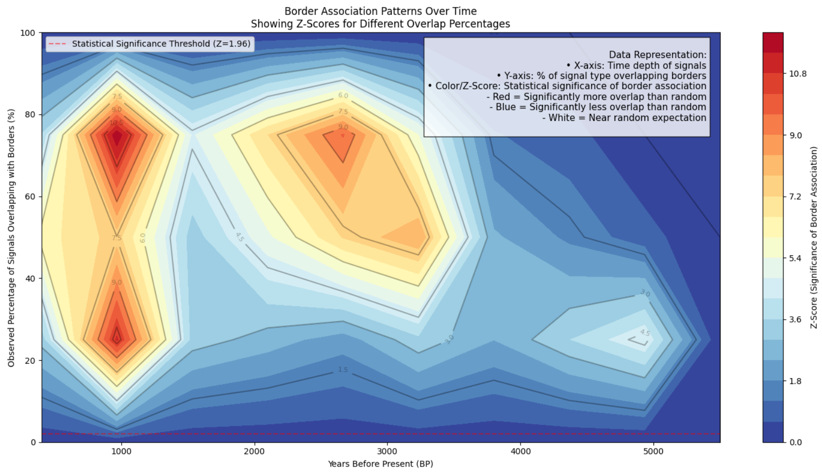

Fig. 3: Temporal Signal Contour: Activity intensity along territorial boundaries over 6000 years mapped within our 3D spatial manifold framework. The clear signal from $(t_0 + 2900)$ to $(t_0 + 1400)$ (900 BCE to 600 CE) coincides with Late Iron Age and Early Christian territorial changes. The analysis reveals deeper Bronze Age signals and Neolithic activity at $(t_0 + 5900)$ (3900 BCE).

RESULTS

Boundaries as Multi-Temporal Constructs

The temporal contour map (Figure 3) reveals signal clusters as early as $(t_0 + 5900)$ (3900 BCE), indicating long-term boundary persistence. The method identifies persistent high-significance bands during Bronze Age ($(t_0 + 3900)$ to $(t_0 + 3400)$) and Medieval periods ($(t_0 + 1400)$ to $(t_0 + 900)$), plus unexpected Neolithic proto-boundaries. Spatial autocorrelation tests confirm these patterns exceed random expectations (p ¡ 0.001).

The boundaries recorded at (t_0+400) (1600 CE) crystallised during the Late Iron Age to Early Medieval period, with Bronze Age reorganisation establishing enduring frameworks and Neolithic markers providing anchor points. The framework successfully detects invasive patterns—Norman motte castles show strong negative correlations with traditional boundaries, correctly identifying this known historical invasion through physical data alone.

Validation Through Territory Connectivity and Ringfort Analysis

Territories associated with power centers exhibited significantly higher connectivity (5.25 vs 2.94 neighboring territories, p = 0.003), with strong correlation between administrative importance and spatial connectivity (r = 0.685).

Random Forest classification (83–94% accuracy) revealed ringfort clustering behaviors encoding the same territorial logic detected temporally. Ringforts within 1.5 km of later boundaries were 22% larger (30.3 m vs 24.8 m diameter) and predominantly HH-type clusters (large-large), mirroring temporal contour high-Z boundary signals at $(t_0 + 1900)$ to (t_0+900) (100 CE–1100 CE). Regional patterns show Cavan's dense HH clusters (0.71/km^2) versus Donegal's LL dominance (small-small, 0.07/km^2), reflecting different federation histories (Table I).

TABLE I: Ringfort spatial signatures

Region	Cluster Type	Density (km^{-2})
Cavan	HH (Large-Large)	0.71
Donegal	LL (Small-Small)	0.07
Leitrim	Mixed	0.54

(a) Tower houses (red dots) $(t_0 + 600)$ to $(t_0 + 400)$

(b) Ringforts $(t_0 + 1600)$ to $(t_0 + 1100)$

Fig. 4: Signal persistence: later tower houses anchor boundaries established by earlier ringfort clusters, despite 700-year gap.

