Design Statement

Problem

Many neuroimaging techniques produce sparse spatial data that, while valuable, are difficult to interpret or visualise comprehensively. For example, stereo electroencephalography (sEEG) provides intracranial recordings with high temporal resolution but limited spatial coverage [1]. At the same time, gene expression data from post-mortem tissue samples offer insight into molecular organisation, but at sparse anatomical points [2]. These data gaps hinder multimodal integration, spatial pattern recognition, and clinical decision-making in domains such as epilepsy surgery and neuropsychiatric research.

Despite advances in interpolation techniques, there is limited consensus on which methods best reconstruct neuroimaging data. Classical methods (e.g., spline, KNN) are computationally simple but may not capture anatomical complexity. Conversely, advanced techniques like Kriging and deep learning promise better accuracy but have trade-offs in computational efficiency and interpretability [3].

State of the Art

Previous work in image inpainting and spatial interpolation has explored statistical and machine learning techniques. Kriging, a geostatistical method [3], is widely used in environmental modelling and has been adapted for neuroimaging due to its spatial interpretability. Multi-Layer Perceptrons (MLPs) [4], particularly with residual connections, can model nonlinear relationships in spatial data and generalise well to diverse inputs. U-Net architectures, initially designed for biomedical image segmentation, have gained popularity for dense volumetric inference tasks due to their encoder-decoder structure and use of skip connections [3, 5].

While promising, each method has its limitations. Kriging struggles with high-dimensional data and computational time [3]. Though efficient, MLPs may fail to capture spatial locality [3], and U-Nets may overfit sparse inputs, generating anatomically implausible outputs. Previous studies have rarely compared these methods directly on sparse neuroimaging tasks, especially across functional (sEEG) and molecular (gene expression) domains.

Aims

This project is designed to evaluate and compare three advanced interpolation methods, Kriging, Residual MLP, and 3D U-Net, on sparse neuroimaging datasets. The study seeks to understand how well each model performs across functional and molecular modalities, specifically using sEEG and gene expression data. The project involves implementing and optimising each method, applying them systematically to both datasets, and assessing their predictive accuracy using a suite of regression-based metrics. In addition to quantitative evaluations, qualitative assessments of spatial plausibility will be conducted to determine each model's ability to reconstruct anatomically faithful outputs. Each advanced method will also be compared against classical interpolation approaches, such as Spline and K-Nearest Neighbours, previously explored by other team members. Ultimately, the aim is to recommend interpolation techniques that best balance predictive performance, computational efficiency, and real-world usability for neuroscientific and clinical applications.

Work Plan

The project will be divided into six phases. The initial phase (Phase 0) will involve conducting a comprehensive literature review to explore existing interpolation and inpainting techniques relevant to sparse neuroimaging data. This review will guide the selection of three advanced methods: Kriging, Residual MLP, and 3D U-Net, which will be the focus of this project due to their potential for spatial accuracy, robustness, and applicability to high-dimensional biomedical data.

In Phase I, the sEEG and gene expression datasets will be acquired, cleaned, and pre-processed. Spatial coordinates will be standardised, and data sparsity will be visualised to inform model setup. During this phase, other group members will develop baseline interpolation methods such as Spline, Polynomial interpolation, and K-Nearest Neighbours. These classical methods will later serve as a benchmark for comparison with the advanced approaches implemented in this project.

Phase II will focus on implementing Kriging using the PyKridge library. Three variogram models, Gaussian, Spherical, and Exponential, will be evaluated to assess how spatial correlation assumptions affect

interpolation performance. Interpolations will be performed over dense 3D grids generated within the convex hull of observed coordinates for both datasets.
In Phase III, machine learning-based methods will be developed. A residual Multi-Layer Perceptron (MLP) will be constructed using PyTorch, incorporating dropouts, batch normalisation, and LeakyReLU activations to improve generalisation and training stability. In parallel, a 3D U-Net will be implemented, featuring an encoder-decoder structure with bottleneck residual blocks and CBAM attention modules for enhanced feature refinement. Both models will be trained using a hybrid loss function that balances accuracy and spatial smoothness. Phase IV will involve evaluating all models through 10-fold cross-validation. Data augmentation techniques such as jitter and Gaussian noise will be applied to simulate real-world variability. Each method's performance will be assessed using quantitative metrics (e.g., R^2, RMSE, MAE) and visual inspection of the reconstructed volumes for spatial plausibility. In Phase V, the advanced methods developed in this project will be compared against the classical techniques implemented by other group members. This comparison will evaluate trade-offs in performance, interpretability, and computational efficiency. Results and visual outputs will be consolidated into a written report with recommendations for the most suitable interpolation approaches in neuroscientific and clinical applications.

Deliverables

The primary outcomes of this project include a reproducible codebase implementing Kriging, MLP, and U-Net for sparse neuroimaging data, a comprehensive quantitative evaluation using metrics such as R^2, RMSE, and MAE, a qualitative assessment of spatial reconstructions and anatomical plausibility, and a comparative benchmark against classical interpolation methods developed by other group members. These deliverables aim to inform academic research and practical applications, including neurosurgical planning and molecular brain planning.

Evaluation

During Phase IV, model performance will be evaluated across multiple dimensions. Predictive accuracy will be quantified using standard regression metrics, including Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Coefficient of Determination (R^2). Spatial fidelity will be assessed through visual inspection of interpolated maps, particularly focusing on whether regional anatomical features are preserved or distorted. Computational efficiency will be measured by recording training and inference times, highlighting models that balance speed with accuracy. Robustness will be tested by applying jitter and noise-based augmentation during training, evaluating how models respond to simulated variability in real-world datasets. Informal feedback on the visual outputs will be collected from peers to validate interpretability.

Project Timeline

Below is the timeline that outlines the key phases of the project, from method selection to final reporting:

	February	March	April	May
Phase 0				
Phase I				
Phase II				
Phase III				
Phase IV				
Phase V				
Report Writing				

7MRI0120 Group Research Project Report

Evaluating Interpolation and In-Painting Techniques for Sparse Neuroimaging Data Using Statistical and Machine Learning Approaches

Supervisors: Konrad Wagstyl and Emma Robinson

Author: Andrea Walker Perez

Co-Authors: Safa Khan, Kunyu Li, Minshen, Yuanting

Acknowledgements

All authors would like to express their gratitude to Dr Konrad Wagstyl and D Emma Robinson, the supervisors, and King's College London for supporting this project.

Abstract

Sparse neuroimaging data, such as stereo electroencephalography (sEEG) and gene expression, present challenges for full-brain analysis due to limited spatial coverage. This study evaluates three advanced interpolation methods, Kriging, Residual Multi-Layer Perceptron (MLP), and 3D U-Net, on sEEG and gene expression datasets to reconstruct missing spatial information. Models were assessed using root mean squared error (RMSE), mean absolute error (MAE), R^2, and visual reconstructions. While U-Net achieved the highest R^2 and fastest runtime, its outputs lacked anatomical plausibility. Kriging and MLP produced more interpretable results but required longer computation. Surprisingly, classical spline interpolation outperformed all advanced methods on both datasets, highlighting that simpler techniques may offer more practical performance in sparse neuroimaging contexts. Model selection should consider numerical accuracy and anatomical realism to support clinical or neuroscientific use.

1. Introduction & Background

Understanding the human brain remains a challenge in neuroscience. While neuroimaging technologies like Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET) provide high-resolution, full-brain coverage [6], other modalities, like stereo electroencephalography (sEEG), and post-mortem gene expression profiling, produce sparse spatial data [1, 7]. These datasets are essential for clinical use, including epilepsy surgery and gene mapping, but their incomplete coverage limits multimodal analysis and interpretation [8].

To address these limitations, interpolation and inpainting techniques estimate values at unobserved spatial locations. Classical methods (e.g., spline, polynomial, KNN) are computationally simple but may fail to capture complex anatomical structures [8, 9]. More advanced techniques like Kriging and deep learning can model spatial relationships more effectively, though they often involve higher computational costs and reduced interpretability [5, 9].

Previous studies typically focus on a single modality or prioritise numerical accuracy, with limited assessment of spatial plausibility, an important factor in biomedical contexts. Few have systematically compared interpolation methods across functional (sEEG) and molecular (gene expression) neuroimaging domains.

Sparse data reconstruction has broad biomedical relevance. More accurate and plausible reconstructions could improve brain atlases, guide neurosurgical planning, and enhance molecular brain mapping [10]. These benefits extend to clinicians, neuroscientists, and developers of AI tools. In low-resource settings, effective interpolation could reduce the need for costly imaging or extensive data collection.

However, using interpolated neuroimaging data introduces ethical and regulatory concerns. Models that produce anatomically implausible results may mislead diagnoses or treatment decisions. Clinical adoption requires interpretable, unbiased, and robust models across diverse populations. Greater involvement of patients in validating these outputs could also support the co-design of tools that align with clinical expectations and end-user needs. Inclusive dataset design is also essential, as interpolation models trained on limited or homogeneous populations risk amplifying existing biases. Future adaptation to underrepresented groups or rare neurological conditions will be critical for ensuring broad applicability and equitable outcomes. For example, gene expression variability across demographic groups highlights the need for inclusive validation [11].

There are also commercial opportunities. Scalable, robust interpolation tools could be integrated into software used by digital health or pharmaceutical companies. However, barriers remain in ensuring regulatory approval, user trust, and integration with existing clinical pipelines. Addressing these challenges is essential for real-world adoption.

By exploring these trade-offs, this project aims to benchmark interpolation methods on sEEG and gene expression data, identifying practical and interpretable approaches that balance quantitative performance, spatial plausibility, computational efficiency, and clinical relevance.

2. Methods and Theory

2.1. Datasets

Two datasets were used:

- **sEEG Dataset**

Derived from a publicly available atlas of intracranial EEG recordings during wakefulness [7], this dataset includes recordings from 1,785 electrode channels distributed across 106 patients. The signals span five frequency bands of interest: delta (0.5-4 Hz), theta (4-7 Hz), alpha (8-12 Hz), beta (13-30 Hz), and gamma (30-80 Hz), with power spectrum estimates computed for each contact. Electrode positions are recorded in MNI space, and the resulting dataset is sparse yet spatially rich across all cortical regions.

- **Gene Expression Data**

This data originates from the Allen Human Brain Atlas [2]. It provides spatially mapped microarray gene expression measurements from six post-mortem brains. The analysis focused on 46 neurotransmitter receptor genes associated with expression levels at discrete MNI-space coordinates. Like the sEEG data, this dataset is sparse but offers critical molecular insights into brain organisation.

Both datasets were pre-processed using standardisation and coordinate normalisation. These datasets allow for evaluating interpolation methods across functional (sEEG) and molecular (gene expression) modalities.

2.2. Kriging

Kriging is a geostatistical method that estimates the value of a variable at unobserved locations based on the spatial autocorrelation of observed values [3].

