Spatio-Temporal Forecasting of PS–InSAR Displacement with a PointNet-Inspired Deep Learning Model

Takayuki Shinohara
shinohara.takayuki@aist.go.jp
Hidetaka Saomoto
h-saomoto@aist.go.jp

National Institute of Advanced Industrial Science and Technology
Tsukuba, Japan

Abstract

Persistent Scatterer InSAR (PS–InSAR) yields a genuine three–dimensional point cloud: each scatterer is identified by fixed coordinates (x, y, z) and an accompanying displacement sequence $\Delta u_1 \dots \Delta u_T$. Most existing forecasting studies treat every series in isolation and, as a result, discard the spatial context that governs tectonic, volcanic, and anthropogenic deformation. We present **PointNet–PSI**, a spatio–temporal model that couples a PointNet–style point cloud encoder with MOMENT, a recent foundation model for general time–series prediction. The permutation–invariant PointNet front–end ingests the unordered PS–InSAR cloud, compresses local geometry and kinematic similarity into latent descriptors, then concatenates these descriptors with the raw displacement history. The enriched embeddings are passed to MOMENT's transformer backbone, which produces multi–step forecasts for every scatterer. In this hybrid design the network learns *where* through spatial aggregation of neighbouring points and *when* through MOMENT's long–range temporal attention, while retaining the large receptive field and data–efficient pre–training advantages of the base model. We validate the approach on the *European Ground Motion Service Basic 2019–2023* vertical–velocity product. We adopt a hindcast protocol: observations from 20192020 serve as context, and all 60 samples of 2021 form the strictly held-out forecast horizon. Compared with strong per–point sequence models (LSTM, Temporal Fusion Transformer, and vanilla MOMENT) and naive PointNet, PointNet–PSI reduces the test RMSE by about 17%.

1 Introduction

Interferometric Synthetic Aperture Radar (InSAR) has revolutionised large-scale geodetic monitoring by providing millimetre-precision measurements of surface displacement irrespective of daylight or cloud cover. Among the available processing chains, *Persistent Scatterer* InSAR (PS–InSAR) [9, 15] extracts coherent phase histories for radar-bright targets that remain stable over hundreds of satellite passes. Typical European constellations deliver $\mathcal{O}(10^6)$ scatterers per orbit, each characterised by fixed spatial coordinates (x, y, z) and a displacement time-series $\Delta u_1, \dots, \Delta u_T$. Recent public releases such as the *European Ground Motion Service (EGMS) Basic 2019–2023* product exposes continent-wide ground-motion

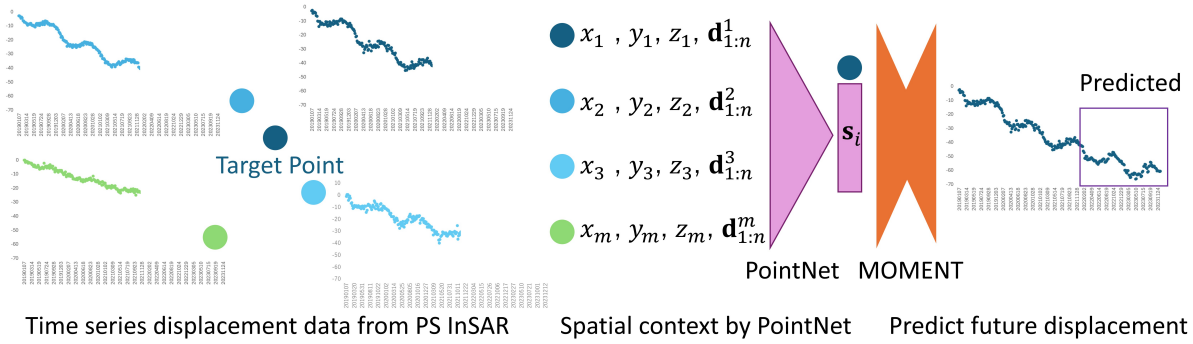

Figure 1: **Overview of our Method.** Given a target PS–InSAR displacement series and those of its nearby scatterers, PointNet [27] distils the neighbours into a permutation–invariant spatial latent vector $\mathbf{s}_i$. This spatial context is fused with the target history and fed to a fine-tuned time-series foundation model called MOMENT [12], enabling joint spatial-temporal reasoning for multi-step ground-motion forecasting.