The ringfort size-clustering patterns spatially "freeze" the same territorial dynamics detected temporally, with HH clusters marking persistent boundaries and LL zones reflecting territorial interiors. This validates our signal model's capacity to decode multi-period territorial logic embedded within the spatial manifold framework, supporting the idea that territorial frameworks visible at $(t_0 + 400)$ have deeper origins.

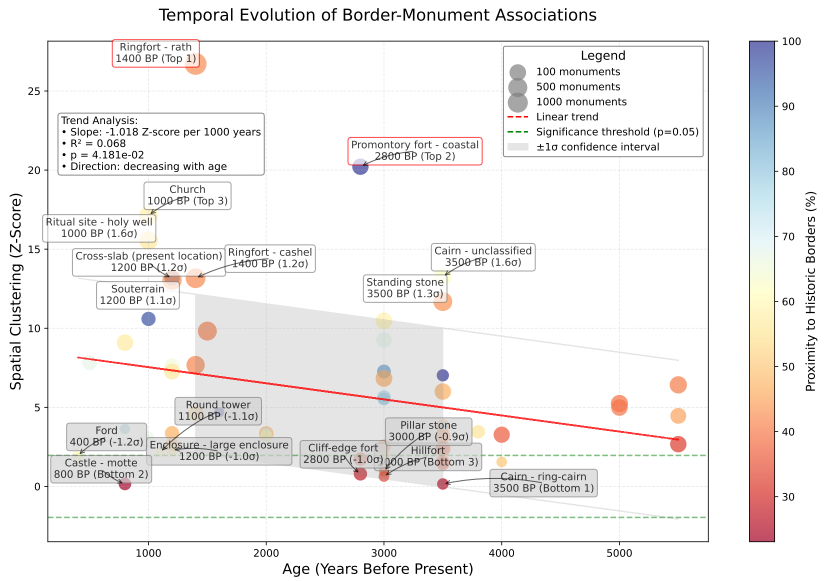

Fig. 5: Temporal evolution of signals showing the burst of activity between $(t_0 + 2900)$ and $(t_0 + 1400)$ (900 BCE-600 CE). Note the negative correlation of Norman mottes, enabling identification of invasive patterns while confirming authentic territorial signals.

IV. Discussion

Results align with known historical patterns, with independent GIS validation confirming inferred structures within our 3D spatial manifold framework. Despite limited data and poor temporal resolution, the approach rapidly identifies broad historical patterns while revealing both continuities and discontinuities—notably the complete absence of border associations with Norman mottes, correctly identifying invasive territorial patterns.

The framework enables three key capabilities: **anomaly detection** through negative correlations (Norman mottes and monastic round towers avoid boundaries, reflecting invasion patterns and functional monastery placement); **temporal resilience** where high-Z boundaries function like technosignatures, with Neolithic signals ($(t_0 + 5900)$, $Z = 5.11$) persisting across millennia, analogous to SETI's enduring technosignature hypothesis [16]; and **universality** through culture-neutral geometric description rather than typological categories.

This model enables archaeology to access signal science mathematics—describing sites using signal fields, noise kernels, and degradation transformations within the 3D manifold framework. The alignment of later tower houses with earlier ringfort clusters demonstrates inherited organisational logic, suggesting signal filtering could theoretically subtract later features to infer earlier boundary states through automated ML models.

Critically, each Gaelic federation exhibits distinct signal fingerprints, supporting bottom-up territorial evolution rather than imposed periodisation. Traditional "Early Medieval ringforts" and "Late Medieval tower houses" appear as different manifestations of persistent spatial logic encoded within the manifold structure. This paradigm provides a generalisable framework for detecting intentional spatial patterning across time—applicable to prehistoric landscapes, colonial cartographies, or planetary surfaces.

[0]Boundary proximity defined as within 1.5 km of territory edges. HH/LL clusters significant at p¡0.01 (Local Moran's I).

V. Limitations and Future Directions

Despite strong results, several limitations constrain this study's scope: **incomplete data coverage** from unavailable datasets (Northern Ireland records) creates edge effects, while inconsistent RMP survey records lack sensor data and standardised measurements; **temporal undersampling** where most signals lack dating, with typology-based $(t_0 + n)$ values introducing uncertainty; and **methodological inconsistencies** from varying survey methods and sparse paired observations between remote-sensed and excavated records.

