Mapping the Dynamics of Atrial Fibrillation with Spatiotemporal Graph Neural Networks

Anonymous Author(s)
Affiliation
Address
email

Abstract

Catheter ablation of persistent Atrial Fibrillation (AF) consists of a one-size-fits-all treatment with limited success. This may be due to our inability to map AF electrical dynamics on the atrial surface with the limited resolution and coverage provided by sequential contact mapping catheters, preventing effective patient phenotyping for personalised, targeted ablation. In this proof-of-concept study, we introduce FIBMAP, a graph recurrent neural network model that reconstructs global AF dynamics from sparse measurements. Trained and validated on 51 non-contact whole atria recordings, FIBMAP reconstructs whole atria dynamics from 10% surface coverage, achieving a 210% lower mean absolute error and a 300% higher performance in recovering phase singularities compared to baseline methods. Clinical utility of FIBMAP is demonstrated on real-world contact mapping recordings, achieving reconstruction fidelity comparable to non-contact mapping. Integrating FIBMAP into clinical practice could enable personalised AF care and improve outcomes.

1 Introduction

Atrial Fibrillation (AF) is the most common cardiac arrhythmia, affecting 37.5 million people worldwide with a lifetime risk of 1 in 3 [1, 2]. The arrhythmia is a major cause of stroke, heart failure and death, with direct costs representing 2.4% of the United Kingdom's £182 billion healthcare budget [3]. Catheter ablation, which aims to restore normal rhythm through controlled destruction of cardiac tissue, is a common treatment approach for AF [4]. However, success rates remain low, particularly for persistent AF where five-year success rates are as low as 50% [5].

The major limitation underlying these poor outcomes is our inability to map AF effectively during invasive electrophysiology procedures, that is, to reconstruct AF electrical dynamics across the atrial surface. Sequential contact mapping, the clinical standard, involves placing a multipolar electrode catheter in contact with the atrial surface (covering less than 10% of the surface per placement) to record electrical signals as a function of time and position. While effective for organised arrhythmias, stitching these non-continuous recordings into coherent maps fails in AF due to chaotic, beat-to-beat variations in wavefront propagation [6]. This limits clinical practice to a one-size-fits-all treatment strategy for a highly heterogeneous disorder [7].

In this proof-of-concept study, we propose a novel solution termed *imputation mapping* that leverages graph deep learning [8, 9] to reconstruct global AF dynamics from the sparse measurements obtained during routine sequential contact mapping. Our solution, FIBMAP, as depicted in Figure 1, introduces a graph recurrent neural network (GRNN) that learns shared representations of AF dynamics across patients while incorporating patient-specific parameters that rapidly personalise the model within the timeframe of the clinical procedure. To this end, our approach aims to bridge the gap between the high-quality sparse measurements available clinically and the comprehensive mapping needed for effective treatment guidance.

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

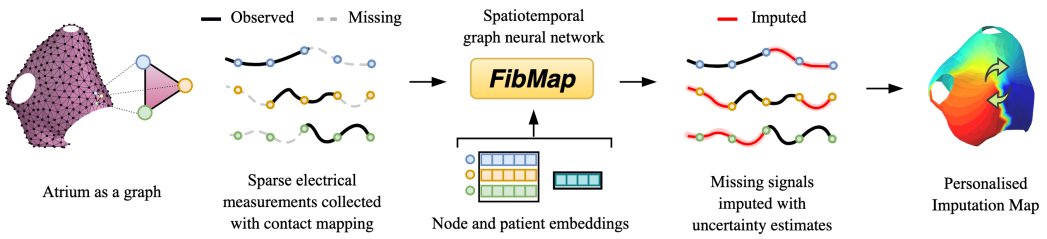

Figure 1: Conceptual overview of FIBMAP.

2 Methods

2.1 Problem formulation and graph representation

We model the atrial surface of patient p as a graph with adjacency matrix $\mathbf{A}^{(p)} \in \mathbb{R}^{N\times N}$, constructed by discretising patient-specific atrial geometry into N nodes connected through surface triangulation (see Figure 1), with adjacency remaining constant over time. Let $x_t^i \in \mathbb{R}$ denote the scalar electrical signal at node i and time t, with $\mathbf{X}_t \in \mathbb{R}^{N\times 1}$ representing the stacked signals across all N nodes and $\mathbf{X}_{t:T}$ indicating the sequence over time interval $[t, T]$. Due to sequential contact mapping, observations are sparse and characterised by a binary mask, $\mathbf{M}_t^{(p)} \in \{0,1\}^{N\times 1}$, indicating valid measurements. The complete graph sequence is represented as $\mathcal{G}_{t:T}^{(p)} = \langle \mathbf{X}_{t:T}^{(p)}, \mathbf{M}_{t:T}^{(p)}, \mathbf{A}^{(p)} \rangle$, a formulation that enables us to leverage recent advances in spatiotemporal graph imputation methods for our task (see Section A for a comprehensive review of related work).

Since clinical procedures measure only electrical signals while tissue properties influencing dynamics remain unmeasured, we introduce trainable embedding parameters to capture these unknown factors: node embeddings $\mathbf{V}^{(p)} \in \mathbb{R}^{N\times q}$ to learn spatially-varying tissue properties (e.g., conductivity, fibrosis) that influence dynamics [10], and patient embeddings $\mathbf{g}^{(p)} \in \mathbb{R}^r$ to capture global patient characteristics that affect overall AF behaviour (e.g., covariates); where q and r represent the node and patient embedding dimensions, respectively.

2.2 Architecture of FIBMAP

FIBMAP employs a bidirectional (Bi)-GRNN, that combines shared parameters, denoted by ϕ, capturing universal AF dynamics with patient-specific embeddings, $\mathbf{V}$ and $\mathbf{g}$, for personalisation. The architecture operates within an encode-process-decode framework, given by

$$\mathbf{H}_t^0 = \text{Enc}_\phi(\mathbf{X}_t, \mathbf{M}_t, \mathbf{V}, \mathbf{g}), \tag{1}$$

$$\mathbf{Z}_{t:T} = \text{Bi-GRNN}_\phi(\mathbf{H}_{t:T}^0, \mathbf{A}), \tag{2}$$

$$\widehat{\mathbf{Y}}_{t:T} = \text{Dec}_\phi(\mathbf{Z}_{t:T}, \mathbf{M}_{t:T}, \mathbf{V}, \mathbf{g}), \tag{3}$$

where the notation $^{(p)}$ has been relaxed. The encoder integrates observed signals, masks, and embeddings into unified node representations, which are then processed bi-directionally through the GRNN. The decoder combines these representations with patient-specific parameters to produce final spatiotemporal reconstructions $\widehat{\mathbf{Y}}_{t:T} \in \mathbb{R}^{N\times Q\times T}$. Specifically, we predict Q quantiles of the electrical signals to enable uncertainty quantification, and optimise the network through a quantile regression loss function [11]. See Sections B.1 and B.2 for more details on the architecture and loss function, respectively.

2.3 Model training and deployment via fine-tuning

The training strategy exploits the fundamental difference between model development and clinical deployment scenarios. During training, we assume dense/fully observed signals $\mathbf{X}_{t:T}$ and simulate sequential contact mapping with systematic data augmentation, varying catheter patch sizes (2.5-50% atrial coverage), spatial overlap between successive patches (0-80%), and dwell times (0.2-4 seconds) to reflect clinical variability (see Section C.2 and Figure 2A for more details). We assume a ground truth is available for training, and use a *whole atria reconstruction loss* to learn how to accurately map from sparse local observations to complete global dynamics. For clinical deployment where only partial observations are available, we fine-tune only the patient-specific embedding parameters, $\mathbf{V}$