archives containing more than twenty million trajectories, enabling unprecedented insight into tectonic creep, volcanic unrest, groundwater draw-down and anthropogenic subsidence. For critical infrastructure and densely populated basins the next logical step is *forecasting*: predicting how these displacements will evolve in the near future to enable early-warning systems and risk-informed urban-planning.

The majority of PS–InSAR forecasting studies feed each displacement series into an independent LSTM or GRU [11, 36, 50, 51], achieving reasonable short-term accuracy but *discarding* the spatial correlations that govern many geophysical processes. Consequently, the predicted fields often appear spatially noisy, violating the physical continuity expected along fault planes or within subsidence bowls. For Small Baseline Subset (SBAS) imagery, which is sampled on a regular grid, spatio-temporal Transformer hybrids have recently proven effective [52], and the Koopman operator [19]-inspired Auto Encoder [31] is also effective for forecasting future displacement of each pixel of SBAS time series image [30]. In contrast, PS–InSAR forms an *unordered* point cloud, rendering pixel-based encoders inapplicable. A natural candidate for such unstructured data is PointNet [26], whose permutation-invariant multilayer perceptrons can learn geometry-aware descriptors directly from 3-D point clouds. Point-cloud networks have already demonstrated superior spatial feature extraction of time series data for per-point full-waveform LiDAR classification tasks [32, 33, 34], suggesting that analogous gains are attainable for PS–InSAR forecasting once spatial context is properly exploited.

The temporal dimension of PS–InSAR forecasting has reaped the benefits of the rapid evolution of sequence modelling. Recurrent networks have largely been superseded by efficient Transformer variants, and, most importantly, large-scale pre-training on hundreds of millions of heterogeneous series has yielded a new class of *foundation models*. Leveraging hierarchical attention, masked-reconstruction objectives, and autoregressive transfer from large language models, these models now define the state of the art in downstream tasks such as forecasting, anomaly detection, and classification. However, input representation for foundation models is confined to a tensor of shape [channels$\times$time]. When deployed on PS–InSAR data, they must still process each displacement series independently, thereby discarding the spatial correlations that govern crustal deformation.

We address this limitation with **PointNet-PSI**, the first architecture that unifies a PointNet-

style point cloud encoder with the Transformer backbone of MOMENT [12]. The permutation-invariant PointNet front end ingests the unordered PS–InSAR point cloud, encodes local geometry and kinematic similarity into compact latent vectors, and concatenates these descriptors with every targets displacement history. The enriched embeddings are then propagated through MOMENTs hierarchical attention stack, enabling the network to learn where through spatial context and when through long-range temporal dynamics, while preserving the data-efficient pre-training and billion-sample scalability of the base model.

2 Related Study

2.1 Time–series Prediction

RNN/LSTM/Transformer. Forecasting dynamical systems with neural networks has been investigated for decades [3, 4]. Long Short Term Memory (LSTM) [14] and Gated Recurrent Units (GRU) [7] improved upon vanilla RNNs, but still suffer from vanishing/exploding gradients [5]. Mitigation strategies span stability analysis [24], unitary weight matrices [2], and antisymmetric parameterisations [6], often trading expressive power for numerical robustness [18]. Physics guided RNNs [16], Hamiltonian neural networks [13], and neural ODE variants embed conservation laws or differential equation structure to improve interpretability and sample efficiency, yet remain difficult to scale beyond modest sequence lengths. Transformer-based forecasters [25, 55, 56] address long-range dependencies through sparse or patchwise attention, but incur quadratic memory in input length and treat each channel independently, thereby ignoring rich spatial context present in geospatial data products like PS–InSAR.