The estimated value $\hat{z}(x_0)$ is expressed as a weighted sum of known values [3]:

$$\hat{z}(x_0) = \sum_{i=1}^{N} \lambda_i Z(x_i) \tag{1}$$

subject to $\sum_{i=1}^{N} \lambda_i = 1$. The weights λ_i minimise the estimation variance and are derived by solving the Kriging system (2), which incorporates the variogram function $\gamma(h)$ that models spatial dependence between points separated by $lag\ h$ [3]. The system ensures unbiasedness and provides optimal weights for interpolation based on spatial autocorrelation [3].

$$\begin{cases} \sum_{j=1}^{N} \lambda_j \gamma\left(\left\|x_i - x_j\right\|\right) + \mu = \gamma(\|x_i - x_0\|), for\ i = 1, 2, \ldots, N \\ \sum_{j=1}^{N} \lambda_j = 1 \end{cases} \tag{2}$$

Three variogram models were evaluated:

- **Gaussian**: Smooth, globally continuous interpolations [12].

$$\gamma(h) = c_0 + c_1 \left[1 - \exp\left(-\frac{h^2}{a^2}\right)\right] \tag{3}$$

- **Spherical**: Compact support, correlation decreases to zero beyond a specific range [12].

$$\gamma(h) = c_0 + c_1 \left[\frac{3h}{2a} - \frac{h^3}{2a^3}\right], h \leq a \tag{4}$$

- **Exponential**: Allows sharper local transitions, favouring short-range spatial correlation [12].

$$\gamma(h) = c_0 + c_1 \left[1 - \exp\left(-\frac{h}{a}\right)\right] \tag{5}$$

Each model captures different spatial behaviours. Kriging was implemented using PyKridge's *OrdinaryKriging3D* [13], and interpolations were conducted within the convex hull of input points to allow volumetric visualisation on both datasets.

PyKrige was selected for its efficient support of 3D variogram modelling and its open-source integration with NumPy, making it well-suited for prototyping spatial models in scientific computing workflows.

2.3. Multi-Layer Perceptron (MLP)

The Multilayer Perceptron (MLP) architecture used in this study incorporates residual connections inspired by recent developments in neural network design [14]. Residual connections enhance gradient flow and enable training deeper networks effectively [14]:

$$y = F(x, \{W_i\}) + x \tag{6}$$

The implemented residual MLP consisted of:

- **Input layer:** Polynomial-expanded spatial coordinates were standardised and used as input [15].
- **Hidden layers:** Five connected hidden layers, each containing 128 neurons, were used. Each hidden layer used batch normalisation [16], dropout (with a dropout rate of 0.3) [17], and LeakyReLU activations [18].
- **Residual connections:** Each hidden layer had residual connections to facilitate training and mitigate issues such as vanishing gradients [14].
- **Output layer:** A final linear layer predicted interpolated values.

The model was trained using the Smooth L1 loss function [19] and optimised with the Adam optimiser [20]. Pytorch [21] was used due to its flexibility, modular architecture design, and compatibility with GPU acceleration, enabling fast experimentation and scalable training across architectures. This architecture was selected for its ability to model nonlinear relationships while remaining computationally efficient. Figure 1 is a schematic representation of the residual MLP architecture implemented in this study.

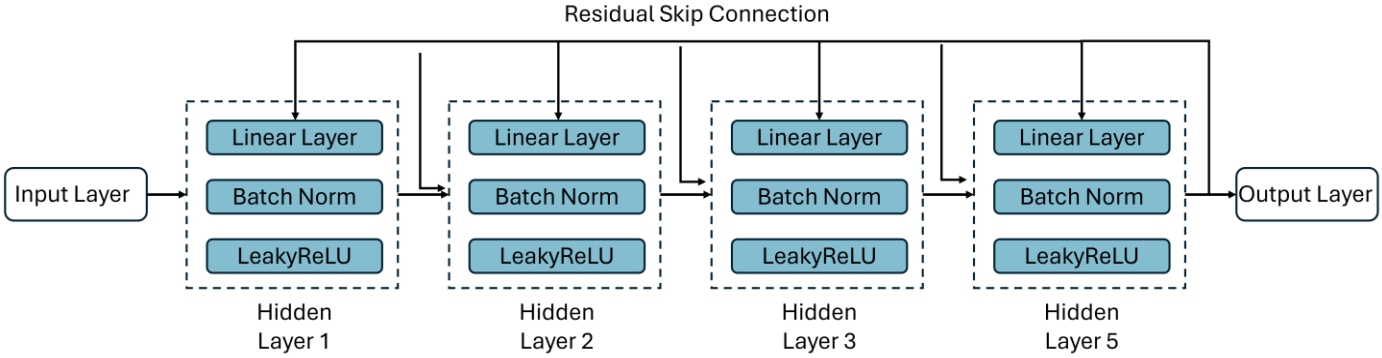


Figure 1 Architecture of the Residual MLP used for neuroimaging data interpolation. The network takes polynomial-expanded and standardised 3D spatial coordinates as input, passes them through five hidden layers with batch normalisation, dropout (0.3), and LeakyReLU activation, and incorporates residual connections to improve training stability. The final output layer predicts the interpolated values at unobserved spatial locations.

2.4. U-Net

The 3D U-Net implemented here was adapted from the inpainting architecture proposed by Liu et al. [22], which incorporates an encoder-decoder structure with residual bottlenecks and Convolutional Block Attention Modules (CBAM) for feature refinement [23].

The implemented 3D U-Net comprises (see Figure 2 for U-Net Architecture):

- **Input Layer:** Two-channel input representing normalised voxel intensity and a binary spatial mask.
- **Encoder Path:** Two encoding stages with 3D convolutions, batch normalisation, ReLU activation, and bottleneck residual blocks. Downsampling via strided convolutions.
- **Bottleneck Layer:** Deepest feature representation using 3D convolutions and residual bottleneck design.
- **Decoder Path:** Two upsampling stages with transpose convolutions. Decoder convolutions with CBAM attention modules for feature refinement [23]. Skip connections from encoder to decoder to retain spatial detail.
- **Output Layer:** Single-channel 3D convolution to predict interpolated voxel intensities.

This architecture was trained with a hybrid loss function combining L1 loss over known voxels [19] and a 3D total variation loss to encourage spatial smoothness [24]. The model was implemented in Pytorch [21]and optimised using Adam [20].

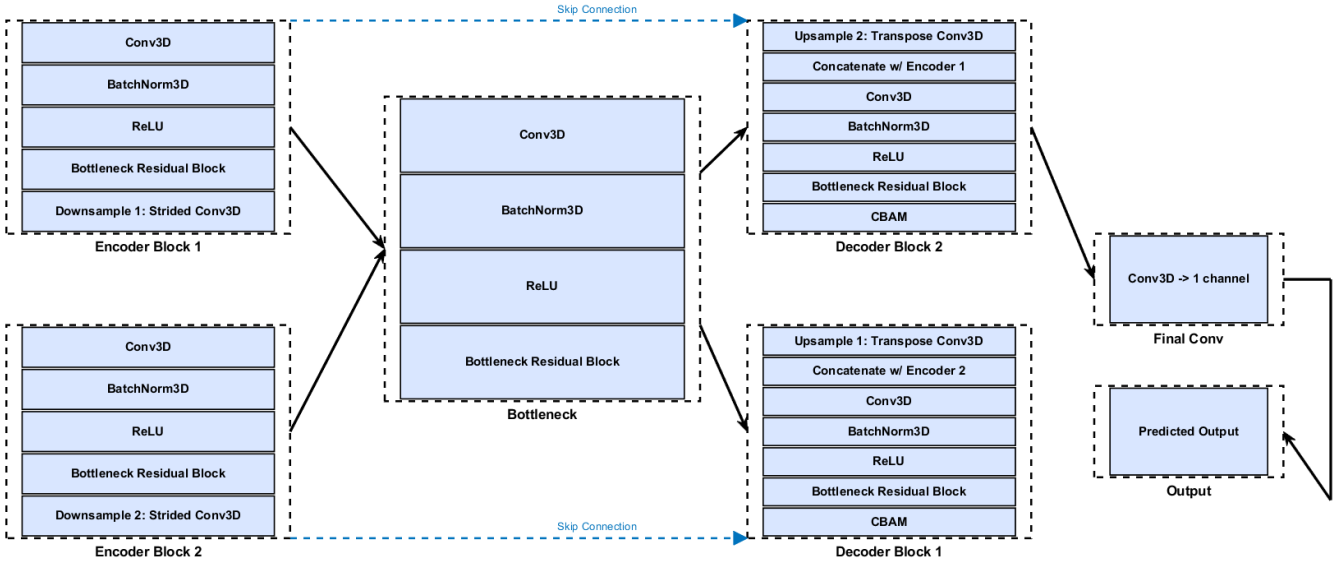


Figure 2 Modular 3D U-Net Architecture

2.5. Evaluation & Validation

To ensure a comprehensive evaluation of interpolation performance, we employed accuracy-based and computational efficiency-based assessments across all methods.

Data Augmentation

To enhance model generalisation and simulate real-world data variability, we applied the following augmentations (Table 1), as similarly motivated in prior neuroimaging and deep learning studies [25-27].

Table 1 Data Augmentation Techniques

Technique	Description
Jitter	Random spatial shifts (±1 voxel) across all three axes.
Gaussian Noise	Additive zero-mean Gaussian noise (σ = 5% of standard deviation)
Jitter + Noise	Combined application of spatial jitter and Gaussian noise perturbation

Cross-Validation

A 10-fold cross-validation scheme was adopted to evaluate generalisation performance. This was repeated per model and augmentation condition, with folds sampled across known (non-zero) voxels. Cross-validation is considered the gold standard for performance estimation in sparse biomedical datasets [28].

Evaluation Metrics

Interpolation accuracy was assessed using the metric listed in Table 2, which provide complementary insights into magnitude and relational performance [29, 30].

Table 2 Accuracy-Based Evaluation Metrics

Metric	Description
Mean Squared Error (MSE)	Measures the average squared difference between predicted and actual values, penalising large deviations.

Mean Absolute Error (MAE)	Computes the average magnitude of absolute errors and is more robust to outliers.
Root Mean Squared Error (RMSE)	Square root of MSE, making it directly interpretable in the same units as the data.
Coefficient of Determination (R^2)	Proportion of variance in the ground truth explained by the model.

Computational Efficiency

In addition to accuracy, we recorded timing statistics for each model:

- Average time per epoch.
- Total time for cross-validation.
- Total time for final training.

These metrics were critical for comparing models in deployment settings. Approaches that deliver similar performance at lower computational cost are more scalable and practical in real-world applications.

3. Results

3.1. Interpolation Performance on sEEG Dataset

3.1.1. Quantitative Performance

U-Net achieved the highest R^2 (0.938), followed by Kriging (~0.35) and MLP (0.20), as shown in Table 3. These trends were consistent across all frequency bands and visualised in Figure 3.

Table 3 Average interpolation metrics across all frequency bands (no augmentation)

Model	R^2 (↑)	MSE (↓)	MAE (↓)	RMSE (↓)
Kriging (Gaussian)	0.353 ± 0.038	0.0031 ± 0.0004	0.0436 ± 0.0042	0.0567 ± 0.0035
Kriging (Spherical)	0.355 ± 0.042	0.0032 ± 0.0005	0.0437 ± 0.0046	0.0569 ± 0.0038
Kriging (Exponential)	0.357 ± 0.040	0.0031 ± 0.004	0.0435 ± 0.0041	0.0568 ± 0.0036
MLP	0.200 ± 0.045	0.0038 ± 0.0004	0.0496 ± 0.0038	0.0616 ± 0.0033
U-Net	0.938 ± 0.047	0.119 ± 0.0003	0.265 ± 0.0032	0.343 ± 0.0029

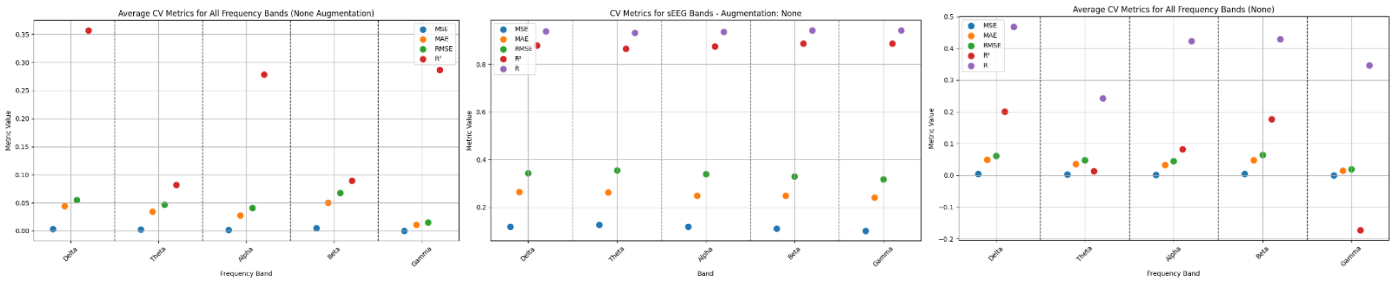


Figure 3 Cross-validation performance across frequency bands (no augmentation) for each interpolation method (Kriging, MLP, and U-Net)

3.1.2. Visual Performance Across Models

Despite high scores, U-Net outputs appeared overly smooth and lacked anatomical fidelity, suggesting overfitting. Kriging and MLP preserve finer regional structure, particularly in Delta, Alpha, and Gamma

bands (see Figures 4-9 for Delta Interpolation). See **Appendix A** for the remaining frequency bands.