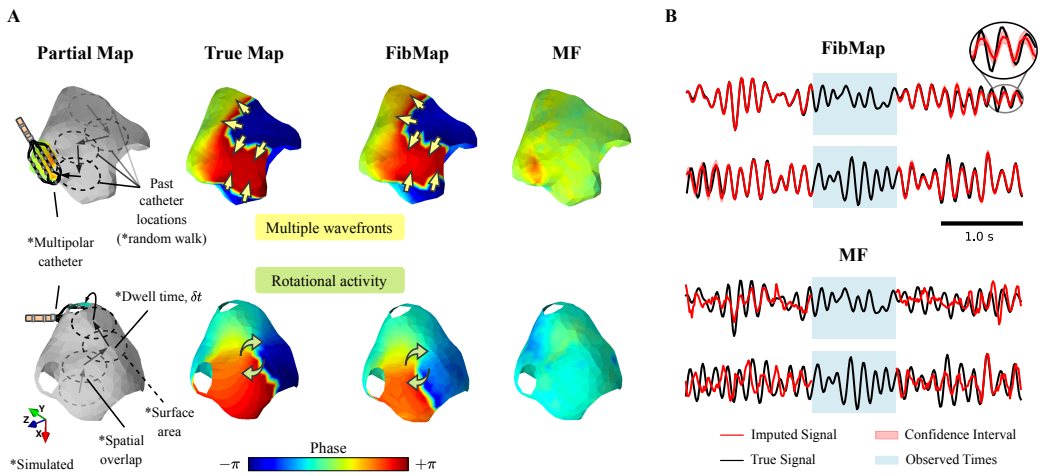

Figure 2: Qualitative results of FIBMAP imputation mapping. A) Phase map snapshots comparing FIBMAP and MF reconstructions against ground truth AcQMap for two patients (rows). FIBMAP reconstructs missing signals (grey regions) from 10% sparse observations using simulated sequential contact mapping (10% surface area, 1s dwell time, no overlap). B) Temporal signal reconstructions at individual nodes for both patients, showing FIBMAP's superior performance over MF in capturing complex AF dynamics.

and $\mathbf{g}$, using an *observed patch reconstruction loss*, while keeping all other architectural parameters, ϕ, fixed (see Figure 4 for the reconstruction loss formulations and Sections C.3 and C.4 for more training and fine-tuning details, respectively).

3 Results

3.1 Dataset and experimental setup

Our evaluation used recordings from 51 persistent AF patients captured with the AcQMap non-contact mapping system. Each AcQMap recording samples $\sim$3500 nodes across the atria at 3000 Hz for 5-20 seconds through inverse mapping of measurements from the atrial centre to the surface using non-contact electrodes [12, 13]. While AcQMap is not used clinically due to high cost, procedural complexity, and reduced spatial resolution inherent to inverse mapping [14, 15, 16], it offers realistic human AF data essential for this proof-of-concept study—unlike computational simulations which may inadequately represent human AF [17, 18]. The cohort (mean age 64$\pm$11 years, 69% male) had typical co-morbidities including hypertension (39%), heart failure (33%), and diabetes (10%); see Section C.1 and Table 2 for complete details. Signals were preprocessed as detailed in Section C.1, and sequential contact mapping was simulated from these whole-atria recordings as described in Section C.2, with patients stratified into train/val/test sets as described in Section C.1.

3.2 Validation of FIBMAP using AcQMap non-contact data

FIBMAP performance on new patients was quantified using our fine-tuning procedure on the test set. Testing involved simulating sequential contact mapping with a catheter surface area of 10% of the atrium, a dwell time of 1 second, and no spatial overlap between successive patches (as shown in Figure 2A). FIBMAP was fine-tuned as detailed in Section C.4, with the observed patch reconstruction loss monitored to perform early stopping. Fine-tuning took 3 hours 40 minutes total, averaging 22 minutes per patient. Performance was then measured against normalised ground truth signals using various metrics. FIBMAP outperformed baseline models in predicting whole atrium signals (Table 1). Specifically, FIBMAP achieved an mean absolute error (MAE) of 0.0574 ± 0.0005, a $2.1\times$ improvement compared to the best baseline imputation model, MF, which had an MAE of 0.1205 ± 0.0012. In mean squared error (MSE) and mean absolute percentage error (MAPE), FIBMAP scored 0.0069 ± 0.0001 and 16.77 ± 0.46, respectively, compared to MFs 0.0254 ± 0.0005 and 32.56 ± 0.56. Baseline models and performance metrics are detailed in Sections C.6 and C.5.

Table 1: Reconstruction performance across transductive, inductive and patient fine-tuned (FT) models.

	Model	MAE ↓	MSE ↓	MRE ↓	MAPE ↓	PS F1 ↑
TRA.	Mean	0.1734±0.0005	0.0507±0.0003	35.08±0.10	47.73±0.33	0.022±0.009
	MF	0.1205±0.0012	0.0254±0.0005	24.37±0.24	32.56±0.56	0.103±0.043
IND.	RNN	0.1393±0.0001	0.0298±0.0000	28.17±0.02	44.06±0.21	0.064±0.022
	Bi-RNN	0.1368±0.0001	0.0290±0.0001	27.66±0.03	42.86±0.19	0.113±0.045
	TTS-Transformer	0.1370±0.0001	0.0290±0.0001	27.71±0.03	43.03±0.20	0.053±0.026
FT	RNN + FT	0.1389±0.0002	0.0296±0.0001	28.09±0.04	43.57±0.51	0.148±0.047
	Bi-RNN + FT	0.1370±0.0000	0.0290±0.0001	27.70±0.02	42.92±0.60	0.183±0.048
	TTS-Transformer + FT	0.1362±0.0001	0.0288±0.0001	27.53±0.03	41.91±0.52	0.258±0.079
	FIBMAP	**0.0574**±0.0005	**0.0069**±0.0001	**11.59**±0.11	**16.77**±0.46	**0.780**±0.053

The clinical significance of these improvements becomes evident when examining the ability to predict phase singularities (PSs)—points where the Hilbert phase (instantaneous phase angle of electrical signals) is undefined—which represent the organising centres of spiral waves that may drive AF and constitute potential ablation targets [19]. FIBMAP achieved an F1 of 0.780 ± 0.053 for PS detection, representing a $3.0\times$ improvement over the next best method, the time-then-space (TTS)-Transformer + fine-tuning (FT) at 0.258 ± 0.079. This improvement in detecting critical AF features suggests that FIBMAP captures not merely signal amplitude but the underlying spatiotemporal dynamics essential for clinical interpretation. See Table 4 for F1-scores across varying tolerance thresholds for PS detection.

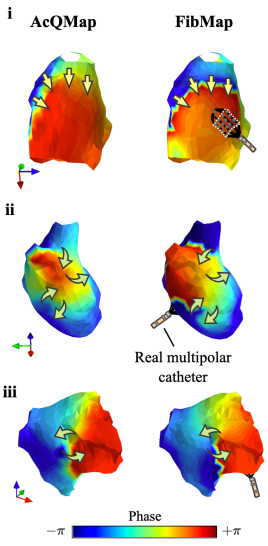

Figure 3: Snapshots of FibMap reconstructions from real HD-Grid recordings vs. AcQMap for three patients i-iii.

3.3 Transferring FibMap to real-world contact mapping

To demonstrate clinical utility, we applied FIBMAP to real-world HD Grid recordings from three patients who also had corresponding AcQMap data (collected non-simultaneously). Using a sliding window cross-correlation framework to handle temporal misalignment (Figure 5A; methods in Section C.7), we found that correlations between AcQMap and FIBMAP reconstructions were consistently higher when comparing the same patient (intra-patient, 99th percentiles: 0.19-0.22) than when comparing different patients (inter-patient: 0.16-0.17) or spatiotemporally shuffled controls ($\approx$0.02); see Figure 5B. Bootstrap confidence intervals confirmed statistical significance across window durations of 0.5-4.0 seconds, demonstrating patient-specific pattern capture (Figure 5C). As shown in Figure 3, FIBMAP successfully reconstructed aspects of AF dynamics including wavefronts and rotational activity from HD Grid data covering $<10\%$ of the atrial surface, revealing patterns invisible in the sparse original measurements.

4 Discussion and Conclusion

We introduce imputation mapping as a novel approach to reconstruct global AF dynamics from sparse clinical measurements, with FIBMAP achieving a $2.1\times$ improvement in reconstruction MAE and a $3.0\times$ improvement in PS detection over baselines when validated on a non-contact ground truth. The graph-based representation of the atrial surface proves crucial in modelling spatial dependencies between electrical signals, while the architectural design capturing shared aspects of AF dynamics across patients combined with personalised embedding components and rapid fine-tuning (22 minutes per patient) makes it feasible to obtain high-quality reconstructions within the duration of the invasive clinical procedure. While early results of imputation mapping on real-world HD Grid recordings display promise through demonstrated patient-specific pattern capture, future work must validate FIBMAP at higher spatial resolutions essential for informing targeted ablation strategies and develop more scalable computational architectures to maintain real-time performance at these resolutions. By bridging the gap between sparse clinical measurements and comprehensive dynamics reconstruction needed for effective treatment guidance, imputation mapping could offer a pathway towards the personalised AF treatment.