Foundation Models. A new generation of large pre-trained time series foundation models has recently emerged. Chronos [1] employs masked reconstruction and contrastive objectives to learn universal temporal representations, while Lag-Llama (ServiceNow) adapts LLaMA weights to autoregressive forecasting. Prophet [39] remains a strong classical baseline that decomposes series into trend, seasonality and holiday effects. MOMENT [12] scales transformer forecasting to the billion-sample regime via hierarchical attention; Moirai [43] hybridises diffusion priors with causal convolution heads; TimeGPT-1 [10] and TimesFM [8] further push parameter counts and training corpora, delivering competitive zero-shot forecasts across hundreds of benchmarks. Despite their versatility, all of these models ingest sequences of the form [channels $\times$ time] and therefore cannot exploit the permutation-invariant *point cloud* nature of PS–InSAR. Our work closes this gap by augmenting MOMENT with a PointNet–style spatial encoder, enabling simultaneous learning of spatial and temporal dependencies within a single foundation model framework.

2.2 Deep learning for 3D Point Clouds

Projection-based approaches. Early attempts to reuse mature image and voxel CNNs map point sets onto regular domains, either as multi-view depth or intensity images [17, 37] or as dense voxels [23, 57]. While these projections permit classical convolutions, they inevitably sacrifice geometric fidelity: fine-scale details are smeared by discretisation and quantisation, and memory footprints scale cubically with resolution.

MLP-based approaches. PointNet [26] inaugurated a contrasting philosophy: operate directly on the raw, unordered coordinates. A shared multilayer perceptron followed by a symmetric pooling operator yields permutation invariance without resorting to hand-crafted neighbourhoods. PointNet++ [28] and its successors introduce hierarchical sampling and local shared MLPs, narrowing the gap to convolution while retaining simplicity [20, 53]. The absence of heavy kernel construction makes these networks lightweight and highly parallelisable.

Convolution-based approaches. A separate line of work strives to endow point clouds with true convolution. Kernel Point Convolution [40] and IPA [22] define continuous kernels anchored to learnable points; PointCNN [21] and PointConv [44] generate filters dynamically from local coordinates. Such designs capture local structures more explicitly than MLPs, but at the cost of elaborate kernel parametrisations and higher computational overhead.

Edge-aware approaches. EdgeConv [42] reinterprets convolution as message passing along edges in a dynamic k-NN graph, stimulating a host of derivatives that refine neighbourhood relationships or integrate attention [35, 41, 45, 46, 47, 48, 49, 54]. Although these methods enrich geometric reasoning, the interaction between local edge cues and global context is often mediated by additional MLPs whose influence is hard to analyse rigorously, and converting rich edge descriptors to normalised attention weights can dilute structural signals.

Because our goal is to couple spatial encoding with the billion-sample scalability of the transformer backbone in MOMENT, we deliberately adopt the original, parameter-efficient PointNet formulation. Its shared MLPs furnish strong geometry-aware descriptors with negligible overhead, making them an ideal front-end for a large time-series foundation model.

3 Proposed Method

Our goal is to predict the future displacement $\hat{\mathbf{u}}_{t+1:t+\tau}$ for every PS–InSAR scatterer points, given its past displacement $\mathbf{d}_{t-\ell+1:t} \in \mathbb{R}^{\ell}$ and the neighbour points of displacement data. To exploit both spatial and temporal information, we propose the PointNet-PSI, which unifies a *PointNet* style point cloud encoder with the *MOMENT* time-series backbone. Figure 2 summarises the architecture.