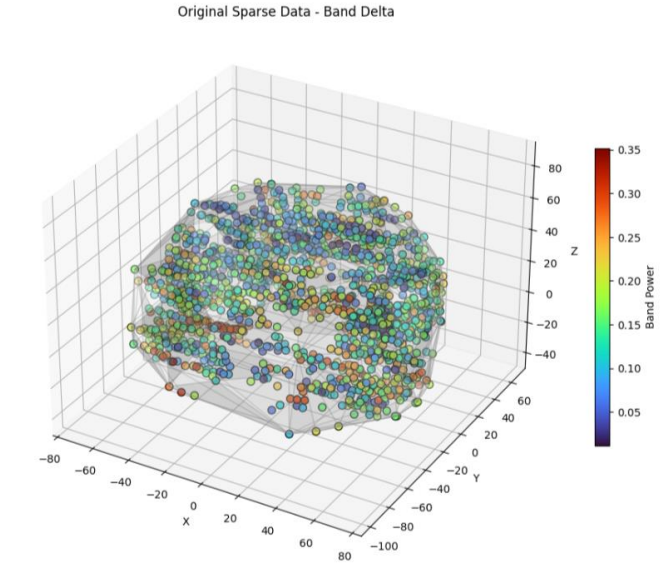


Figure 4 Visualisation of Original Sparse Data of Delta Band

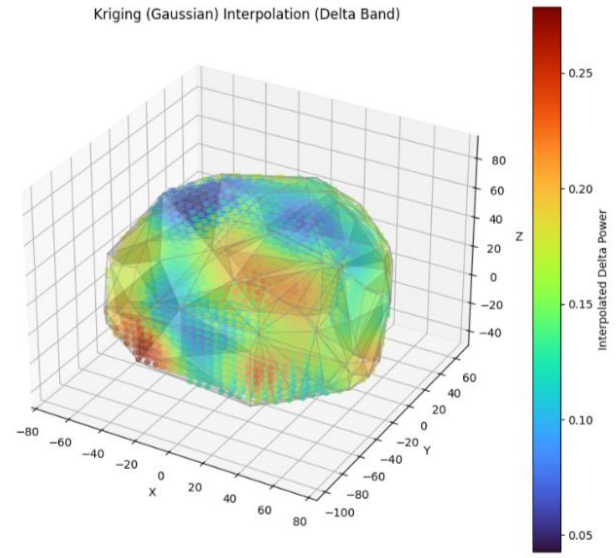


Figure 5 In-Painting visualisation using Gaussian Kriging

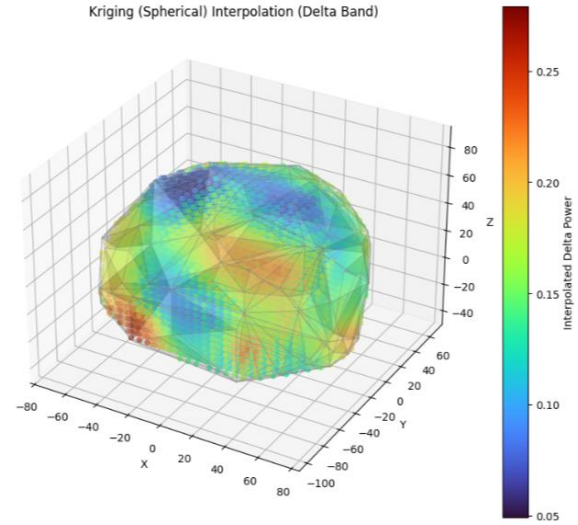


Figure 6 In-Painting Visualisation using Spherical Kriging

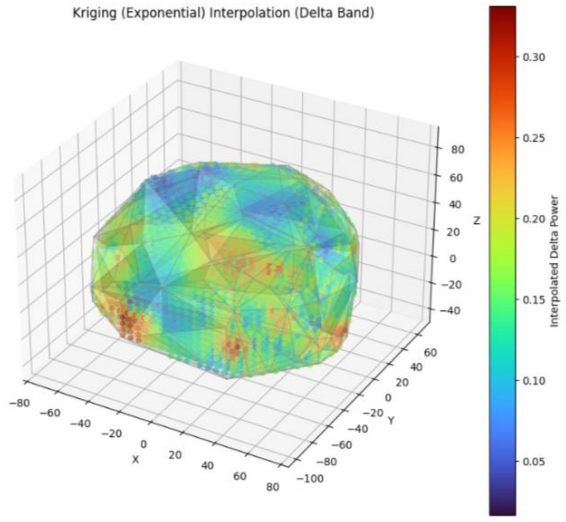


Figure 7 In-Painting Visualisation using Exponential Kriging

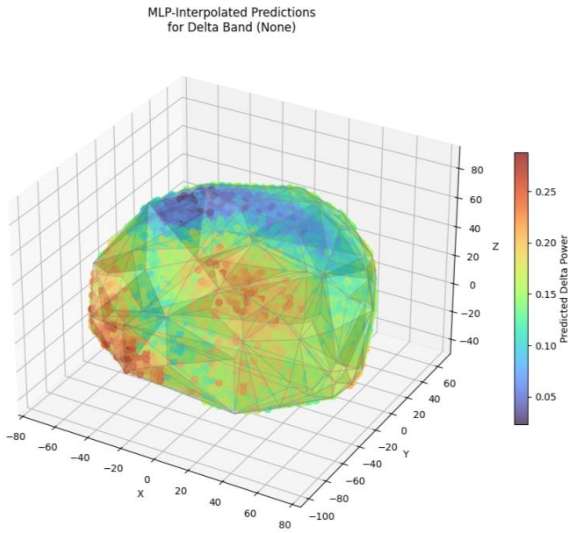


Figure 8 In-Painting Visualisation using MLP

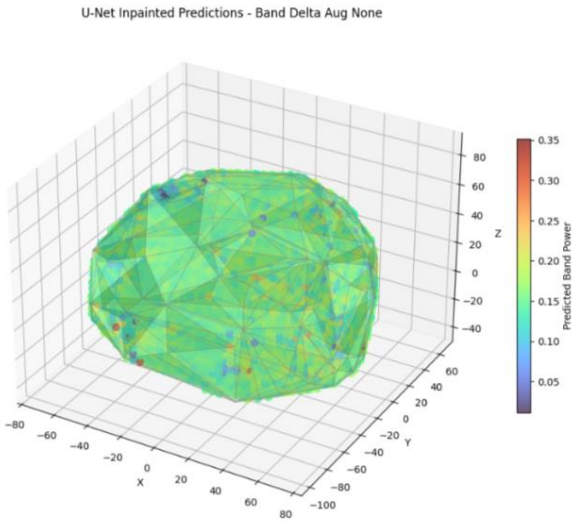


Figure 9 In-Painting Visualisation using U-Net

3.2. Interpolation Performance on Gene Expression Dataset

Due to Kriging's computational cost, only five genes were evaluated across all methods to keep consistency.

3.2.1. Quantitative Performance

As shown in Table 4, U-Net again led numerically ($R^2 = 0.878$), while MLP and Spherical Kriging achieved moderate scores (0.10-0.12). Gaussian Kriging underperformed due to instability on sparse inputs ($R^2 = -11.4$). CV metrics for the gene CHRM1 are illustrated in Figure 10.

Table 4 Average interpolation metrics across the first five gene expressions (no augmentation)

Model	R^2 (↑)	MSE (↓)	MAE (↓)	RMSE (↓)
Kriging (Gaussian)	-11.399± 22.911	11.730 ± 21.591	1.441 ±1.380	2.193 ± 2.462
Kriging (Spherical)	0.1202 ± 0.0529	0.8688 ± 0.05266	0.7253 ± 0.0239	0.9301 ± 0.0287
Kriging (Exponential)	0.0800 ± 0.0673	0.9040 ± 0.0659	0.7399 ± 0.0325	0.9489 ± 0.0356
MLP	0.101 ± 0.060	0.890 ± 0.118	0.729 ± 0.042	0.941 ± 0.062
U-Net	0.878 ± 0.028	0.120 ± 0.027	0.260 ± 0.029	0.345 ± 0.039