References

[1] Giuseppe Lippi, Fabian Sanchis-Gomar, and Gianfranco Cervellin. "Global epidemiology of atrial fibrillation: An increasing epidemic and public health challenge". In: *International Journal of Stroke* 16.2 (2021), pp. 217–221.

[2] Liping Mou et al. "Lifetime risk of atrial fibrillation by race and socioeconomic status: ARIC study (Atherosclerosis Risk in Communities)". In: *Circulation: Arrhythmia and Electrophysiology* 11.7 (2018), e006350.

[3] S Stewart, N Murphy, A Walker, A McGuire, and J J V McMurray. "Cost of an emerging epidemic: An economic analysis of atrial fibrillation in the UK". In: *Heart* 90.3 (2004), pp. 286–292.

[4] Karl-Heinz Kuck et al. "Cryoballoon or radiofrequency ablation for paroxysmal atrial fibrillation". In: *New England Journal of Medicine* 374.23 (2016), pp. 2235–2245.

[5] Daniel Scherr et al. "Five-year outcome of catheter ablation of persistent atrial fibrillation using termination of atrial fibrillation as a procedural endpoint". In: *Circulation: Arrhythmia and Electrophysiology* 8.1 (2015), pp. 18–24.

[6] Ozan Özgül et al. "High-density and high coverage composite mapping of repetitive atrial activation patterns". In: *Computers in Biology and Medicine* 159 (2023), p. 106920.

[7] Fu Siong Ng, Balvinder S. Handa, Xinyang Li, and Nicholas S. Peters. "Toward mechanism-directed electrophenotype-based treatments for atrial fibrillation". In: *Frontiers in Physiology* 11 (2020).

[8] Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. "The graph neural network model". In: *IEEE Transactions on Neural Networks* 20.1 (2008), pp. 61–80.

[9] Michael M. Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. "Geometric deep learning: Going beyond Euclidean data". In: *IEEE Signal Processing Magazine* 34.4 (2017), pp. 18–42.

[10] Andrea Cini, Ivan Marisca, Daniele Zambon, and Cesare Alippi. "Taming local effects in graph-based spatiotemporal forecasting". In: *Proceedings of the 37th International Conference on Neural Information Processing Systems* 36 (2024).

[11] Roger Koenker and Kevin F Hallock. "Quantile regression". In: *Journal of Economic Perspectives* 15.4 (2001), pp. 143–156.

[12] Natasja MS De Groot et al. "Critical appraisal of technologies to assess electrical activity during atrial fibrillation: A position paper from the European heart rhythm association and European society of cardiology working group on eCardiology in collaboration with the heart rhythm society, Asia pacific heart rhythm society, Latin American heart rhythm society and computing in cardiology". In: *EP Europace* 24.2 (2022), pp. 313–330.

[13] Rui Shi et al. "Validation of dipole density mapping during atrial fibrillation and sinus rhythm in human left atrium". In: *Clinical Electrophysiology* 6.2 (2020), pp. 171–181.

[14] Sanjiv M. Narayan et al. "Treatment of atrial fibrillation by the ablation of localized sources: CONFIRM (Conventional Ablation for Atrial Fibrillation With or Without Focal Impulse and Rotor Modulation) trial". In: *Journal of the American College of Cardiology* 60.7 (2012), pp. 628–636.

[15] Eric Buch et al. "Long-term clinical outcomes of focal impulse and rotor modulation for treatment of atrial fibrillation: A multicenter experience". In: *Heart Rhythm* 13.3 (2016), pp. 636–641.

[16] Caroline H. Roney et al. "Spatial resolution requirements for accurate identification of drivers of atrial fibrillation". In: *Circulation: Arrhythmia and Electrophysiology* 10.5 (2017), e004899.

[17] Jordi Heijman, Henry Sutanto, Harry J G M Crijns, Stanley Nattel, and Natalia A Trayanova. "Computational models of atrial fibrillation: achievements, challenges, and perspectives for improving clinical care". In: *Cardiovascular Research* 117.7 (Apr. 2021), pp. 1682–1699. ISSN: 0008-6363. DOI: 10.1093/cvr/cvab138.

[18] Byounghyun Lim et al. "In situ procedure for high-efficiency computational modeling of atrial fibrillation reflecting personal anatomy, fiber orientation, fibrosis, and electrophysiology". In: *Scientific Reports* 10.1 (Feb. 2020). ISSN: 2045-2322. DOI: 10.1038/s41598-020-59372-x.

[19] Xinyang Li et al. "Standardised framework for quantitative analysis of fibrillation dynamics". In: *Scientific Reports* 9.1 (2019), p. 16671.

[20] Antonio Ortega, Pascal Frossard, Jelena Kovaevi, José M. F. Moura, and Pierre Vandergheynst. "Graph signal processing: Overview, challenges, and applications". In: *Proceedings of the IEEE* 106.5 (2018), pp. 808–828.

[21] Ljubia Stankovi et al. "Data analytics on graphs. Part II: Signals on graphs". In: *Foundations and Trendső in Machine Learning* 13.2-3 (2020), pp. 158–331. ISSN: 1935-8237.

[22] Ljubia Stankovi et al. "Data analytics on graphs. Part III: Machine learning on graphs, from graph topology to applications". In: *Foundations and Trendső in Machine Learning* 13.4 (2020), pp. 332–530. ISSN: 1935-8237.

[23] Thomas N. Kipf and Max Welling. “Semi-supervised classification with graph convolutional networks”. In: *Proceedings of the International Conference on Learning Representations* (2017).

[24] Ziwei Zhang, Peng Cui, and Wenwu Zhu. “Deep learning on graphs: A survey”. In: *IEEE Transactions on Knowledge and Data Engineering* 34.1 (2020), pp. 249–270.

[25] Andrea Cini, Ivan Marisca, Daniele Zambon, and Cesare Alippi. “Graph deep learning for time series forecasting”. In: *arXiv preprint arXiv:2310.15978* (2023).

[26] Ming Jin et al. “A survey on graph neural networks for time series: Forecasting, classification, imputation, and anomaly detection”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2024).

[27] Jianfei Gao and Bruno Ribeiro. “On the equivalence between temporal and static equivariant graph representations”. In: *Proceedings of the 39th International Conference on Machine Learning* (2022), pp. 7052–7076.

[28] Andrea Cini, Ivan Marisca, Filippo Maria Bianchi, and Cesare Alippi. “Scalable spatiotemporal graph neural networks”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 37. 2023. Chap. 6, pp. 7218–7226.

[29] Youngjoo Seo, Michaël Defferrard, Pierre Vandergheynst, and Xavier Bresson. “Structured sequence modeling with graph convolutional recurrent networks”. In: *Proceedings of the 25th International Conference on Neural Information Processing*. Springer, 2018, pp. 362–373.

[30] Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. “Diffusion convolutional recurrent neural network: Data-driven traffic forecasting”. In: *Proceedings of the International Conference on Learning Representations* (2018).

[31] Ivan Marisca, Andrea Cini, and Cesare Alippi. “Learning to reconstruct missing data from spatiotemporal graphs with sparse observations”. In: *Proceedings of the 36th International Conference on Neural Information Processing Systems* (2022), pp. 32069–32082.

[32] Weiwei Jiang and Jiayun Luo. “Graph neural network for traffic forecasting: A survey”. In: *Expert Systems with Applications* 207 (2022), p. 117921.

[33] Remi Lam et al. “Learning skillful medium-range global weather forecasting”. In: *Science* 382.6677 (2023), pp. 1416–1421.

[34] Andrea Cini, Danilo Mandic, and Cesare Alippi. “Graph-based time series clustering for end-to-end hierarchical forecasting”. In: *Proceedings of the 41st International Conference on Machine Learning* (2024).

[35] Andrea Cini, Ivan Marisca, and Cesare Alippi. “Filling the g_ap_s: Multivariate time series imputation by graph neural networks”. In: *Proceedings of the International Conference on Learning Representations* (2022).

[36] Xueyan Yin, Feifan Li, Yanming Shen, Heng Qi, and Baocai Yin. “NodeTrans: A graph transfer learning approach for traffic prediction”. In: *arXiv preprint arXiv:2207.01301* (2022).

[37] Ruslan Salakhutdinov and Andriy Mnih. “Bayesian probabilistic matrix factorization using Markov chain Monte Carlo”. In: *Proceedings of the 25th International Conference on Machine Learning*. 2008, pp. 880–887.