3.1 PointNet-PSI

3.1.1 PointNet–based Spatial Feature Extraction

For a target scatterer i with Cartesian 3D coordinates $\mathbf{p}_i = (x_i, y_i, z_i)$ we first collect all neighbouring scatterers that fall inside a fixed search radius[1] $r = 200$ m,

$$\mathcal{N}_i \;=\; \big\{\, j \;\big|\; \|\mathbf{p}_j - \mathbf{p}_i\|_2 < r,\; j \neq i \big\}.$$

[1] At the 25 m ground sampling of EGMS, a radius of $r = 200$ m encloses at most $|\mathcal{N}_i| \approx 64$ scatterers under uniform spacing.

Neighbour representation. Let ℓ be the length of the *context window* in time steps that the temporal backbone receives (Section 3.1.2 fixes $\ell = 120$). For every neighbour $j \in \mathcal{N}_i$ we concatenate its 3-D position with its normalised displacement history $\mathbf{d}^{(j)}_{t-\ell+1:t} \in \mathbb{R}^{\ell}$:

$$\mathbf{x}_j = \left[x_j,\, y_j,\, z_j,\, \mathbf{d}^{(j)}_{t-\ell+1:t}\right] \in \mathbb{R}^{3+\ell}.$$

Shared MLP and symmetric pooling. PS–InSAR neighbours form an *unordered* set. We therefore adopt a permutation-invariant set encoder (shared MLP + symmetric pooling) to summarise local geometry and kinematic similarity without constructing a graph at every step. This late-fusion summary allows the temporal backbone to retain the input statistics it was pre-trained on, avoiding distribution shifts that arise when spatial features are injected into the encoder tokens.

Following PointNet [26] we apply a shared multilayer perceptron $f_\theta : \mathbb{R}^{3+\ell} \longrightarrow \mathbb{R}^d$, where d is the latent feature dimension (*e.g.*, $d = 128$), to every $\mathbf{x}_j$ and then aggregate the unordered set with a channel-wise Max Pooling operation(MAX):

$$\mathbf{s}_i \;=\; \underset{j\in\mathcal{N}_i}{\mathrm{MAX}} f_\theta(\mathbf{x}_j) \;\in \mathbb{R}^d.$$

The resulting spatial latent vector $\mathbf{s}_i$ is a permutation-invariant descriptor that captures both the local geometry of the point cloud and the kinematic similarity of the surrounding displacement traces. It is subsequently fused with MOMENTs temporal representation in the forecasting head (Section 3.1.2).

3.1.2 MOMENT–based Temporal Forecasting

Background on MOMENT. *Massive Online Multiscale Encoder for Time-series* (MOMENT) is a foundation model pre-trained on 900 M unlabelled sequences drawn from finance, energy, meteorology and industrial sensors. Let C be the number of input channels, T the (padded) context length (§4.1), and D the hidden dimension of the transformer backbone ($D = 768$ in the `moment-base` checkpoint). MOMENT introduces two architectural ideas that make transformer forecasting practical at this scale.

1. **Patch tokenisation.** Rather than operating on raw time steps, MOMENT groups the sequence into fixed-length patches $\mathbf{p}_k \in \mathbb{R}^{C\times L_{\text{patch}}}$ with $L_{\text{patch}}=8$ by default. Each patch is linearly projected to a D-dimensional token, reducing the effective sequence length by a factor L_{patch} and hence the quadratic cost of self-attention.

2. **Hierarchical attention.** A pyramid of J transformer stages ($J=4$ in our model) processes the tokens at progressively coarser resolutions, while a *global memory* attends to every stage. The resulting receptive field scales as $\mathcal{O}(N \log N)$ with $N = T/L_{\text{patch}}$ tokens, capturing both short-range fluctuations and multi-year trends.

Pre-training combines a masked-patch objective with a next-series contrastive task, producing representations that transfer robustly to downstream domains such as PS–InSAR.