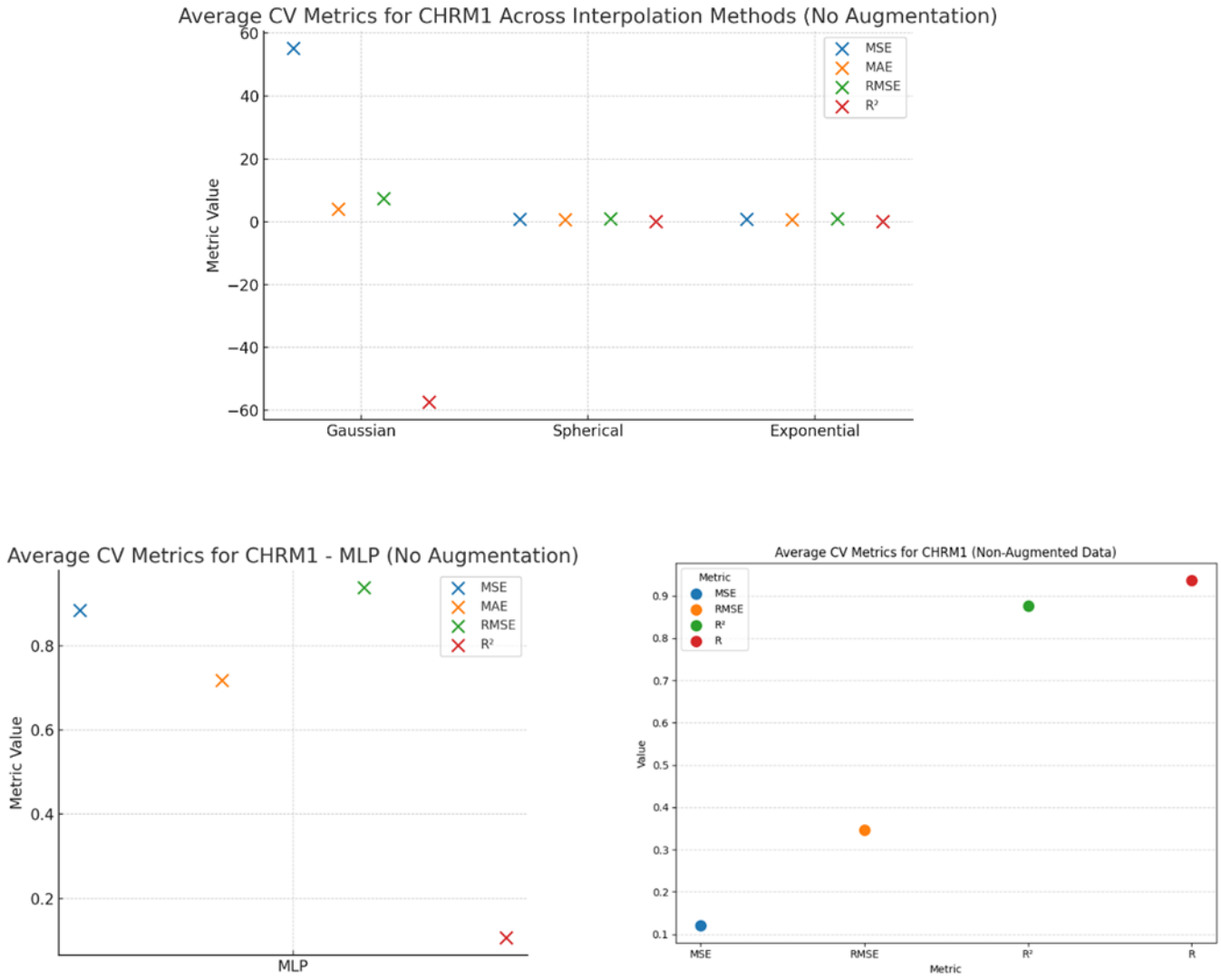


Figure 10 CHM1 Metrics Across Methods

3.2.2. Visual Performance Across Models

Visual results showed U-Net reconstructions were smooth but misaligned with known patterns. MLP and Kriging offered more spatially faithful outputs (Figures 11-15).

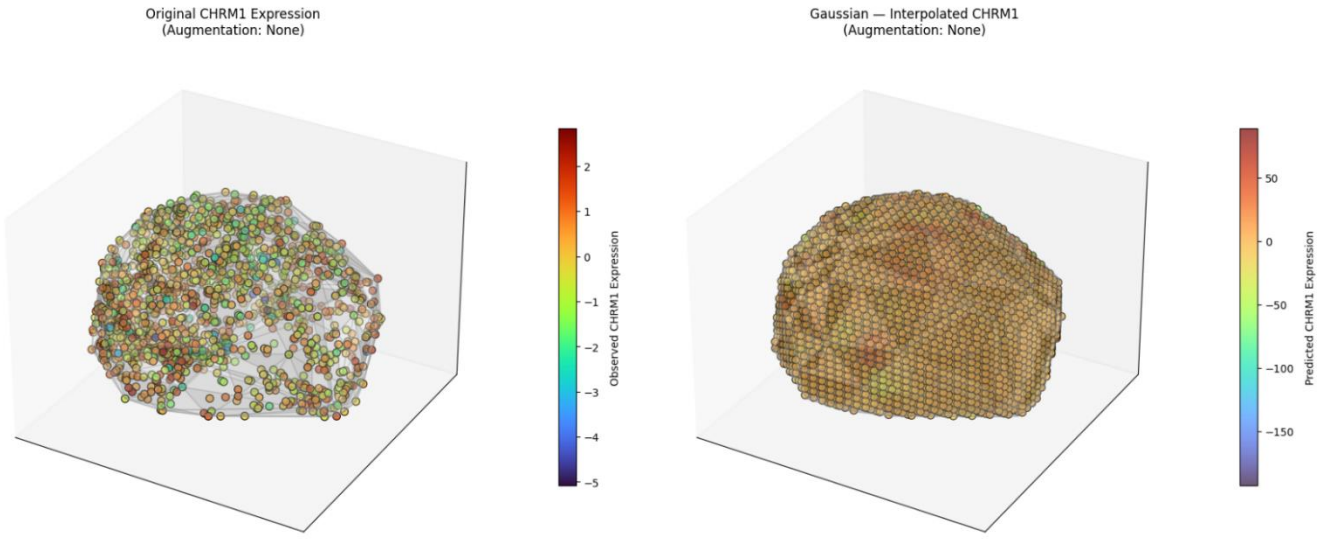


Figure 11 CHRM1 Original sparse data (left) and interpolated data using Gaussian Kriging with no augmentation

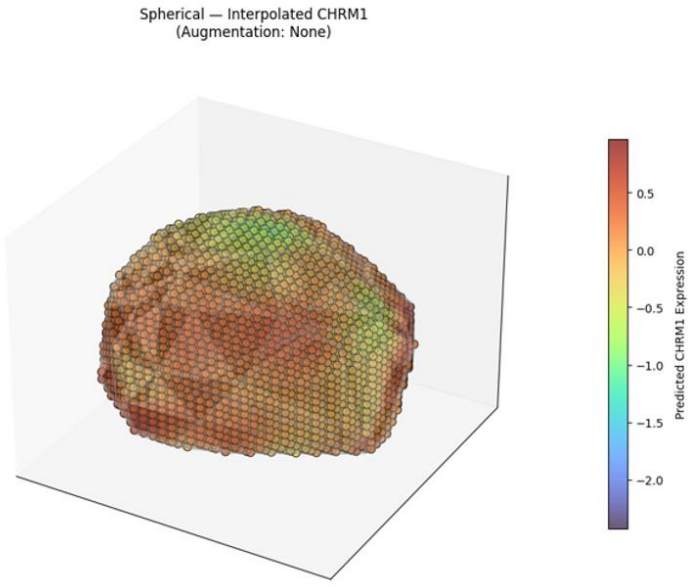


Figure 12 In-Painting Visualisation using Spherical Kriging

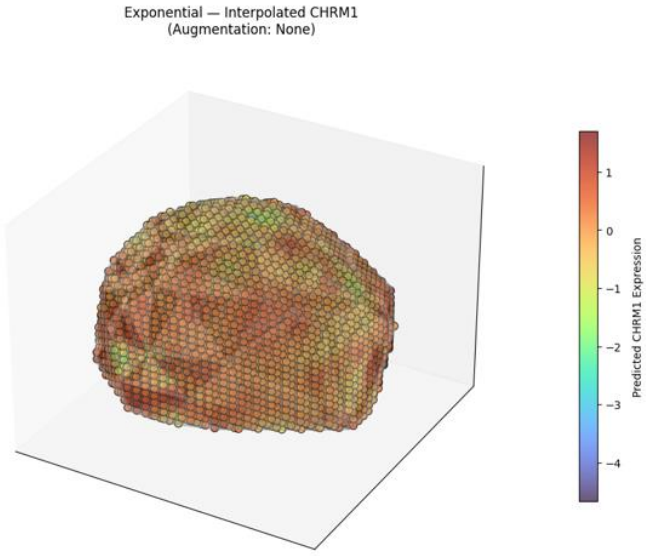


Figure 13 In-Painting Visualisation using Exponential Kriging

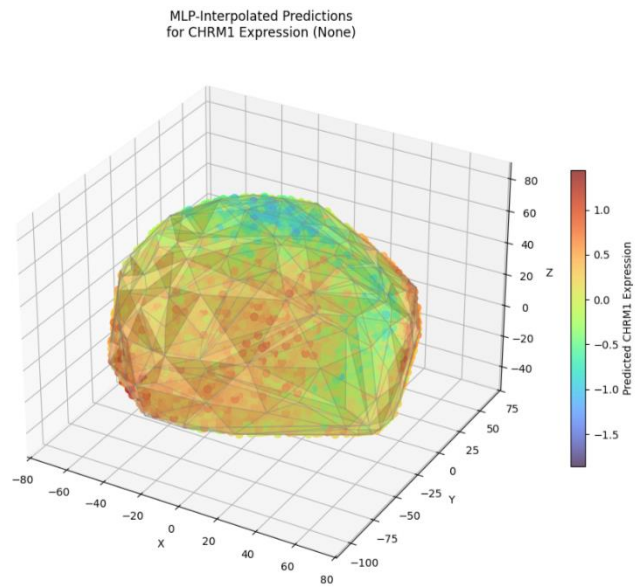


Figure 14 In-Painting Visualisation using MLP

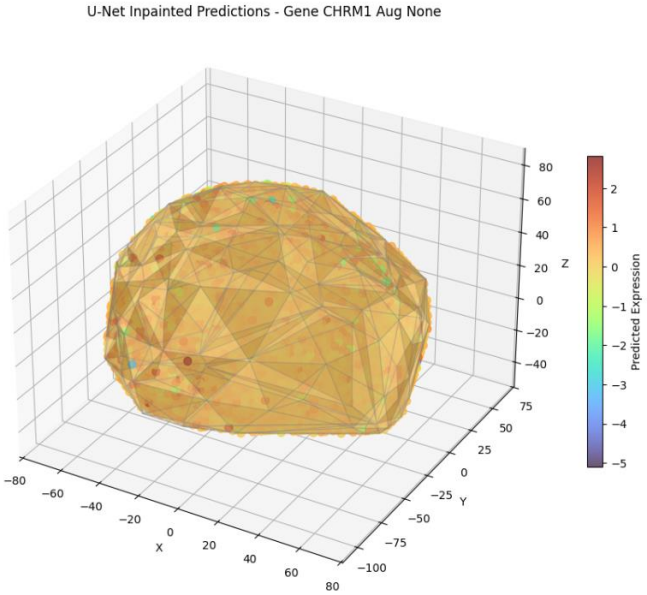


Figure 15 In-Painting Visualisation using U-Net

3.3. Data Augmentation Impact

Data augmentation strategies were applied to both datasets, sEEG and gene expression, to evaluate model robustness. Table 5 provides a representative summary of cross-validation performance on the Delta band from the sEEG dataset.

Table 5 Summary of Cross-Validation Metrics for Delta Band and All Augmentation Conditions on sEEG data

Augmentation	Model	R^2	MSE	RMSE	MAE
None	Kriging (Gaussian)	0.330	0.003	0.057	0.045
	Kriging (Spherical)	0.350	0.003	0.056	0.045
	Kriging (Exponential)	0.392	0.003	0.054	0.042
	MLP	0.200	0.004	0.062	0.045
	U-Net	0.879	0.120	0.343	0.265
Jitter	Kriging (Gaussian)	0.330	0.003	0.056	0.045
	Kriging (Spherical)	0.351	0.003	0.056	0.045
	Kriging (Exponential)	0.390	0.003	0.054	0.042
	MLP	0.212	0.004	0.061	0.049
	U-Net	0.870	0.127	0.352	0.268
Noise	Kriging (Gaussian)	0.328	0.003	0.056	0.045
	Kriging (Spherical)	0.352	0.003	0.056	0.045
	Kriging (Exponential)	0.391	0.003	0.054	0.042
	MLP	0.160	0.004	0.063	0.051
	U-Net	0.867	0.130	0.358	0.276
Jitter + Noise	Kriging (Gaussian)	0.329	0.003	0.057	0.045
	Kriging (Spherical)	0.352	0.003	0.056	0.045
	Kriging (Exponential)	0.391	0.003	0.054	0.042
	MLP	0.168	0.004	0.063	0.051
	U-Net	0.874	0.123	0.350	0.270

3.4. Computational Efficiency

U-Net trained in under 10 seconds, significantly faster than MLP (~100s) and Kriging (~150s). See Table 6 for full timing breakdown.