[38] Daniel Lee and H Sebastian Seung. “Algorithms for non-negative matrix factorization”. In: *Proceedings of the 13th International Conference on Neural Information Processing Systems* 13 (2000).

[39] Deng Cai, Xiaofei He, Jiawei Han, and Thomas S Huang. “Graph regularized nonnegative matrix factorization for data representation”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33.8 (2010), pp. 1548–1560.

[40] Hsiang-Fu Yu, Nikhil Rao, and Inderjit S Dhillon. “Temporal regularized matrix factorization for high-dimensional time series prediction”. In: *Proceedings of the 30th International Conference on Neural Information Processing Systems* 29 (2016).

[41] Wei Cao et al. “BRITS: Bidirectional recurrent imputation for time series”. In: *Proceedings of the 32nd International Conference on Neural Information Processing Systems* 31 (2018).

[42] Jinsung Yoon, Daniel Jarrett, and Mihaela Van der Schaar. “Time-series generative adversarial networks”. In: *Proceedings of the 33rd International Conference on Neural Information Processing Systems* 32 (2019).

[43] Yukai Liu, Rose Yu, Stephan Zheng, Eric Zhan, and Yisong Yue. “NAOMI: Non-autoregressive multiresolution sequence imputation”. In: *Proceedings of the 33rd International Conference on Neural Information Processing Systems* 32 (2019).

[44] Wenjie Du, David Côté, and Yan Liu. “SAITS: Self-attention-based imputation for time series”. In: *Expert Systems with Applications* 219 (2023), p. 119619.

[45] Yusuke Tashiro, Jiaming Song, Yang Song, and Stefano Ermon. “CSDI: Conditional score-based diffusion models for probabilistic time series imputation”. In: *Proceedings of the 35th International Conference on Neural Information Processing Systems* 34 (2021), pp. 24804–24816.

[46] J Alcaraz and N Strodthoff. “Diffusion-based time series imputation and forecasting with structured state space models”. In: *Transactions on Machine Learning Research* (2023).

[47] Alexander Jenkins, Zehua Chen, Fu Siong Ng, and Danilo Mandic. "Improving diffusion models for ECG imputation with an augmented template prior". In: *arXiv preprint arXiv:2310.15742* (2023).

[48] Pablo Montero-Manso and Rob J Hyndman. "Principles and algorithms for forecasting groups of time series: Locality and globality". In: *International Journal of Forecasting* 37.4 (2021), pp. 1632–1653.

[49] Xavier Bresson and Thomas Laurent. "Residual gated graph convnets". In: *arXiv preprint arXiv:1711.07553* (2017).

[50] James Atwood and Don Towsley. "Diffusion-convolutional neural networks". In: *Proceedings of the 30th International Conference on Neural Information Processing Systems* 29 (2016).

[51] Mikhail Belkin and Partha Niyogi. "Laplacian eigenmaps for dimensionality reduction and data representation". In: *Neural Computation* 15.6 (2003), pp. 1373–1396.

[52] L. Wieser et al. "Detection of Phase Singularities in Triangular Meshes". In: *Methods of Information in Medicine* 46.06 (2007), pp. 646–654. ISSN: 2511-705X. DOI: 10.3414/me0427.

[53] Ashish Vaswani et al. "Attention is all you need". In: *Advances in Neural Information Processing Systems*. 2017, pp. 5998–6008.

[54] Paul J Besl and Neil D McKay. "Method for registration of 3-D shapes". In: *Proceedings of Sensor fusion IV: Control Paradigms and Data Structures* 1611 (1992), pp. 586–606.

A Background and related work

Building on graph signal processing methods [20, 21], graph deep learning [9, 22] has generalised successful deep learning architectures, such as convolutional neural networks, to the irregular graph domain [23, 24]. Spatio-temporal graph neural networks (ST-GNNs) refer to the class of deep learning architectures designed to analyse time-varying graph signals [25, 26]. ST-GNNs can be categorised into time-then-space [27, 28] or time-and-space models [29, 30, 31], which denote separate or joint processing of the space and time dimensions, respectively. A notable example of a time-and-space ST-GNN is the GRNN, which replaces the fully connected layers in a recurrent neural network (RNN) with graph convolutions [29, 30]. ST-GNNs have achieved state-of-the-art performance in tasks such as forecasting [32, 33, 34] and missing data imputation [31, 35].

ST-GNNs are inherently global models, sharing parameters across space and time. Global ST-GNNs can be used for zero-shot transfer and inductive learning on unseen graphs, however, they might fail to account for spatial variations in dynamics. Node embeddings have been recently introduced to learn these local effects in ST-GNNs [10]. Whilst hybrid global-local models often outperform global architectures, recent research has focussed on improving their performance in an inductive setting using transfer learning [10, 36].

Other imputation methods have been applied to time series, ranging from MF methods [37, 38] and their counterparts with graph and temporal regularisation [39, 40], to deep learning techniques employing RNNs [41], generative adversarial networks [42, 43], transformers [44], or most recently denoising diffusion probabilistic models [45, 46, 47].

B FIBMAP

B.1 Architecture

FIBMAP is instantiated as a Bi-GRNN, a non-linear state-space model designed to propagate information from valid observations across space and time. Our solution is composed of two spatiotemporal encoder blocks and a decoder. The spatiotemporal encoders operate in two different directions, processing the sequence in both a forward (*fwd*) and backward (*bwd*) direction, respectively. FIBMAP's architecture extends the framework proposed in [35] by introducing local (node) and global (patient) embedding parameters to address the AF mapping problem. These modifications enable generalisation across patients and efficient transfer to new patients via fine-tuning.

To balance the trade-off between personalised reconstruction and generalisation across a spectrum of AF dynamics with a single model, FIBMAP is designed to perform personalised reconstruction by providing patient-specific parameters, $\mathbf{V}^{(p)}$ and $\mathbf{g}^{(p)}$, as additional input to the encoder and decoder components of the architecture. All other model parameters are shared across patients to learn the common aspects of the dynamics (such as the physics of the problem). We begin by defining the spatiotemporal encoder block of the architecture, before explaining the decoder.

B.1.1 The spatiotemporal encoder

First, the observed data at time t is encoded as

$$\mathbf{H}_t^0 = \text{MLP}_{\text{enc}}([\mathbf{X}_t^{(p)} \odot \mathbf{M}_t^{(p)} \| \mathbf{M}_t^{(p)} \| \mathbf{G}^{(p)} \| \mathbf{V}^{(p)}]), \tag{4}$$

where the symbols $\odot$ and $\|$ denote the Hadamard product and concatenation operator, respectively, and $\mathbf{G}^{(p)} \in \mathbb{R}^{N \times r}$ is a matrix where copies of $\mathbf{g}^{(p)}$ are stacked across nodes. The encoding step constructs a representation of the observed and missing values alongside the patient-specific parameters in a common embedding space, $\mathbf{H}_t^0 \in \mathbb{R}^{N \times d}$. Next, the embedded data is processed sequentially using a gated-GRNN, where the k-th layer is given by

$$\mathbf{Z}_t^k = \mathbf{H}_{t-1}^k, \tag{5a}$$
$$\mathbf{R}_t^k = \sigma(\text{MP}_r^k([\mathbf{Z}_t^k \| \mathbf{H}_{t-1}^k], \mathcal{E})), \tag{5b}$$
$$\mathbf{U}_t^k = \sigma(\text{MP}_u^k([\mathbf{Z}_t^k \| \mathbf{H}_{t-1}^k], \mathcal{E})), \tag{5c}$$
$$\mathbf{C}_t^k = \tanh(\text{MP}_c^k([\mathbf{Z}_t^k \| \mathbf{R}_t^k \odot \mathbf{H}_{t-1}^k], \mathcal{E})), \tag{5d}$$
$$\mathbf{H}_t^k = \mathbf{U}_t^k \odot \mathbf{H}_{t-1}^k + (1 - \mathbf{U}_t^k) \odot \mathbf{C}_t^k. \tag{5e}$$

$\text{MP}_r^k(\cdot)$, $\text{MP}_u^k(\cdot)$ and $\text{MP}_c^k(\cdot)$ denote the message passing neural network (MPNN) layers for the reset, update, and candidate gates, respectively, and the activation functions $\sigma(\cdot)$ and $\tanh(\cdot)$ denote the sigmoid and hyperbolic tangents. The hidden state of the gated-GRNN at each layer k is initialised as a linear function of the node embeddings

$$\mathbf{H}_{t=0}^k = \mathbf{V}^{(p)} \mathbf{W}_{\text{init}}^k + \mathbf{b}_{\text{init}}^k, \tag{6}$$

where $\mathbf{W}_{\text{init}}^k \in \mathbb{R}^{q \times d}$ and $\mathbf{b}_{\text{init}}^k \in \mathbb{R}^d$, which takes the different characteristics of each time series into account when initialising the state, thus reducing the need to rely on a long observation history [48].