Patch Embedding. After REVIN normalisation, the displacement context is partitioned into $N = T/L_{\text{patch}}$ non-overlapping patches. Each flattened patch $\mathbf{x}_t \in \mathbb{R}^{C L_{\text{patch}}}$ is projected to the model dimension

$$\mathbf{v}_t \;=\; W_{\text{patch}}\mathbf{x}_t \;+\; \mathbf{b}_{\text{patch}} \quad \in \mathbb{R}^D,$$

augmented with sinusoidal positional encodings, and passed through a 0.1 dropout layer. The resulting sequence $\{\mathbf{v}_t\}_{t=1}^{N}$ forms the input to the encoder.

Hierarchical Transformer Encoder. MOMENT replaces vanilla ViT blocks with a T5-style stack. Each of the L=24 blocks contains multi-head self-attention (16 heads, 64-d sub-spaces) with learned relative-position biases, followed by a gated GELU feed-forward network of width $2.75D$. In the upper half of the stack, keys and values are computed on a sub-sampled token stream, reducing both memory and run-time to $\mathcal{O}(N \log N)$ while retaining a wide receptive field. All sub-layers are preceded by layer normalisation and followed by 0.1 dropout.

Forecasting Head. The encoder outputs $\mathbf{H} \in \mathbb{R}^{N \times D}$. We flatten $\mathbf{H}$ along the temporal axis, apply 0.1 dropout, and obtain the τ-step forecast via

$$\hat{\mathbf{u}}_{t+1:t+\tau} \;=\; W_{\text{head}}\,\text{Flatten}(\mathbf{H}) \;+\; \mathbf{b}_{\text{head}} \qquad \in \mathbb{R}^{\tau},$$

where τ=60 in all experiments. When PointNet conditioning is enabled, the spatial latent $\mathbf{s}_i$ is concatenated *only at the head stage*. This late fusion preserves the pre-trained encoder's token distribution and delegates spatial–temporal fusion to the head.

3.2 Optimisation

The entire network is trained end-to-end with the mean-squared error(MSE) loss

$$\mathcal{L} \;=\; \frac{1}{S\,\tau} \sum_{i=1}^{S} \left\| \hat{\mathbf{u}}^{(i)}_{t+1:t+\tau} - \mathbf{u}^{(i)}_{t+1:t+\tau} \right\|_2^2, \tag{1}$$

where S denotes the mini-batch size and τ is the prediction horizon (§4.1).[2] Optimisation uses AdamW with $\beta_1 = 0.9$, $\beta_2 = 0.95$ and a cosine learning-rate schedule. Training is performed in mixed precision (bfloat16) and the total gradient $\mathbf{g}$ is clipped to $\|\mathbf{g}\|_2 \leq 5$ to prevent an explosion. We monitor the validation RMSE after each epoch and *store a checkpoint only when this metric improves*, ensuring that the best-performing model is retained for final evaluation. All symbols appearing in EQ (1) have been introduced: S (mini-batch size), τ (forecast horizon), t (index of the last context step), and $\hat{\mathbf{u}}$ and $\mathbf{u}$ (forecast and ground-truth trajectories).

4 Experimental Results

4.1 Dataset

We use the publicly available European Ground Motion Service 2019–2023 product [29], which provides annual vertical displacement rates for the entire continent. To keep the study tractable yet geophysically diverse, we select tiles covering volcanic and anthropogenic deformation areas. After filtering out short or corrupted series, everyone expressed as a 180-length displacement trace (2019:01 → 2021:12).

[2] Throughout, $\hat{\mathbf{u}}$ and $\mathbf{u}$ are the predicted and true displacement sequences, respectively.