Table 6 Training and Validation Times

Method	Data	CV Average (s)	CV Total (s)	Final Total (s)
Gaussian Kriging	sEEG	68.6761	686.7978	155.9281
	Gene Expression	69.3055	693.0919	164.3776
Spherical Kriging	sEEG	77.7852	777.8849	155.5656
	Gene Expression	77.8717	778.7464	164.5861
Exponential Kriging	sEEG	68.3156	683.1907	156.0414
	Gene Expression	68.1892	681.9254	1633.0179
MLP	sEEG	0.2158	506.7822	103.1889
	Gene Expression	0.2149	499.3499	104.3317

U-Net	sEEG	0.0194	53.3342	9.6405
	Gene Expression	0.0195	51.1975	9.6927

3.5. Comparison with Classical Methods

As shown in Table 7 and Table 8 and Figure 16 and Figure 17, Spline interpolation performed best overall, especially on the gene dataset ($R^2 = 0.6130$), where most methods struggled. For sEEG, Exponential Kriging and Spline achieved the highest R^2 values (0.3915 and 0.3831). MLP showed balanced performance but required more time, particularly on the gene data. Kriging was the most computationally expensive methods across both datasets.

Table 7 Cross-Validation Metrics for Each Method and Dataset

Method	Data	MSE	MAE	RMSE	R^2
Exponential Kriging	**sEEG (Delta)**	**0.0029**	**0.0423**	**0.0537**	**0.3915**
	Gene (CHRM1)	0.8951	0.7215	0.9452	0.0817
Linear	sEEG (Delta)	0.0032	0.0431	0.0561	0.3341
	Gene (CHRM1)	1.0134	0.7706	1.0046	-0.0425
KNN	sEEG (Delta)	0.0031	0.0436	0.0553	0.3542
	Gene (CHRM1)	0.8964	0.7254	0.9452	0.0830
Exponential	sEEG (Delta)	0.0043	0.0497	0.0655	0.0892
	Gene (CHRM1)	1.2818	0.8526	1.1305	-0.3221
Chebyshev Polynomial	sEEG (Delta)	0.0039	0.0505	0.0621	0.1900
	Gene (CHRM1)	0.9278	0.7398	0.9610	0.0613
Spline	sEEG (Delta)	0.0008	0.0202	0.0286	0.3831
	Gene (CHRM1)	**0.4154**	**0.4621**	**0.6270**	**0.6130**
MLP	sEEG (Delta)	0.0039	0.0493	0.0493	0.1880
	Gene (CHRM1)	0.8910	0.7281	0.9341	0.1221

Table 8 Timing Summary for Each Method and Dataset

Method	Data	CV Average (s)	CV Total (s)	Final (s)
Exponential Kriging	**sEEG (Delta)**	**63.1686**	**631.6860**	**95.3289**
	Gene (CHRM1)	**61.4861**	**614.8609**	**92.6404**
Linear	sEEG (Delta)	0.0817	0.8175	0.0892
	Gene (CHRM1)	0.0704	0.7043	0.1016
KNN	sEEG (Delta)	0.0021	0.0212	0.0390
	Gene (CHRM1)	0.0022	0.0217	0.0399
Exponential	sEEG (Delta)	0.0122	0.1220	0.6982
	Gene (CHRM1)	0.0122	0.0122	0.7327
Chebyshev Polynomial	sEEG (Delta)	0.0025	0.0246	0.0079
	Gene (CHRM1)	0.0024	0.0243	0.0084
Spline	sEEG (Delta)	0.0053	0.0548	0.0152
	Gene (CHRM1)	0.2032	2.0319	0.6531
MLP	sEEG (Delta)	0.0018	0.0178	0.0175
	Gene (CHRM1)	1.713	17.1252	2.2793

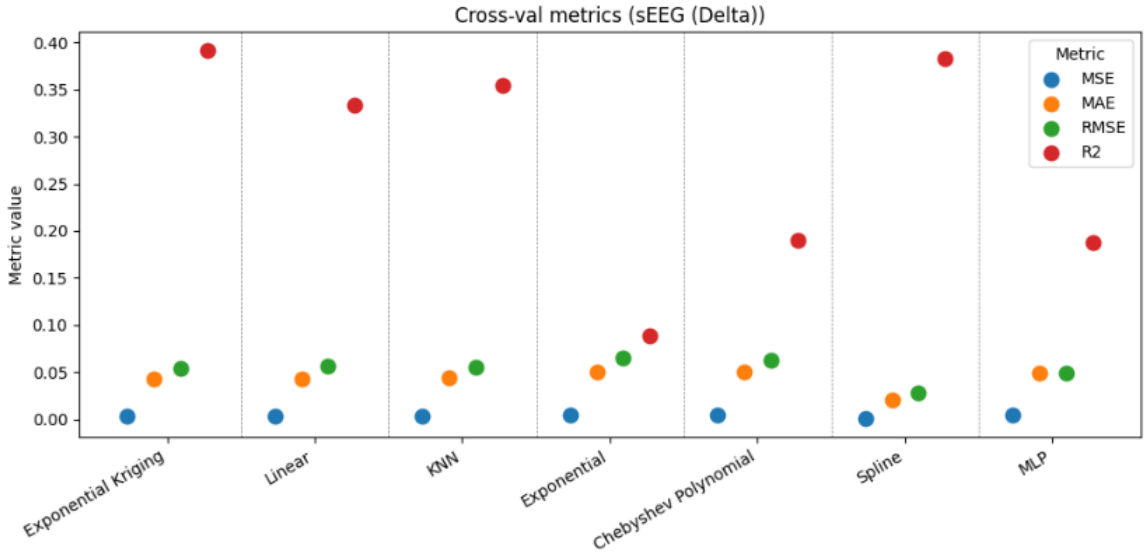


Figure 16 Plot of the Average Metrics for Each Method with no Augmentation Applied using sEEG data

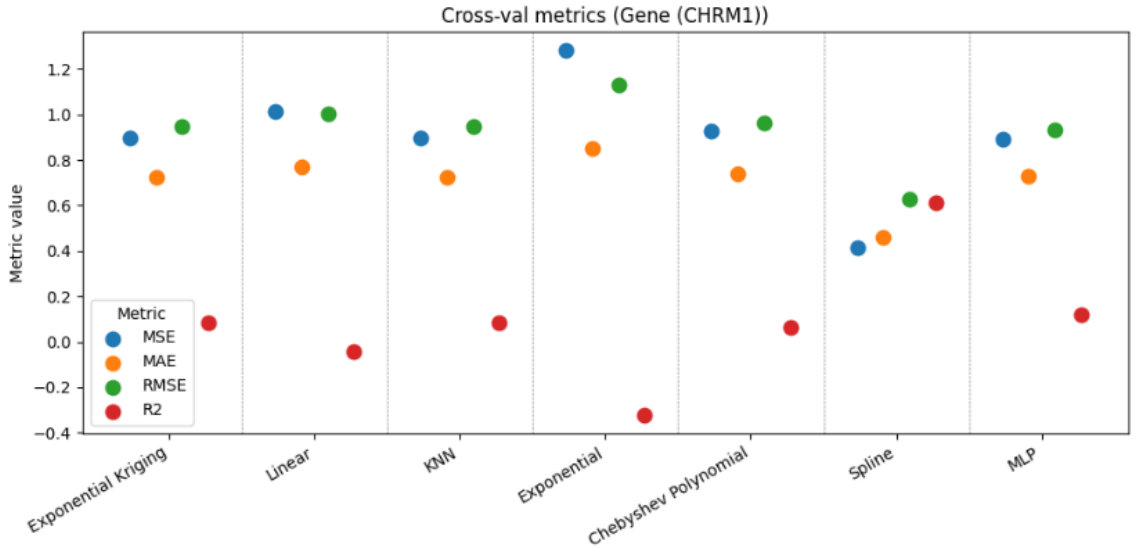


Figure 17 Plot of the Average Metrics for Each Method with no Augmentation Applied using gene expression data

4. Discussion

This study evaluated three advanced interpolation methods, Kriging, Multi-layer perceptron (MLP), and U-Net, on sparse neuroimaging datasets (sEEG and gene expression). While quantitative metrics (e.g., R^2, RMSE) suggested that U-Net was the top-performing model, a closer visual inspection revealed essential limitations that affect its practical suitability for neuroimaging tasks.

4.1. Model Comparisons

U-Net consistently achieved the highest R^2 scores across both sEEG (0.938) and gene expression data (0,878) as shown in Table 3 and Table 4, respectively. It also outperformed other methods in terms of computational efficiency, requiring less than 10 seconds for complete training and validation, as shown in Table 6. However, its visual outputs often lacked anatomical or functional plausibility, failing to reproduce region-specific signal variations in the original sparse data (As seen in Figure 15).

In contrast, Kriging and MLP produced more spatially accurate and interpretable reconstructions. Although their quantitative scores were lower, both methods better preserved regional features and variability. Kriging's performance was relatively stable across variogram models, and MLP showed slightly better alignment with ground truth than U-Net in some visualisations.

These findings suggest that U-Net may be overfitting to data patterns during training, learning to minimise error globally but failing to generalise well to unseen or irregular input spaces. This overfitting may be amplified by the sparsity of input data, which reduces the number of meaningful spatial cues available for

learning. Without sufficient regularisation or anatomical constraints, the model may over-amplify local biases in the training set, leading to smooth but biologically implausible reconstructions. This issue has also been noted in sparse medical imaging domains, where U-Net struggles without explicit spatial priors [22, 31]

Importantly, when compared with baseline interpolation techniques such as Spline (Section 3.5), we observed that classical methods can still be competitive. For example, Spline interpolation achieved an R^2 of 0.6130, as reported in Table 7, outperforming both MLP and Kriging in gene expression data. This highlights the value of benchmarking new models against simpler alternatives to avoid unnecessary model complexity when gains are marginal.

4.2. Strengths and Limitations

Kriging offers high spatial interpretability and exemplary performance in smaller datasets but comes with significant computational cost. For instance, it required several minutes to process a single fold during cross-validation, as seen in Table 8. Its reliance on distance-based correlation also limits scalability.

MLP provides a good trade-off between accuracy, efficiency, and visual reliability. Its residual architecture improved training stability, though it struggled more with gene expression data, likely due to noise and biological heterogeneity.

While efficient and quantitatively strong, U-Net lacked the granularity needed for accurate spatial reconstruction. Its design may better suit dense image-like data rather than sparse pointwise inputs without strong regularisation. Additionally, the complexity of U-Net architectures poses challenges for explainability and reproducibility, particularly in clinical workflows that demand transparency. In contrast, methods like Kriging, though computationally heavier, offer clearer spatial reasoning, making them more amenable to regulatory acceptance and quality assurance. These concerns reflect borader issues in clinical AI adoption, where intepretability and trust and important [32]

4.3. Implications

The results highlight the importance of combining quantitative metrics with spatial inspection in neuroimaging (see Section 3). A high R^2 alone may not guarantee meaningful reconstructions, especially in clinical or neuroscientific contexts where anatomical accuracy is critical.

Moreover, these findings suggest that application priorities should guide model choice: U-Net speed may be advantageous for research involving exploratory visualisation or real-time use. Kriging or MLP may be more appropriate for tasks requiring interpretability and anatomical faithfulness (e.g., surgical planning, atlas generation). From a socio-economic perspective, accurate interpolation can reduce reliance on high-density imaging, which is often expensive and inaccessible in low-resource settings. Clinically interpretable reconstructions can support more equitable diagnostics and research, especially when applied to diverse or underrepresented populations. Ensuring models perform reliably across demographic groups is essential for responsible deployment in healthcare.