Message passing layers The MPNN layers for each gate in the GRNN, denoted as MP, compute the hidden state of the i-th node as

$$\mathbf{h}_{t,i}^{k+1} = \gamma^k \left(\mathbf{h}_{t,i}^k, \sum_{j \in \mathcal{N}(i)} \rho^k(\mathbf{o}_i^k, \mathbf{o}_j^k, e_{ij}) \right), \tag{7}$$

where $\mathbf{o}_i^k \in \mathbb{R}^{2d}$ represents the input of the gate at layer k, $\mathcal{N}(i)$ denotes the set of nodes connected to i, $\rho^k(\cdot)$ is a message function, $\gamma^k(\cdot)$ is an update function, and the summation serves as a permutation invariant aggregation over neighbouring nodes. Specifically, the message function $\rho^k(\cdot)$ is implemented as

$$\mathbf{m}_{ij}^k = \text{MLP}_{\text{msg}}^k([\mathbf{o}_i^k \| \mathbf{o}_j^k]), \tag{8a}$$
$$\alpha_{ij}^k = \sigma(\mathbf{W}_{\text{msg}-\alpha}^k \mathbf{m}_{ij}^k), \tag{8b}$$
$$\mathbf{m}_{ij}^k = \alpha_{ij}^k \mathbf{m}_{ij}^k, \tag{8c}$$

where the messages along non-zero edges e_{ij} are weighted according to an inferred scalar value α_{ij}^k, which resides on the interval [0, 1]. The scalar value α_{ij}^k allows for anisotropic message passing, which assists in learning latent edge attributes such as the coupling strength between nodes. Messages are then aggregated at node i using the summation, $\mathbf{m}_i^k = \sum_{j \in \mathcal{N}(i)} \mathbf{m}_{ij}^k$. Finally, the update function $\gamma^k(\cdot)$ is implemented with a residual skip connection as

$$\mathbf{u}_i^k = \text{MLP}_{\text{update}}^k([\mathbf{h}_i^k \| \mathbf{m}_i^k]), \tag{9a}$$
$$\mathbf{h}_{t,i}^{k+1} = \mathbf{u}_i^k + \mathbf{W}_{\text{skip}}^k \mathbf{o}_i^k, \tag{9b}$$

where $\mathbf{W}_{\text{skip}} \in \mathbb{R}^{d \times 2d}$ and $\mathbf{h}_i^{k+1} \in \mathbb{R}^d$. Note that the parameters of the MPNN layers, MP, are defined separately for each gate and layer. This message-passing mechanism loosely resembles the gated graph convolution introduced in [49].

Iterative imputations To learn effective state space representations while processing the data sequentially, missing values must be accounted for. To do this, first and second-stage imputations as in [35] are performed iteratively within the spatiotemporal encoder block. The first stage imputation performs one-step-ahead prediction via a linear readout as

$$\widehat{\mathbf{X}}_{t+1}^{(1)} = \mathbf{H}_t^K \mathbf{W}_{\text{stage-1}} + \mathbf{b}_{\text{stage-1}}, \tag{10}$$

which is used in place for the missing values. The second stage imputation acts as a type of regularisation, estimating the value at node i using the previous hidden states, as well as the masks and first-stage imputations at the neighbouring nodes. The second stage imputation involves first computing an intermediate representation of each node, given by

$$\mathbf{s}_{t+1,i} = \gamma_{\text{stage-2}} \left(\mathbf{h}_{t,i}^K, \sum_{j \in \mathcal{N}(i)} \rho_{\text{stage-2}}([\hat{\mathbf{x}}_{t+1,j}^{(1)} \| \mathbf{h}_{t,j}^K \| \mathbf{m}_{t+1,j}], e_{ij}) \right), \tag{11}$$

where $\mathbf{s}_{t+1,i} \in \mathbb{R}^d$ and the message passing is implemented using a diffusion graph convolution [50]. Next, a linear readout layer performs the second stage imputation as

$$\widehat{\mathbf{X}}_{t+1}^{(2)} = [\mathbf{S}_{t+1} \| \mathbf{H}_t^K] \mathbf{W}_{\text{stage-2}} + \mathbf{b}_{\text{stage-2}}, \tag{12}$$

which is used again in place of the missing values in $\mathbf{X}_{t+1}$.

The process detailed so far in the spatiotemporal encoding block is repeated for $\mathbf{X}_{t+1}$, to learn representations $\mathbf{S}_{t+2}$ and $\mathbf{H}_{t+1}^K$, and so forth, until a set of representations $\mathbf{S}_{t:t+W}$ and $\mathbf{H}_{t:t+W}^K$, and imputations $\widehat{\mathbf{X}}_{t:t+W}^{(1)}$ and $\widehat{\mathbf{X}}_{t:t+W}^{(2)}$ are calculated for the whole window length W. To summarise the spatiotemporal encoding block detailed in this section, we use the following shorthand notation

$$\langle \mathbf{S}_{t:t+W}, \mathbf{H}_{t:t+W}^K, \widehat{\mathbf{X}}_{t:t+W}^{(1)}, \widehat{\mathbf{X}}_{t:t+W}^{(2)} \rangle = \text{ST-Encoder}(\mathbf{X}_{t:t+W}^{(p)}, \mathbf{M}_{t:t+W}^{(p)}, \mathbf{g}^{(p)}, \mathbf{V}^{(p)}). \tag{13}$$

B.1.2 The decoder

The sequences are encoded in both the forward and backward directions, to form

$$\langle \mathbf{S}_{\text{fwd}}, \mathbf{H}_{\text{fwd}}^K, \widehat{\mathbf{X}}_{\text{fwd}}^{(1)}, \widehat{\mathbf{X}}_{\text{fwd}}^{(2)} \rangle = \text{ST-Encoder}(\mathbf{X}_{t:t+W}^{(p)}, \mathbf{M}_{t:t+W}^{(p)}, \mathbf{g}^{(p)}, \mathbf{V}^{(p)}), \tag{14}$$

and

$$\langle \mathbf{S}_{\text{bwd}}, \mathbf{H}_{\text{bwd}}^K, \widehat{\mathbf{X}}_{\text{bwd}}^{(1)}, \widehat{\mathbf{X}}_{\text{bwd}}^{(2)} \rangle = \text{ST-Encoder}(\mathbf{X}_{t+W:t}^{(p)}, \mathbf{M}_{t+W:t}^{(p)}, \mathbf{g}^{(p)}, \mathbf{V}^{(p)}), \tag{15}$$

respectively, where $t + W : t$ denotes the time reversed/backward sequence. The next stage is to decode the hidden state representations from the encodings in both directions, to perform a final imputation at time t'. This is done with the following function

$$\widehat{\mathbf{Y}}_{t'} = \text{MLP}_{\text{dec}}([\mathbf{S}_{\text{fwd} \leq t'} \| \mathbf{H}_{\text{fwd} \leq t'}^K \| \mathbf{S}_{\text{bwd} \geq t'} \| \mathbf{H}_{\text{bwd} \geq t'}^K \| \mathbf{M}_{t'}^p \| \mathbf{V}^{(p)} \| \mathbf{g}^{(p)}]), \tag{16}$$

where the notation fwd $\leq t'$ refers to an index of the hidden states at time t' (which includes information from times $\leq t'$) and similarly, bwd $\geq t'$ refers to index at time t' (which includes information from times $\geq t'$). The decoder takes a form similar to that in [35] except patient-specific parameters are also provided to facilitate personalisation. We denote the final imputed values for the entire window length as $\widehat{\mathbf{Y}}_{t:t+W}$.

B.2 Optimisation

To quantify the uncertainty of the reconstructed dynamics, FIBMAP is formulated as a quantile regression [11]. In general, a quantile regression aims to estimate the conditional quantile function, $Q_Y(\tau|\mathbf{B})$, which represents the τ-th quantile of the response variable, Y, given the predictor variables, $\mathbf{B}$. This is given by $Q_Y(\tau|\mathbf{B}) = \inf\{y \in \mathbb{R} | P(Y \leq y|\mathbf{B}) \geq \tau\}$, where $P(Y \leq y|\mathbf{B})$ is the conditional cumulative distribution function of the response variable, Y, given the predictors $\mathbf{B}$, and $\inf$ represents the infimum of the set.