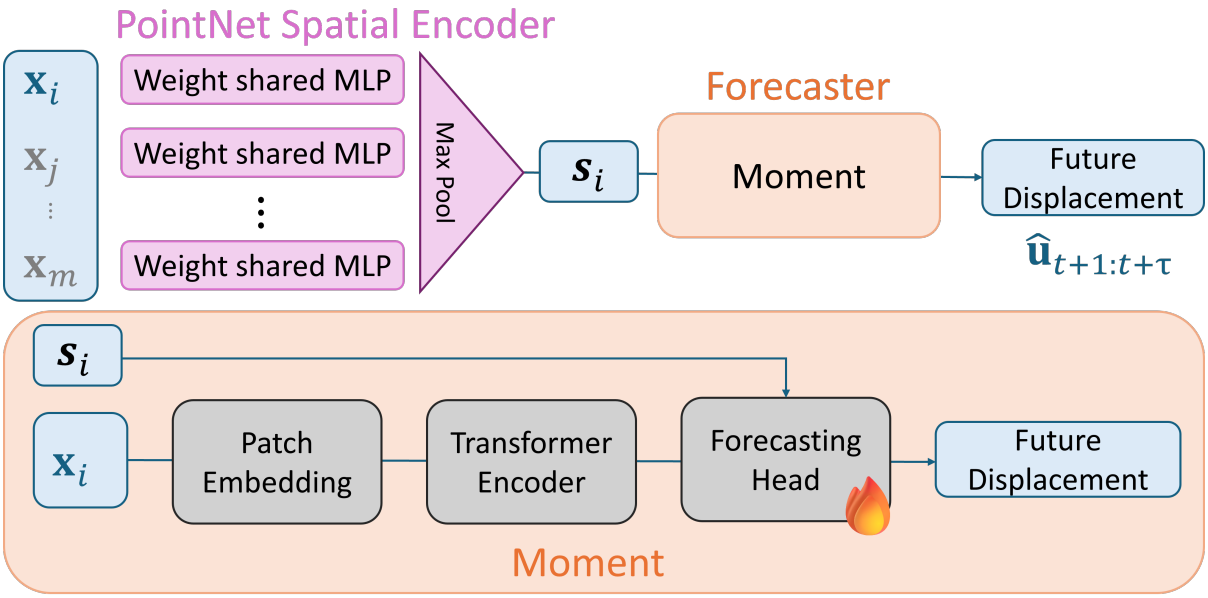

Figure 2: **PointNet-PSI architecture.** The target PS-InSAR displacement track $\mathbf{x}_i$, together with its spatial neighbourhood tracks $\mathbf{x}_j, \ldots, \mathbf{x}_m$, is first processed by a **PointNet**–based spatial encoder, which compresses the local geometry and kinematics into a spatial latent vector $\mathbf{s}_i$. The raw displacement history of the target point is then fed to the **MOMENT** forecaster. The forecasting head fuses the temporal representation with $\mathbf{s}_i$ and outputs the τ-step displacement forecast ($\hat{\mathbf{u}}_{t+1:t+\tau}$).

Following the geophysical convention of hindcast evaluation, we divide the time axis, not the spatial domain: observations from **2019–2020** (2×60 acquisitions, ESA Sentinel-1 revisits each track roughly every six days; after PS processing the average tile still yields $\sim$ 60 valid frames/year.) form the *context*, while all 60 frames of **2021** constitute the prediction horizon. Series are randomly assigned to 70% training, 15% validation, and 15% blind test sets.

Every context displacement data is zero-padded on the right to 256 steps–the fixed input length required by MOMENT–and a per-split Z-score[3] normalisation is applied using training statistics only.

4.2 Results and Discussion

Quantitative comparison. Table 1 lists root-mean-square error (RMSE) on the blind-test year 2021. **PointNet-PSI** delivers the lowest errors, outperforming all per-point baselines (LSTM, Informer, vanilla MOMENT) as well as graph-based variants. *Vanilla MOMENT* already beats LSTM and Informer, underscoring the strength of recent foundation models for generic time-series forecasting; its large-scale pre-training and hierarchical attention give it a clear advantage even when spatial cues are absent. Nevertheless, stripping away the PointNet block and feeding zero-padded tracks directly to MOMENT raises the RMSE to 3.30 mm and yields spatially noisy displacement fields, confirming that explicit spatial learning remains essential. PointNet without MOMENT baseline recovered part of the lost accuracy but still trails PointNet-PSI, highlighting the benefit of our time series foundation model.