5. Conclusion & Future Work

5.1. Conclusion

This study explored and compared three advanced interpolation methods, Kriging, Multi-Layer Perceptron (MLP), and U-Net, for reconstructing sparse neuroimaging data from sEEG and gene expression datasets. Quantitative results indicated that U-Net achieved the highest R^2 scores and was the most computationally efficient model. However, qualitative inspection revealed signs of overfitting, with U-Net failing to recover anatomically accurate structures in some cases.

Kriging and MLP demonstrated more spatially consistent reconstructions, preserving regional signal patterns at the cost of slightly lower numerical accuracy. Kriging provided interpretable outputs but was computationally intensive, while MLP offered a good balance between efficiency and performance.

When compared against classical interpolation methods evaluated by other group members, Spline interpolation achieved the best performance on both datasets. These results highlight that simple, non-parametric approaches can still outperform complex models in specific sparse settings, especially when the data structure aligns well with their underlying assumptions.

These findings underscore the need to evaluate interpolation methods through error metrics, spatial consistency, and practical applicability, especially in biomedical domains where interpretability is essential. However, over-reliance on interpolated maps without understanding the limitations of the input data poses risks, including anatomical inaccuracy and demographic bias. This aligns with recent calls for more rigorous evaluation of data-driven methods in health contexts, especially concerning demographic fairness and responsible deployment [33]. To ensure safe clinical use, caution must be taken when using outputs in downstream analyses or decision-making.

Looking ahead, integrating robust interpolation tools into neurosurgical workflows, brain atlas construction, or clinical decision-support systems could accelerate precision medicine and improve access to high-quality neuroimaging insights, particularly in data-scarce environments.

5.2. Future Work

There are several promising directions for future research. One avenue is the development of hybrid models that combine the spatial interpretability of Kriging with the learning flexibility of neural networks [34]. Improving regularisation strategies in U-Net, through dropout, spatial priors, or attention masking, may also help mitigate overfitting and enhance anatomical accuracy.

Graph-based ML models like Graph Neural Networks (GNNs) [35, 36] could offer a more natural framework for handling sparse, irregular neuroimaging data, particularly for gene expression. Expanding gene evaluations to include more diverse genetic markers and biological priors could further improve robustness. Finally, validating these methods on downstream tasks, such as brain region classification or connectivity modelling, would provide a more practical measure of their utility in neuroscientific workflows and clinical pipelines.

6. References

[1] P. Kahane, E. Landré, L. Minotti, S. Francione, and P. Ryvlin, "The Bancaud and Talairach view on the epileptogenic zone: a working hypothesis," (in eng), *Epileptic Disord,* vol. 8 Suppl 2, pp. S16-26, Aug 2006.

[2] E. H. Shen, C. C. Overly, and A. R. Jones, "The Allen Human Brain Atlas: comprehensive gene expression mapping of the human brain," (in eng), *Trends Neurosci,* vol. 35, no. 12, pp. 711-4, Dec 2012, doi: 10.1016/j.tins.2012.09.005.

[3] N. Battalgazy, R. Valenta, P. Gow, C. Spier, and G. Forbes, "Addressing Geological Challenges in Mineral Resource Estimation: A Comparative Study of Deep Learning and Traditional Techniques," *Minerals*, vol. 13, no. 7, doi: 10.3390/min13070982.

[4] C. Li *et al.*, "Spatio-temporal MLP network for seizure prediction using EEG signals," *Measurement,* vol. 206, p. 112278, 2023/01/01/ 2023, doi: https://doi.org/10.1016/j.measurement.2022.112278.

[5] R. Yousef *et al.*, "U-Net-Based Models towards Optimal MR Brain Image Segmentation," (in eng), *Diagnostics (Basel),* vol. 13, no. 9, May 4 2023, doi: 10.3390/diagnostics13091624.

[6] M. E. Raichle, "A brief history of human brain mapping," (in eng), *Trends Neurosci,* vol. 32, no. 2, pp. 118-26, Feb 2009, doi: 10.1016/j.tins.2008.11.001.

[7] B. Frauscher *et al.*, "Atlas of the normal intracranial electroencephalogram: neurophysiological awake activity in different cortical areas," (in eng), *Brain,* vol. 141, no. 4, pp. 1130-1144, Apr 1 2018, doi: 10.1093/brain/awy035.

[8] H. Zhu, T. Li, and B. Zhao, "Statistical Learning Methods for Neuroimaging Data Analysis with Applications," (in eng), *Annu Rev Biomed Data Sci,* vol. 6, pp. 73-104, Aug 10 2023, doi: 10.1146/annurev-biodatasci-020722-100353.

[9] L. French and P. Pavlidis, "Relationships between gene expression and brain wiring in the adult rodent brain," (in eng), *PLoS Comput Biol,* vol. 7, no. 1, p. e1001049, Jan 6 2011, doi: 10.1371/journal.pcbi.1001049.

[10] N. Koide-Majima, S. Nishimoto, and K. Majima, "Mental image reconstruction from human brain activity: Neural decoding of mental imagery via deep neural network-based Bayesian estimation," *Neural Networks,* vol. 170, pp. 349-363, 2024/02/01/ 2024, doi: https://doi.org/10.1016/j.neunet.2023.11.024.

[11] E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," *Nature Medicine,* vol. 25, no. 1, pp. 44-56, 2019/01/01 2019, doi: 10.1038/s41591-018-0300-7.

[12] P. K. Kitanidis, *Introduction to Geostatistics: Applications in Hydrogeology*. Cambridge: Cambridge University Press, 1997.

[13] GeoStat-Framework. "pykrige.ok3d.OrdinaryKriging3D - PyKrige 1.7.0 documentation." https://geostat-framework.readthedocs.io/projects/pykrige/en/stable/generated/pykrige.ok3d.OrdinaryKriging3D.html (accessed May 14, 2025).

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *arXiv,* vol. 1512.03385, 2015. [Online]. Available: https://arxiv.org/abs/1512.03385.

[15] B. R. A. Rahimi, "Random features for large-scale kernel machines," presented at the Advances in Neural Information Processing Systems 20 (NeurIPS 2007), 2007. [Online]. Available: https://papers.nips.cc/paper_files/paper/2007/file/013a006f03dbc5392effeb8f18fda755-Paper.pdf.

[16] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," *arXiv preprint arXiv:1502.03167,* 2015. [Online]. Available: https://arxiv.org/pdf/1502.03167.

[17] G. H. N. Srivastava, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *Journal of Machine Learning Research,* vol. 15, pp. 1929-1958, 2014. [Online]. Available: https://www.cs.toronto.edu/~rsalakhu/papers/srivastava14a.pdf.

[18] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical Evaluation of Rectified Activations in Convolutional Network," *arXiv preprint arXiv:1505.00853,* 2015. [Online]. Available: https://arxiv.org/pdf/1505.00853.

[19] R. Girshick, "Fast R-CNN," *arXiv preprint arXiv:1504.08083,* 2015. [Online]. Available: https://arxiv.org/pdf/1504.08083.

[20] D. P. Kingma and J. L. Ba, "ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION," *arXiv preprint arXiv:1412.6980,* 2015. [Online]. Available: https://arxiv.org/pdf/1412.6980.

[21] A. Paszke, "Pytorch: An imperative style, high-performance deep learning library," *arXiv preprint arXiv:1912.01703,* 2019.

[22] L. Liu and Y. Liu, "Load image inpainting: An improved U-Net based load missing data recovery method," *Applied Energy,* vol. 327, p. 119988, 2022/12/01/ 2022, doi: https://doi.org/10.1016/j.apenergy.2022.119988.

[23] S. P. Woo, Jongchan; Lee, Joon-Young; Kweon, In So, "CBAM: Convolutional Block Attention Module," *arXiv preprint,* vol. arXiv:1807.06521, 2018. [Online]. Available: https://arxiv.org/abs/1807.06521.

[24] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena,* vol. 60, no. 1, pp. 259-268, 1992/11/01/ 1992, doi: https://doi.org/10.1016/0167-2789(92)90242-F.

[25] K. Kamnitsas *et al.*, "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Medical Image Analysis,* vol. 36, pp. 61-78, 2017/02/01/ 2017, doi: https://doi.org/10.1016/j.media.2016.10.004.

[26] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *Journal of Big Data,* vol. 6, no. 1, p. 60, 2019/07/06 2019, doi: 10.1186/s40537-019-0197-0.

[27] P. Chlap, H. Min, N. Vandenberg, J. Dowling, L. Holloway, and A. Haworth, "A review of medical image data augmentation techniques for deep learning applications," (in eng), *J Med Imaging Radiat Oncol,* vol. 65, no. 5, pp. 545-563, Aug 2021, doi: 10.1111/1754-9485.13261.

[28] G. Varoquaux, "Cross-validation failure: Small sample sizes lead to large error bars," *NeuroImage,* vol. 180, pp. 68-77, 2018/10/15/ 2018, doi: https://doi.org/10.1016/j.neuroimage.2017.06.061.

[29] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature," *Geosci. Model Dev.,* vol. 7, no. 3, pp. 1247-1250, 2014, doi: 10.5194/gmd-7-1247-2014.

[30] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," (in eng), *PeerJ Comput Sci,* vol. 7, p. e623, 2021, doi: 10.7717/peerj-cs.623.

[31] Y. Han and J. C. Ye, "Framing U-Net via deep convolutional framelets: Application to sparse-view CT," *IEEE transactions on medical imaging,* vol. 37, no. 6, pp. 1418-1429, 2018.

[32] S. Tonekaboni, S. Joshi, M. D. McCradden, and A. Goldenberg, "What clinicians want: contextualizing explainable machine learning for clinical end use," in *Machine learning for healthcare conference*, 2019: PMLR, pp. 359-380.

[33] I. Y. Chen, E. Pierson, S. Rose, S. Joshi, K. Ferryman, and M. Ghassemi, "Ethical Machine Learning in Healthcare," (in eng), *Annu Rev Biomed Data Sci,* vol. 4, pp. 123-144, Jul 2021, doi: 10.1146/annurev-biodatasci-092820-114757.

[34] M. Mohammadpour, H. Roshan, M. Arashpour, and H. Masoumi, "Machine learning assisted Kriging to capture spatial variability in petrophysical property modelling," *Marine and Petroleum Geology,* vol. 167, p. 106967, 2024/09/01/ 2024, doi: https://doi.org/10.1016/j.marpetgeo.2024.106967.

[35] T. Liu, Z.-Y. Fang, Z. Zhang, Y. Yu, M. Li, and M.-Z. Yin, "A comprehensive overview of graph neural network-based approaches to clustering for spatial transcriptomics," *Computational and Structural Biotechnology Journal,* vol. 23, pp. 106-128, 2024/12/01/ 2024, doi: https://doi.org/10.1016/j.csbj.2023.11.055.