In this work, conditional quantile functions for $\tau \in \mathcal{C}$, where $\mathcal{C} = \{\tau_1, \tau_2, \ldots, \tau_{|\mathcal{C}|}\}$, are predicted for each of the imputed values. This is done by computing the following masked quantile/pinball loss function

$$\mathcal{L}(\widehat{\mathbf{Y}}_{t:t+T}, \widetilde{\mathbf{Y}}_{t:t+T}, \bar{\mathbf{M}}_{t:t+T}) = \frac{\sum_{h=t}^{t+T} \sum_{i=1}^{N} \bar{m}_{t,i} \mathcal{L}_{t,i}(\hat{\mathbf{y}}_{t,i}, y_{t,i})}{\sum_{h=t}^{t+T} \sum_{i=1}^{N} \bar{m}_{t,i}}, \tag{17}$$

Table 2: Patient demographics and clinical characteristics. Values are mean ± SD or n (%). The asterisk denotes the presence of missing values; the number missing is shown as n^*.

	Total ($N = 51$)
Demographics	
Mean age (years)	64 ± 11
Male sex	35 (69%)
Mean BMI	29 ± 5
Comorbidities	
Hypertension	20 (39%)
Diabetes mellitus	5 (10%)
Heart failure	17 (33%)
CHA_2DS_2-VASc score >2	14 (27%)
Medications	
Beta-blockers	33^* (87%, $n^* = 13$)
Amiodarone	11^* (29%, $n^* = 13$)
Anticoagulants	38^* (100%, $n^* = 13$)
Statins	16^* (42%, $n^* = 13$)

where $\widehat{\mathbf{Y}}_{t:t+T} \in \mathbb{R}^{N\times|\mathcal{C}|}$, $\mathbf{Y}_t \in \mathbb{R}^{N\times 1}$, and $\bar{\mathbf{M}}_t \in \{0,1\}^{N\times 1}$, represent the predicted values, target values, and the evaluation mask, respectively, at time t. The function $\mathcal{L}_{t,i}$ computes the average pinball loss computed at time t and node i, which is given by

$$\mathcal{L}_{t,i}(\hat{\mathbf{y}}_{t,i}, y_{t,i}) = \frac{1}{|\mathcal{C}|}\sum_{c=1}^{|\mathcal{C}|}(y_{t,i} - \hat{\mathbf{y}}_{t,i}[c])(\tau_c - 1\{y_{t,i} - \hat{\mathbf{y}}_{t,i}[c] < 0\}), \tag{18}$$

where c refers to the channel of $\hat{\mathbf{y}}_{t,i}$ (prediction vector at time t and node i) which contains the prediction of the c-th conditional quantile, $Q_y(\tau_c|\ldots)$ at time t and node i, and $1\{y_{t,i} - \hat{\mathbf{y}}_{t,i}[c] < 0\}$ represents the indicator function, which is equal to 1 if $y_{t,i} - \hat{\mathbf{y}}_{t,i}[c] < 0$, else it is equal to 0. In practice, we modify the decoder to predict a value for each quantile (rather than a single value) and compute the average pinball loss from each predicted quantile value as in (18).

For FIBMAP, the following loss function is minimised

$$\mathcal{L}_{\text{loss}} = \mathcal{L}(\widehat{\mathbf{Y}}_{t:t+T}, \mathbf{X}_{t:t+T}, \bar{\mathbf{M}}_{t:t+T}) \tag{19}$$

$$+ \mathcal{L}(\widehat{\mathbf{X}}^{(1),\text{fwd}}_{t:t+T}, \mathbf{X}_{t:t+T}, \bar{\mathbf{M}}_{t:t+T}) + \mathcal{L}(\widehat{\mathbf{X}}^{(2),\text{fwd}}_{t:t+T}, \mathbf{X}_{t:t+T}, \bar{\mathbf{M}}_{t:t+T}) \tag{20}$$

$$+ \mathcal{L}(\widehat{\mathbf{X}}^{(1),\text{bwd}}_{t:t+T}, \mathbf{X}_{t:t+T}, \bar{\mathbf{M}}_{t:t+T}) + \mathcal{L}(\widehat{\mathbf{X}}^{(1),\text{bwd}}_{t:t+T}, \mathbf{X}_{t:t+T}, \bar{\mathbf{M}}_{t:t+T}), \tag{21}$$

where $\mathcal{L}$ is given in (17) and each component of the loss is specific to a different imputation stage and processing direction. The specification of the evaluation mask, $\bar{\mathbf{M}}$, differs depending on training and fine-tuning as detailed in the next sections.

C Data and implementation details

C.1 Dataset

Our dataset comprises recordings from 51 patients with persistent AF, obtained using the AcQMap system. The system uses non-contact electrodes to sense intra-cavitary electrograms, from which dipole density measurements (charge density in Coulombs/cm) are inversely derived across the entire atria [12, 13]. Each recording samples approximately 3500 nodes at 3000 Hz for 5-20 seconds. All recordings were obtained prior to ablation at two United Kingdom centres between 2016 and 2023. The patient cohort (mean age 64 ± 11 years, 69% male) had standard cardiovascular risk factors and were on guideline-directed medical therapy (see Table 2 for more details).

For pre-processing, these signals are first resampled spatially to a resolution consisting of 500 nodes. This is done by k-means clustering of the 3D node coordinates into $k = 500$ non-overlapping clusters,

where signals are resampled using the mean average signal within each cluster for each time point. Temporal resampling to 70 Hz is conducted through a combination of low pass filtering and down sampling. Low-pass filtering is applied to each time series at 70 Hz to prevent aliasing, and then the filtered signal is downsampled by reducing the sampling rate proportionally. Finally, the time series are normalised across space and time by applying min-max normalisation, where the minimum and maximum values are determined across all nodes and times, ensuring that the amplitude of the resampled signals is scaled consistently between 0 and 1.

The dataset was stratified to ensure a spectrum of AF dynamics within training (70%), validation (10%), and test (20%) sets. The Shannon entropy was first computed for the resampled time series at each node to do this. The cumulative distribution function (CDF) of Shannon entropy across the atrium was then created for each patient, serving as a measure to compare the similarity of the spatiotemporal dynamics between patients. The similarity between patients was assessed by computing KolmogorovSmirnov (KS) distance between CDFs for each pair of patients, where a smaller KS distance represents more similar entropy distributions/spatiotemporal dynamics. From this, groups of similar patients were identified by clustering a Laplacian eigenmap [51] using k-means, where the number of clusters was determined using the elbow method. This involved forming a weighted adjacency matrix by applying a radial basis function kernel to the reciprocal of the KS distances and thresholding. Finally, the training, validation, and test sets were formed by performing stratified sampling across these clusters, maintaining a consistent spectrum of AF dynamics within each set.

C.2 Simulating sequential contact mapping

A sequential contact mapping strategy is simulated from the non-contact recordings as a self-avoiding walk of the multipolar catheter (represented as a patch of observations) on the atrial surface; whereby the catheter surface area, dwell time, and spatial overlap, could be controlled. The self-avoiding walk is defined by first providing the catheter surface area as a fraction of the total area, where its reciprocal (1/area) is rounded up to the nearest integer to give the number of patches required to sample the whole atrium. Then, the atrium is split into disjoint patches by k-means clustering the 3D node coordinates, with k equal to the number of patches. Spatial overlap between patches is simulated by adding additional clusters between adjacent patches and sharing node sets in proportion to the overlap required. A self-avoiding random walk ensures the path does not revisit the same region twice. This is simulated by sampling a patch randomly, then sampling the remaining patches without replacement with a probability proportional to the distance between the current and remaining patches. The resulting sequence of patches is converted to an observation mask, $\mathbf{M}$, by assigning a unity value if nodes are within the observed patch, zero otherwise, and repeating these values such that each patch is observed for a duration equal to the specified dwell time. Except during the sensitivity analysis, the sequence of patches is repeated until the length of the available recording is met.