[3]Z-score normalisation: $a' = (a - \mu)/\sigma$ computed per split using training statistics.

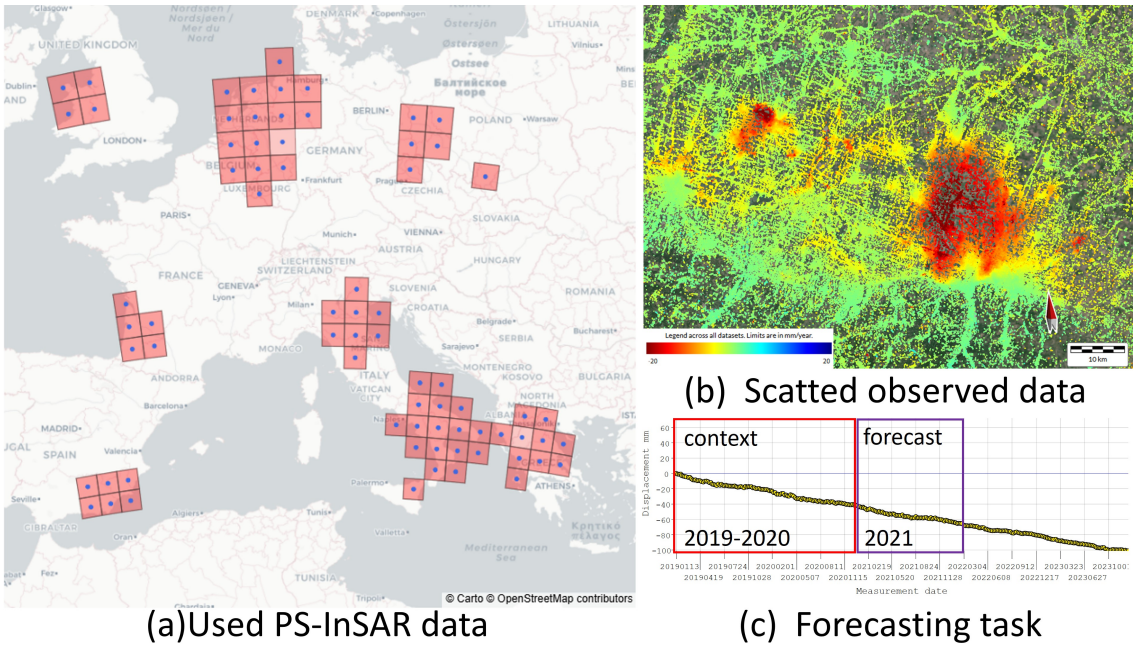

Figure 3: **Dataset.** (a) Spatial coverage of the 100 km Œ 100 km EGMS PS-InSAR tiles used in this work (red outlines, blue centroids). The selected data captures a broad range of tectonic, volcanic and anthropogenic deformation. (b) Irregularly scattered point cloud-like vertical-velocity map (mm yr^{-}1, red = subsidence, blue = uplift). (c) Example displacement trace. The first 120 frames (2019–2020) serve as context for the model, and the next 60 frames (2021) are withheld for forecasting. ©OpenStreetMap contributors. ©Carto.

Model	RMSE [mm]	Foundation Model	Spatial Feature
LSTM [14]†	3.87	-	-
Informer [55]†	3.42	-	-
Vanilla MOMENT [12]†	3.28	✓	-
Vanilla PointNet⋆	3.96	-	✓
PointNet–PSI (ours)	**2.71**	✓	✓

Table 1: **Forecast accuracy on EGMS 2021** (lower is better). † = our implementation; ⋆ = PointNet without Moment.