Future work should integrate radiocarbon-dated sites whose standardised format aligns with our signal-based approach, addressing the core challenge of training algorithms without comparing observed signals $s_{\text{obs}}(\mathbf{x})$ to original forms. Archaeology can benefit from probabilistic dating using measurable signal properties (size, material, degradation) as temporal priors, alongside techniques like OLE (Optimal Linear Estimation) for estimating phenomenon occurrence from partial records [17].

Field practices should adapt by recording core attributes (position $\mathbf{x}$, dimensions, materials) and publishing raw signal observations alongside interpretations. The temporal-to-spatial embedding within the manifold framework promises ML separation of different signal components in complex sites, enabling reverse reconstruction—a form of "error correction" to reveal original signals at inception, greatly enhancing archaeological reconstruction through more complex and accurate inference of the past.

VI. Conclusion

Traditional typological and era-based data structures obstruct computational archaeological inference at scale. Our signal-based approach offers a scalable alternative grounded in physical principles, designed for inference and uncertainty quantification within a 3D spatial manifold framework where time becomes spatial distance along motion vectors.

This study demonstrates that significant archaeological features can be recovered independent of historical records, suggesting this approach can support extensive research while dramatically increasing data handling scale. The signal field paradigm offers three key advantages: explicit degradation modelling enables learning from partial sites; unified language integrates diverse data types from LiDAR to ancient texts; and automated pattern recognition through advanced ML techniques. Most importantly, it transforms archaeological data into formats interoperable with planetary and astronomical sciences.

Methodologically, this enables bottom-up categorisation where ML groups features by fundamental physical attributes rather than imposed typologies, naturally accommodating archaeological challenges of interference, masking, and high false positive rates. By creating structured, typology-neutral data that separates observations from inferences and models time as spatial vectors, we transform archaeological records into dynamic, queryable spatiotemporal fields compatible with modern signal processing and ML frameworks.

A. Abbreviations and Acronyms

GIS Geographic Information System
LiSA Local Indicators of Spatial Association
ML Machine Learning
KDE Kernel Density Estimation
KNN K-Nearest Neighbor
RMP Record of Monuments and Places (Ireland)

Acknowledgment

Editng and latex code generation: [1] OpenAI, ”ChatGPT (Mar 14 version) [Large language model],” 2023. [Online]. Available: https://chat.openai.com [2] DeepSeek, ”DeepSeek Chat (May 2024 version) [Large language model],” 2024. [Online]. Available: https://chat.deepseek.com

References

[1] D. Davis, “Theoretical Repositioning of Automated Remote Sensing Archaeology: Shifting from Features to Ephemeral Landscapes,” *Journal of Computer Applications in Archaeology*, vol. 4, no. 1, p. 94, May 2021. [Online]. Available: https://journal.caa-international.org/article/10.5334/jcaa.72/

[2] A. Ribeiro, “The attitude towards science in the changing panorama of archaeological theory,” *Revista de Antropologia*, vol. 61, pp. 353–371, 01 2021.

[3] D. S. Davis, “Defining what we study: The contribution of machine automation in archaeological research,” *Digital Applications in Archaeology and Cultural Heritage*, vol. 18, p. e00152, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2212054820300515

[4] K. W. Kintigh, J. H. Altschul, M. C. Beaudry, R. D. Drennan, A. P. Kinzig, T. A. Kohler, W. F. Limp, H. D. G. Maschner, W. K. Michener, T. R. Pauketat, P. Peregrine, J. A. Sabloff, T. J. Wilkinson, H. T. Wright, and M. A. Zeder, “Grand challenges for archaeology,” *Proceedings of the National Academy of Sciences*, vol. 111, no. 3, p. 879–880, Jan. 2014. [Online]. Available: http://dx.doi.org/10.1073/pnas.1324000111