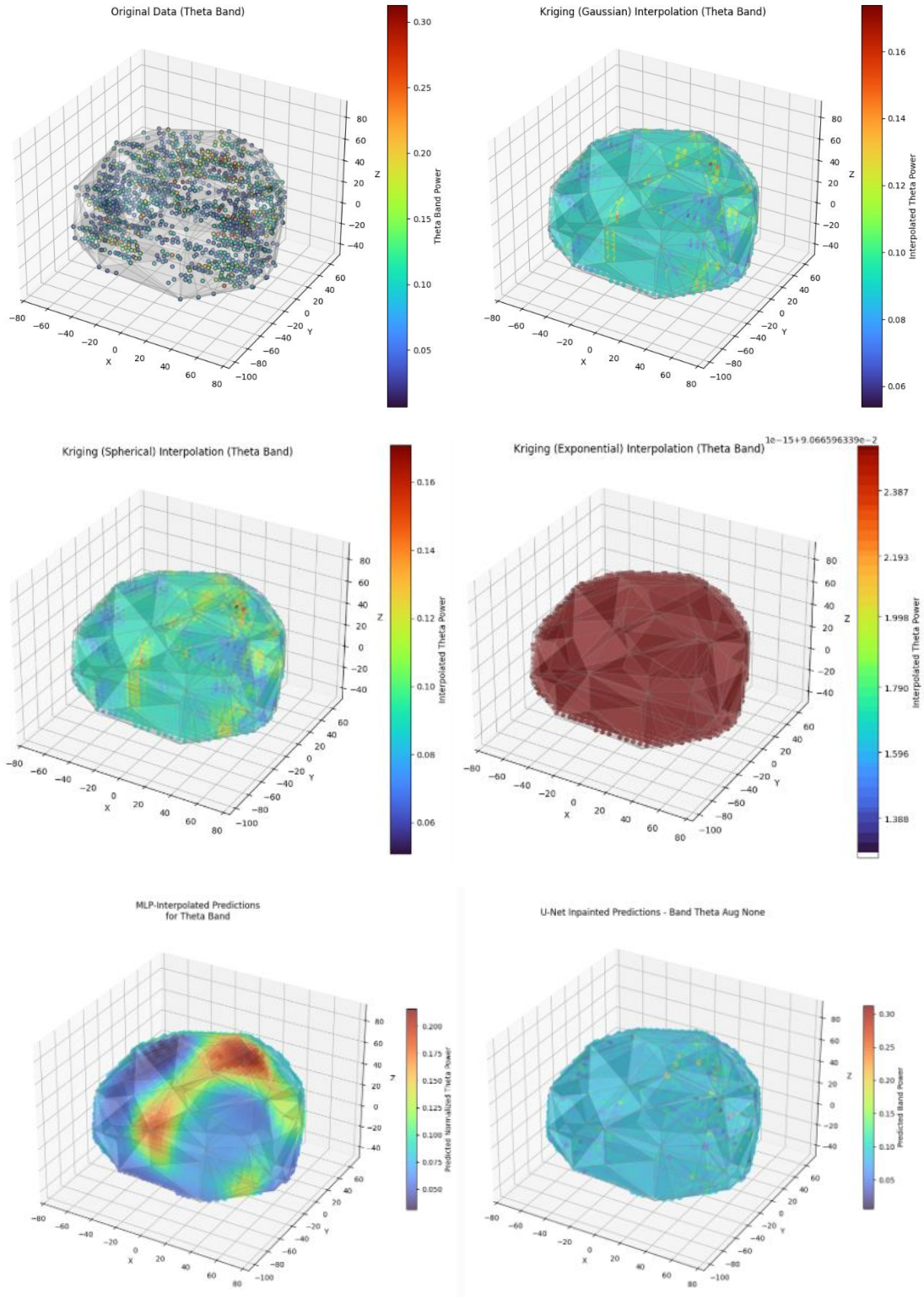
[36] L. Wang, K. Li, and X. P. Hu, "Graph convolutional network for fMRI analysis based on connectivity neighborhood," (in eng), *Netw Neurosci,* vol. 5, no. 1, pp. 83-95, 2021, doi: 10.1162/netn_a_00171.

Appendix A – Visualisation Across all Frequency Bands

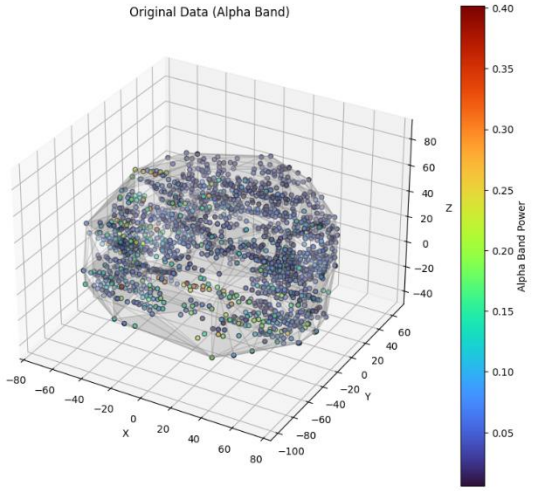
This appendix presents qualitative comparisons of Kriging (Gaussian, Spherical, Exponential), MLP and U-Net interpolation results across four frequency bands (Theta, Alpha, Beta, Gamma). Note that Delta Band is shown in the main text. The original sparse data is shown for reference.

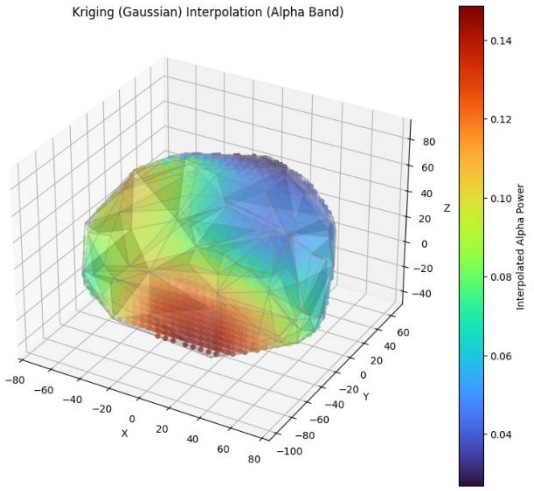
Each subfigure displays the interpolated power distribution on a 3D brain surface mesh for a given frequency band and variogram. These visualisations help assess the anatomical plausibility and smoothness of reconstructions.

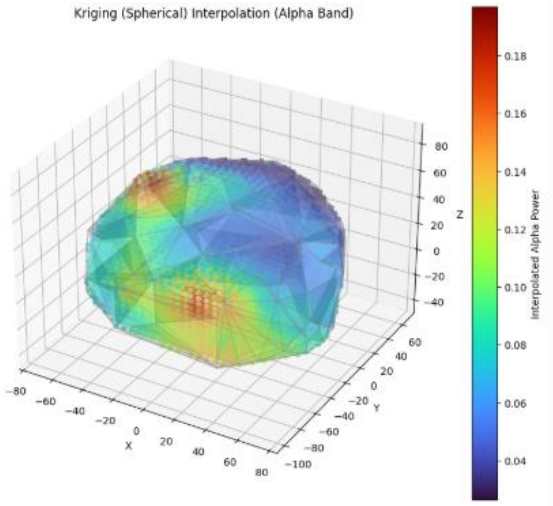
Theta Band:

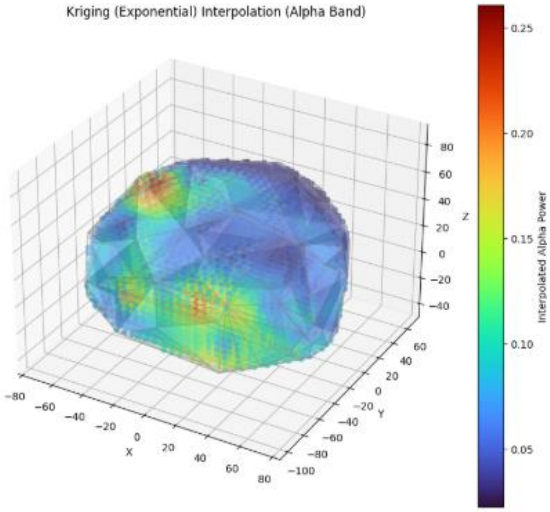


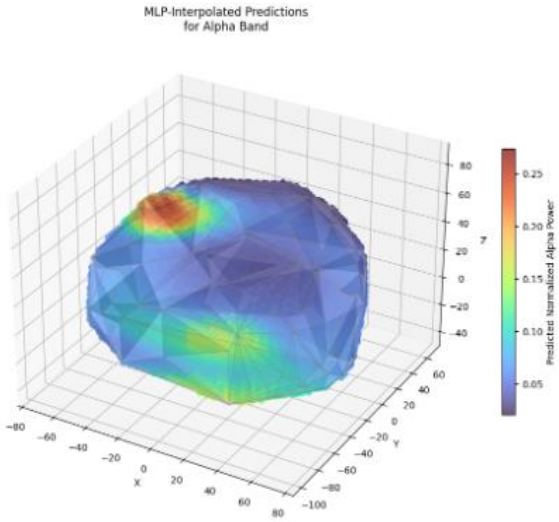
Alpha Band:

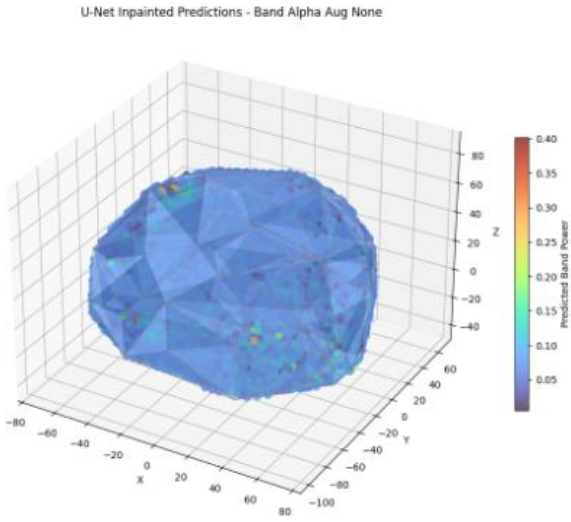

Original Data (Alpha Band)
Alpha Band Power
X
Y
Z


Kriging (Gaussian) Interpolation (Alpha Band)
Interpolated Alpha Power
X
Y
Z

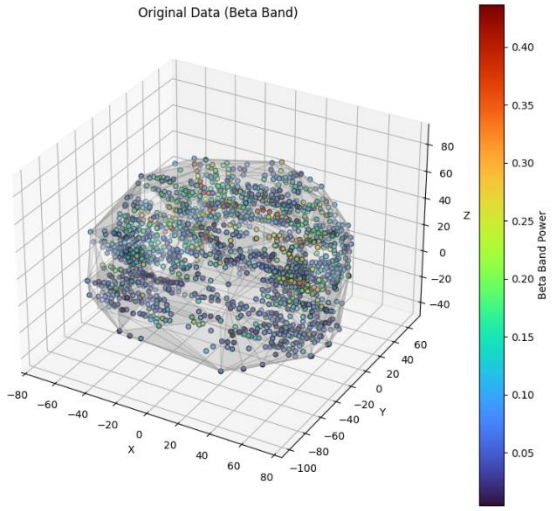

Kriging (Spherical) Interpolation (Alpha Band)
Interpolated Alpha Power
X
Y
Z

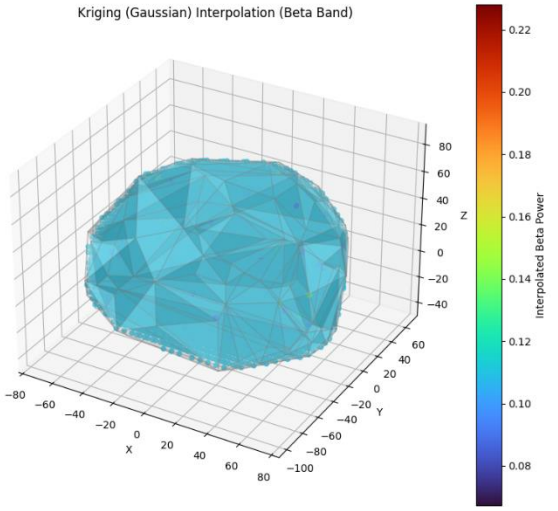

Kriging (Exponential) Interpolation (Alpha Band)
Interpolated Alpha Power
X
Y
Z

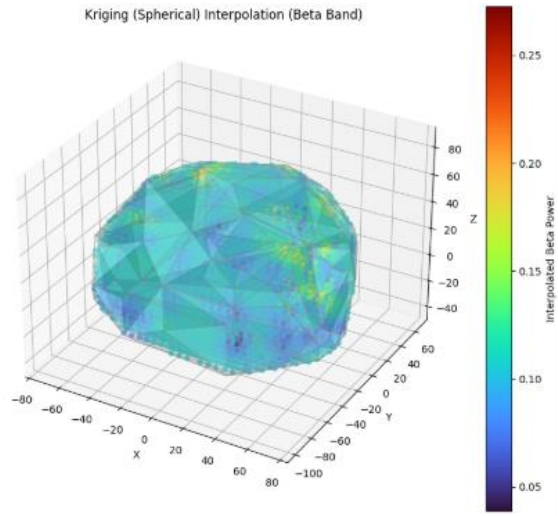

MLP-Interpolated Predictions
for Alpha Band
Predicted Normalized Alpha Power
X
Y
Z