C.3 Training FIBMAP

The training procedure of FIBMAP leverages the whole atria ground truth signals of each patient to learn a robust function for reconstructing whole atrium dynamics from data collected in a sequential contact mapping fashion. To do this, a self-supervised approach is employed, wherein self-avoiding walks of the multipolar catheter are sampled at random to form the observation mask, $\mathbf{M}$, and observed input data, $\mathbf{X}$, of each training sample, alongside a range of catheter surface areas (2.5 50% of the atrium), dwell times (0.2 - 4 seconds) and spatial overlaps (0 3 additional clusters between adjacent patches). At each training iteration, batches of size B are formed by collating samples from different patients, whereby B inputs $\mathbf{X}^{(p)}_{t:t+W}, \mathbf{M}^{(p)}_{t:t+W}, \mathbf{A}^{(p)}, \mathbf{V}^{(p)}, \mathbf{g}^{(p)}$ are collated for a temporal input window size, W, sampled from the original time series following a sliding window approach with unit stride.

Imputation is performed within the temporal window and a whole atria reconstruction loss is used to optimise FIBMAP during training, whereby (21) is optimised using an evaluation mask, $\bar{\mathbf{M}}_{t:t+W}$, with all nodes and times in the input window equal to unity (see Figure 4). This approach ensures a robust function for reconstructing whole atria maps that generalise across the distribution of multipolar catheter paths and parameters of sequential contact mapping such as dwell time.

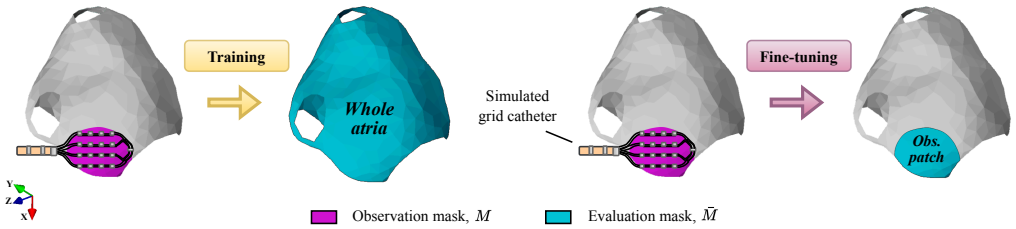

Figure 4: Definition of the observation and evaluation masks in FIBMAP training and fine-tuning. During training, FIBMAP is optimised to perform whole atria mapping using a whole atria evaluation mask to compute the loss function. During fine-tuning, only the patient-specific parameters of FIBMAP are optimised using an observed patch reconstruction loss.

A hyperparameter search is conducted for the parameters shown in Table 3, by first splitting the time series of each patient sequentially, with the first 85% of time steps being used for training, the second 5% for validation, and the final 10% for testing. For the validation and test sequences, a fixed self-avoiding walk of the multipolar catheter is used with a surface area of 10%, a dwell time of 1 second, and no spatial overlap. Each configuration in the hyperparameter search is conducted for 100 epochs, whereby the MAE across the remaining atria for the validation sequence is monitored with early stopping. All experiments were performed on a NVIDIA RTX A5000 graphics processing unit. The best-performing set of hyperparameters are chosen by computing the MAE loss across the remaining atria for the test sequence and selecting the minimum loss. The best-performing configuration for training FIBMAP was found to be $B = 16$, $W = 40$, $d = 64$, $q = 64$, $r = 16$, $K = 1$, and 1024 hidden layer neurons. This configuration was retrained for 500 epochs, where a learning rate of 0.0009 and a cosine scheduler were used. Training took a total of 31 hours. The result is a pre-trained FibMap model, which performs accurate and robust whole atria reconstruction from partial observations across a spectrum of dynamics found in patients and variations in multipolar mapping.

C.4 Fine-tuning FIBMAP

In clinical practice, only the patches of multipolar catheter measurements are available from sequential contact mapping. The rest of the atrium remains unobserved. To make FIBMAP a feasible clinical solution for imputation mapping, a fine-tuning procedure is introduced to enable accurate whole atria reconstruction on new and unseen patients when only partial observations of the dynamics are available.

Our fine-tuning procedure preserves essential knowledge for whole atria reconstruction acquired during training by fixing the parameters of the pre-trained FIBMAP model, while quickly personalising the model to new patients by optimising only the node and patient embedding parameters using an observed patch reconstruction loss (see Figure 4 for an illustration of the observation and evaluation masks used during fine-tuning, which are defined to be equal). The observed patch reconstruction loss is defined using this evaluation mask in (21). This restricts the optimisation during fine-tuning to only the patches of multipolar catheter measurements.

Our fine-tuning procedure was configured on the validation set, which also has ground truth whole atria signals available for patients. Validation of the fine-tuning procedures aims to evaluate the relationship between the observed patch and whole atria reconstruction losses. Again, the time series of each patient was split sequentially, with the first 85% of time steps being used for training, the second 5% for validation, and the final 10% for testing.

A random hyperparameter search was conducted across learning rates [0.0005, 0.005] and batch sizes [16, 32, 64, 128]. Each learning rate in the hyperparameter search was conducted for 100 epochs, whereby the MAE across the remaining atria for the validation sequence was monitored with early stopping. For the validation and test sequences, a fixed self-avoiding walk of the multipolar catheter was used with a surface area of 10%, a dwell time of 1 second, and no spatial overlap. The best-performing set of hyperparameters was chosen by computing the MAE across the remaining atria for the test sequence and selecting the minimum loss. The best-performing configuration for fine-tuning FIBMAP was found to have a learning rate of 0.005 and a batch size of 16.

Table 3: Hyperparameter values tested during FIBMAP training.

Hyperparameter	**Range tested**
Batch size, B	[16, 32, 64]
Input window length, W	[20, 30, 40, 50]
Hidden size, d	[64, 128, 256]
Node embedding size, q	[16, 32, 64]
Patient embedding size, r	[16, 32, 64]
Layers, K	[1, 2, 3]
Hidden layer neurons in MLP_{enc}, MLP_{dec}	[128, 256, 512, 1024]

Finally, the performance of FIBMAP imputation mapping on new and unseen patients, through our fine-tuning setup, was quantified on the test set patients. A sequential time series split was not used during testing, instead, all observed measurements were used. FIBMAP was fine-tuned for 100 epochs using the configuration found in validation, and the observed patch reconstruction loss was monitored to perform early stopping. The performance metrics, detailed next, were computed across the remaining atria to assess test set performance.

C.5 Performance metrics

Quantitative assessment of the reconstructed imputation maps was performed using several metrics: MAE, mean relative error (MRE) and MAPE. These metrics evaluate the fidelity of the reconstructed whole atria maps, $\widehat{\mathbf{Y}}$, against the ground truth, $\mathbf{X}$, for the mask, $\bar{\mathbf{M}}$, which represents the logical binary complement of the observation mask, $\mathbf{M}$. The averaged performance metrics were computed via (17), where $\mathcal{L}_{t,i}$ was computed using each of our evaluation metrics. Metrics were computed for each model trained or fine-tuned across 5 different seeds (changing both the self-avoiding walk and model parameter initialisations). The average and standard deviation across these seeds were reported for each metric.

Additionally, the PS F1-score was used to evaluate the accuracy of tracking PSs in the reconstructed phase maps compared to the ground truth. PSs represent points where the Hilbert phase is undefined and mark the organising centres of spiral waves that may drive AF, making their accurate detection clinically relevant for ablation guidance.

Manual annotation For each patient in the test set and each imputation model, PSs were manually labelled by two independent trained observers using a custom graphical user interface. Manual annotation was chosen over automated detection methods due to the poor performance of existing automated algorithms [52] at the reduced spatial and temporal resolutions used in this proof-of-concept study (500 nodes, 70 Hz). Automated methods typically require higher resolution data to reliably identify PS locations, making manual expert annotation the gold standard for evaluation at these scales.

Given the labour-intensive nature of manual annotation, only the central 70 frames (1 second) of each reconstructed phase map were annotated for each patient and model combination. Since annotating PSs for each model seed across multiple random initializations was computationally infeasible, the mean and standard deviation of F1-scores were computed from 1000 bootstrap samples.

Detection thresholds PS detection was evaluated at four clinically relevant precision levels with varying spatial and temporal tolerances: High Precision (≈0.5cm, 0.1s), Moderate Precision (≈1.0cm, 0.2s), Standard Precision (≈1.5cm, 0.3s), and Lenient Precision (≈2.0cm, 0.4s). A PS was considered correctly detected if its location in the reconstructed map fell within the specified spatial tolerance (measured in graph hops) and temporal tolerance compared to its location in the ground truth map. Results across all thresholds are shown in Table 4, with FIBMAP achieving superior performance at all precision levels.

Inter-observer agreement and validation Inter-observer agreement was quantified to validate the reliability of manual annotations, achieving an F1-score of 0.827 ± 0.243 and Cohen's κ of

Table 4: PS detection F1-scores at clinically relevant thresholds.