[5] D. A. Vakoch, Ed., *Archaeology, Anthropology, and Interstellar Communication*. Washington, D.C.: NASA History Division, 2014. [Online]. Available: https://www.nasa.gov/connect/ebooks/archaeology-anthropology-and-interstellar-communication

[6] J. Calder, R. Coil, A. Melton, P. J. Olver, G. Tostevin, and K. Yezzi-Woodley, “Use and misuse of machine learning in anthropology,” 2022. [Online]. Available: https://arxiv.org/abs/2209.02811

[7] L. Anselma and P. Terenziani, “Preferences in Temporal Relational Databases,” *IEEE Access*, vol. 12, pp. 65 418–65 427, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/10521646/

[8] T. McLaughlin, N. Whitehouse, R. Schulting, M. McClatchie, and P. Barratt, “The changing face of neolithic and bronze age ireland: A big data approach to the settlement and burial records,” *Journal of World Prehistory*, Jun. 2016.

[9] K. Phelan and D. Riordan, “Detection of ringforts from aerial photography using machine learning,” in *2020 31st Irish Signals and Systems Conference (ISSC)*. Letterkenny, Ireland: IEEE, Jun. 2020, pp. 1–6. [Online]. Available: https://ieeexplore.ieee.org/document/9180159/

[10] Y. Qiang and N. Van De Weghe, “Re-Arranging Space, Time and Scales in GIS: Alternative Models for Multi-Scale Spatio-Temporal Modeling and Analyses,” *ISPRS International Journal of Geo-Information*, vol. 8, no. 2, p. 72, Feb. 2019. [Online]. Available: http://www.mdpi.com/2220-9964/8/2/72

[11] X. Liu, O. Kounadi, and R. Zurita-Milla, “Incorporating Spatial Autocorrelation in Machine Learning Models Using Spatial Lag and Eigenvector Spatial Filtering Features,” *ISPRS International Journal of Geo-Information*, vol. 11, no. 4, p. 242, Apr. 2022. [Online]. Available: https://www.mdpi.com/2220-9964/11/4/242

[12] S. M. Kazemi, R. Goel, S. Eghbali, J. Ramanan, J. Sahota, S. Thakur, S. Wu, C. Smyth, P. Poupart, and M. A. Brubaker, “Time2vec: Learning a vector representation of time,” *CoRR*, vol. abs/1907.05321, 2019. [Online]. Available: http://arxiv.org/abs/1907.05321

[13] M. Schneider, N. Greifzu, L. Wang *et al.*, “An end-to-end machine learning approach with explanation for time series with varying lengths,” *Neural Computing and Applications*, vol. 36, pp. 7491–7508, 2024. [Online]. Available: https://doi.org/10.1007/s00521-024-09473-9

[14] W. J. Smyth, “ISBN 9781859183977 - Map-making, Lndscapes And Memory: Colonial And Early Modern Ireland,” 2006. [Online]. Available: https://isbnsearch.org/isbn/9781859183977

[15] F. Menéndez-Marsh, M. Al-Rawi, J. Fonte, R. Dias, L. J. Gonçalves, L. G. Seco, J. Hipólito, J. P. Machado, J. Medina, J. Moreira, T. Do Pereiro, M. Vázquez, and A. Neves, “Geographic Information Systems in Archaeology: A Systematic Review,” *Journal of Computer Applications in Archaeology*, vol. 6, no. 1, May 2023. [Online]. Available: https://account.journal.caa-international.org/index.php/up-j-jcaa/article/view/104

[16] J. T. Wright, J. Haqq-Misra, A. Frank, R. Kopparapu, M. Lingam, and S. Z. Sheikh, “The case for technosignatures: Why they may be abundant, long-lived, highly detectable, and unambiguous,” *The Astrophysical Journal Letters*, vol. 927, no. 2, p. L30, mar 2022. [Online]. Available: http://dx.doi.org/10.3847/2041-8213/ac5824

[17] A. Key, D. Roberts, and I. Jarić, “Reconstructing the full temporal range of archaeological phenomena from sparse data,” *Journal of Archaeological Science*, vol. 135, p. 105479, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0305440321001497