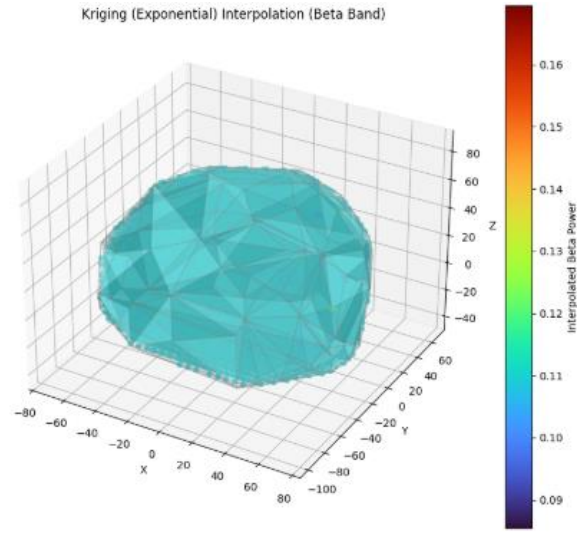

U-Net Inpainted Predictions - Band Alpha Aug None
Predicted Band Power
X
Y
Z

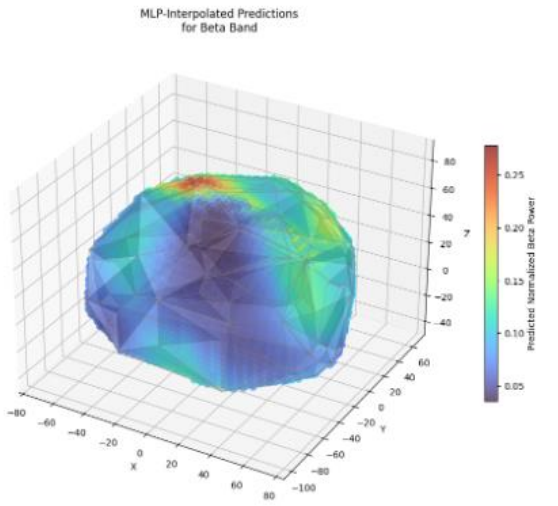
Beta Band:

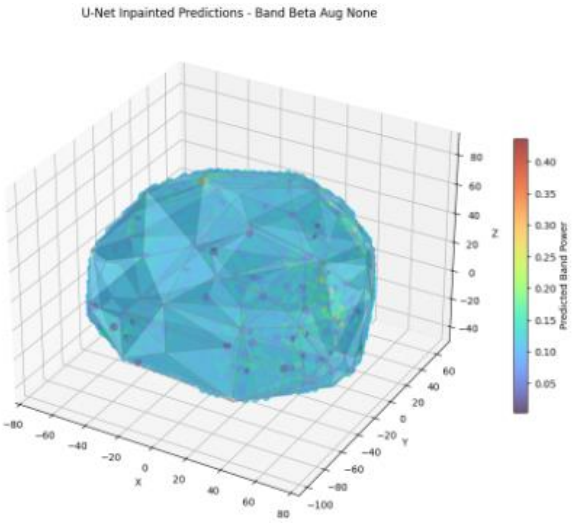

Original Data (Beta Band)
Beta Band Power
X
Y
Z


Kriging (Gaussian) Interpolation (Beta Band)
Interpolated Beta Power
X
Y
Z

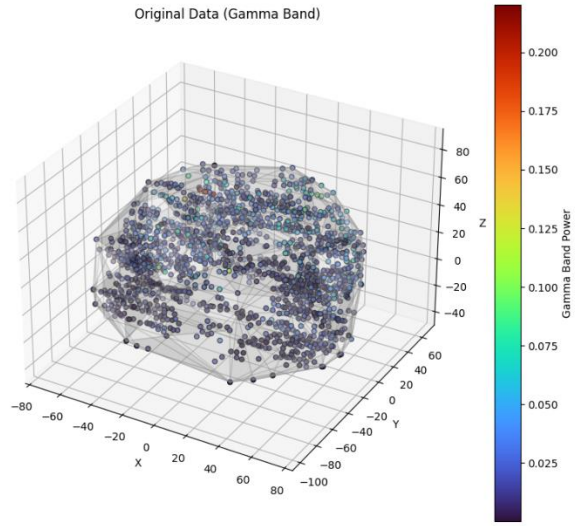

Kriging (Spherical) Interpolation (Beta Band)
Interpolated Beta Power
X
Y
Z

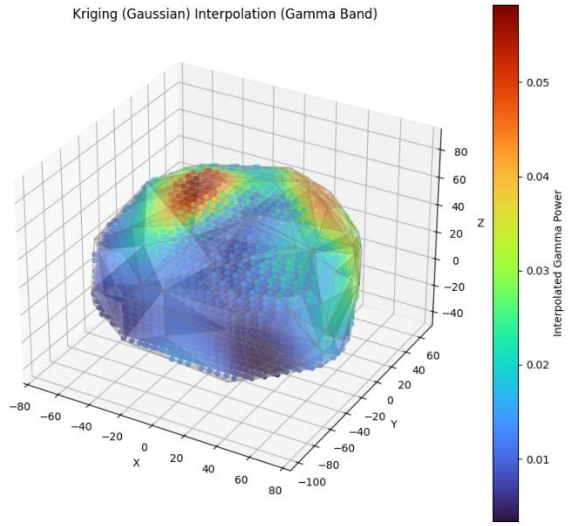

Kriging (Exponential) Interpolation (Beta Band)
Interpolated Beta Power
X
Y
Z

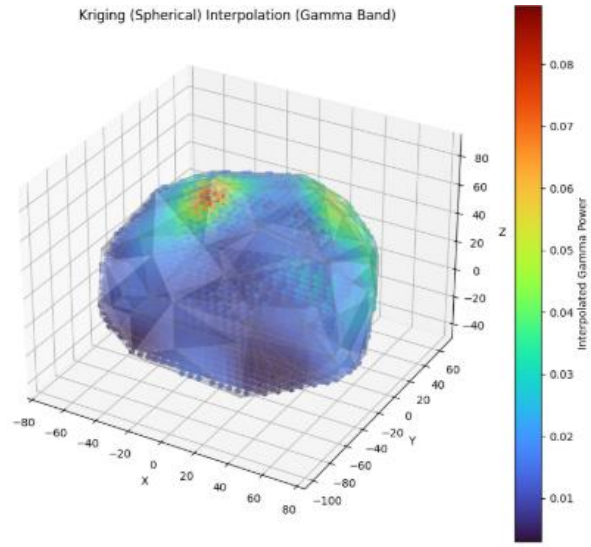

MLP-Interpolated Predictions
for Beta Band
Predicted Normalized Beta Power
X
Y
Z

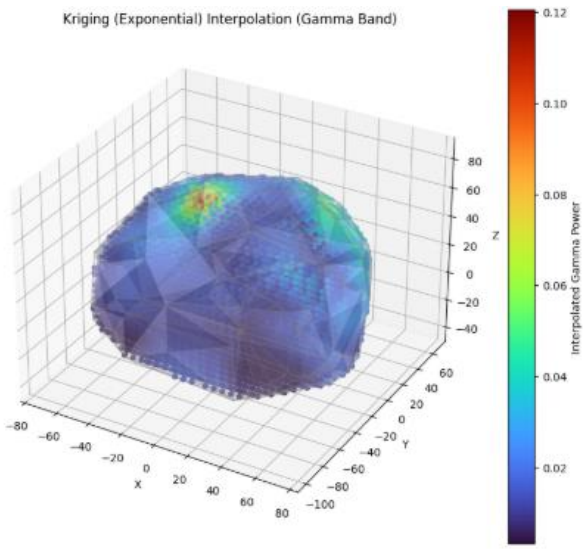

U-Net Inpainted Predictions - Band Beta Aug None
Predicted Band Power
X
Y
Z

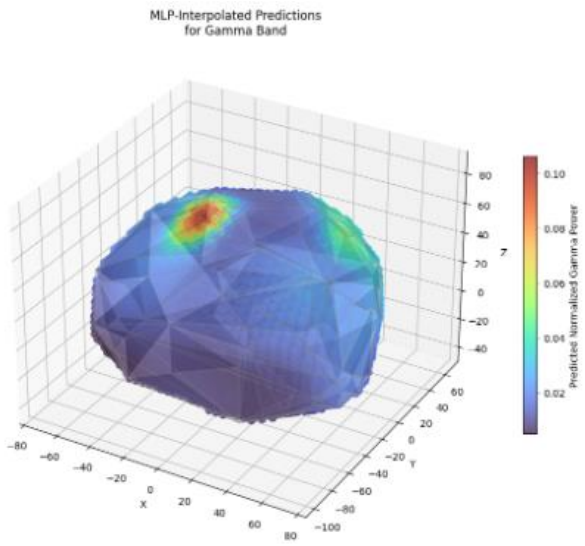
Gamma Band:

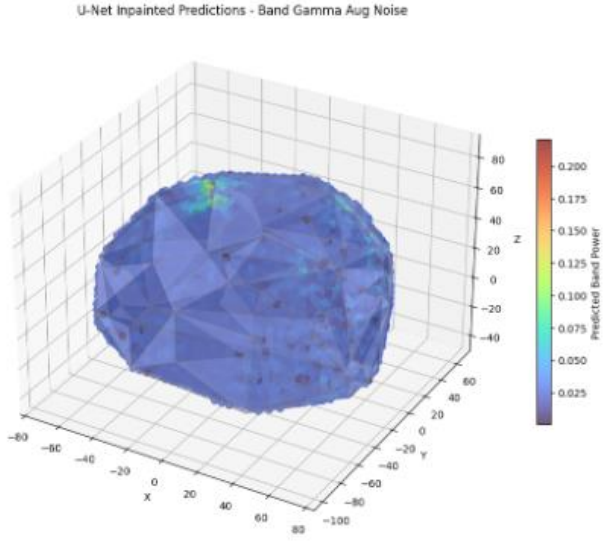

Original Data (Gamma Band)
Gamma Band Power
X
Y
Z


Kriging (Gaussian) Interpolation (Gamma Band)
Interpolated Gamma Power
X
Y
Z


Kriging (Spherical) Interpolation (Gamma Band)
Interpolated Gamma Power
X
Y
Z


Kriging (Exponential) Interpolation (Gamma Band)
Interpolated Gamma Power
X
Y
Z


MLP-Interpolated Predictions
for Gamma Band
Predicted Normalized Gamma Power
X
Y
Z


U-Net Inpainted Predictions - Band Gamma Aug Noise
Predicted Band Power
X
Y
Z