Method	High Precision (≈0.5cm, 0.1s)	Moderate Precision (≈1.0cm, 0.2s)	Standard Precision (≈1.5cm, 0.3s)	Lenient Precision (≈2.0cm, 0.4s)
Transductive				
Mean	0.009±0.006	0.022±0.009	0.038±0.010	0.047±0.010
MF	0.049±0.025	0.103±0.043	0.149±0.060	0.206±0.099
Inductive				
RNN	0.053±0.022	0.064±0.022	0.070±0.025	0.080±0.027
Bi-RNN	0.082±0.032	0.113±0.045	0.132±0.047	0.150±0.047
TTS-Transformer	0.044±0.022	0.053±0.026	0.060±0.025	0.075±0.024
Fine-tuned				
RNN + FT	0.108±0.040	0.148±0.047	0.182±0.055	0.205±0.061
Bi-RNN + FT	0.133±0.044	0.183±0.048	0.229±0.054	0.263±0.065
TTS-Transformer + FT	0.193±0.074	0.258±0.079	0.310±0.098	0.348±0.105
FIBMAP	**0.555±0.066**	**0.780±0.053**	**0.903±0.036**	**0.964±0.021**

Table 5: Inter-observer agreement and comparison with automated detection for PS detection.

Validation Metric	F1-Score	Cohen's κ
Inter-Observer Agreement	**0.827±0.243**	**0.885±0.187**
Automated vs Observer 1	0.238±0.130	0.352±0.148
Automated vs Observer 2	0.292±0.151	0.392±0.132

Values represent mean ± standard deviation across patients.

0.885 ± 0.187 across patients (Table 5), indicating excellent agreement beyond chance at the High Precision tolerance level (≈0.5cm, 0.1s). To further validate the superiority of manual annotation, we compared automated detection performance using the 3D topological charge algorithm [52] against both observers, finding substantially lower agreement (F1-scores of 0.238 and 0.292 respectively), confirming that manual expert annotation provides more reliable ground truth labels for PS detection evaluation at these resolutions.

C.6 Baseline models

We introduce additional baseline models for the imputation mapping task:

1. Mean, which performs imputation using the node-level average;
2. MF with rank = 10;
3. Univariate RNN, which performs imputation based solely on the node-level signals;
4. Univariate Bi-RNN.
5. TTS-Transformer, which applies temporal self-attention followed by spatial self-attention using 8-head multi-head attention mechanisms [53].

Mean and MF baseline models are employed solely on the test set due to their transductive nature. The RNN, Bi-RNN, and TTS-Transformer models were initially trained using MAE loss function and followed identical hyperparameter settings and training-test protocols as FIBMAP.

To evaluate patient-specific adaptation, we implemented fine-tuned (FT) variants (RNN + FT, Bi-RNN + FT, TTS-Transformer + FT) where pre-trained models are fine-tuned for 50 epochs on each patient using a learning rate of 0.001. We employed the same observed patch reconstruction loss for fine-tuning as used in FIBMAP. Selected layers were fine-tuned: only input-to-hidden weights/biases and readout layers for RNN models; and input encoder, readout, and normalisation layers for TTS-Transformer. This preserves learned representations while enabling patient-specific adaptation.

C.7 Imputation mapping from EnSite Precision HD Grid recordings

From our test cohort, three AF patients had both EnSite Precision HD Grid and AcQMap recordings collected non-contemporaneously before ablation. For these patients, AcQMap recordings were 20 seconds in duration, while EnSite Precision HD Grid recordings were significantly longer (14, 19, and 20 minutes). EnSite Precision HD Grid data consists of sequential contact mapping recordings from 16 electrodes arranged in a grid array, along with the roving 3D coordinates of each electrode within the atrium.

To ensure compatibility with FIBMAP, the EnSite Precision HD Grid recordings were pre-processed. Sparse electrogram recordings were first mapped to a uniform discretisation of the atrial surface using a nearest-neighbour approach with a 3 mm radius to interpolate the signals between electrodes. An observation mask identified active recording periods, after which signals were normalised using the maximum peak-to-peak voltage. The data was then resampled spatially to a resolution of 500 nodes using k-means clustering and resampled temporally to 70 Hz using a combination of low-pass filtering and downsampling. A final min-max normalisation ensured all signals fell within 0 and 1. From these processed measurements, imputation maps were generated using our fine-tuning procedure, whereby the observed patch reconstruction loss was monitored to perform early stopping.

We developed a sliding window cross-correlation framework to compare FIBMAP reconstructions against AcQMap 'ground truth' recordings. While temporal alignment was not possible between the non-contemporaneous recordings, spatial alignment was achieved between AcQMap and imputation map by centring the vertices of the geometries around the origin, applying rigid registration using the iterative closest point algorithm [54], and projecting data between geometries using a $k = 5$ nearest-neighbours regression. Using sliding window lengths from 0.5 to 4.0 seconds and a constant stride of 0.1 seconds, we computed the Pearson correlations between Hilbert phases of processed signals across all nodes and times within window pairs. This generated a cross-correlation matrix characterising the spatiotemporal similarity between recordings, which were flattened and plotted as distributions for analysis.

To validate that FIBMAP captured patient-specific dynamics, we performed three types of correlations: intra-patient comparisons between AcQMap and FIBMAP from the same patient; inter-patient comparisons between AcQMap and FIBMAP from different patients; and random baseline intra-patient comparisons between AcQMap and a spatiotemporally shuffled imputation map. For each patient, the 99th percentile of the intra, inter and shuffled distributions were plotted and statistical significance was assessed by computing the confidence intervals via bootstrapping ($n = 1000$ rounds of 10000 resamples).

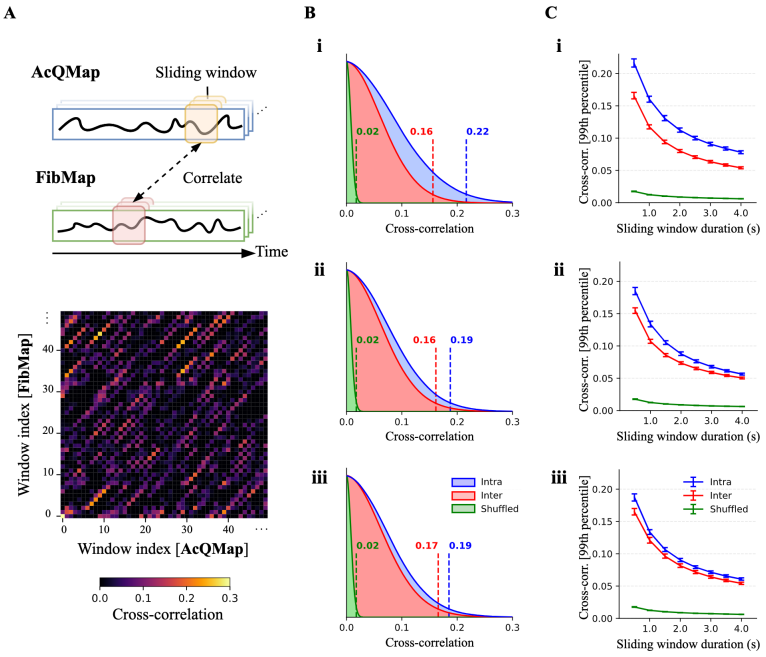

Figure 5: Validation of FIBMAP imputation maps from EnSite Precision HD Grid Mapping against non-simultaneous ground truth AcQMap recordings. A) Sliding window cross-correlation analysis between AcQMap and FIBMAP phase signals, enabling comparison between non-simultaneous recordings by measuring the similarity between AF patterns across different time windows. The correlation matrix below shows pairwise comparisons between all possible window combinations. For three patients (indexed i-iii), B) shows the kernel density estimates of pairwise cross-correlations computed between AcQMap and FIBMAP maps from the same patient (intra), different patients (inter), and spatiotemporally shuffled maps (random baseline), using 0.5-second sliding windows. Vertical dashed lines indicate the 99th percentile, with consistently higher values for intra-patient (0.19-0.22) versus inter-patient (0.16-0.17) correlations and shuffled baseline (0.02), demonstrating patient-specific pattern capture. C) Temporal robustness analysis showing the 99th percentile of cross-correlations against sliding window duration. Non-overlapping confidence intervals (computed via bootstrap sampling) between intra-patient and other distributions confirm that FIBMAP can capture patient-specific dynamics across different temporal scales.