Qualitative evaluation. Figure 4 illustrates six representative test-series generated by our PointNet-PSI. Each panel shows the 120-step input context (solid navy), the 60-step ground-truth continuation (dotted blue), and our forecast (dashed red). The examples were deliberately chosen to span the spectrum of ground-motion behaviours found in the EGMS corpus. In both monotonic subsidence and uplift the network extrapolates the long-term trend almost perfectly, preserving both slope and absolute magnitude. For seasonally modulated subsidence, PointNet-PSI reproduces the annual oscillation that is super-imposed on the downward drift, showing that the temporal backbone retains high-frequency information even after spatial conditioning. When the motion is seasonally modulated uplift the predicted displacement closely matches the phase and amplitude of the observed cycle, confirming that the model does not merely fit a linear trend. The final pair of examples contains pronounced non-linear transients–abrupt rate changes and episodic accelerations that are notoriously difficult to forecast. Although small timing discrepancies remain, the predictions capture both

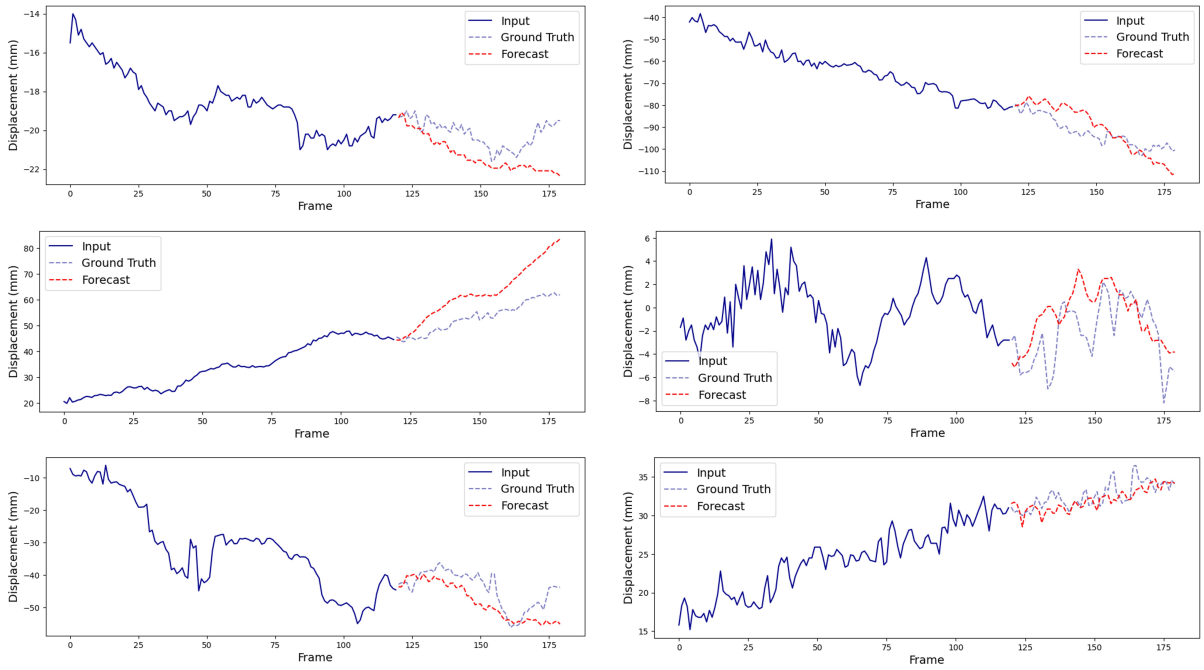

Figure 4: **Qualitative result of test data.** Ground true 2021 displacements (blue dashed line) and PointNet-PSI (red line).

the onset and the sign of these events, whereas per-point baselines typically revert to the mean once the context ends.

Across all six panels, the red dashed forecasts remain spatially coherent with their surroundings (not shown), validating the benefit of the PointNet-derived neighbourhood descriptor. Taken together with the quantitative gains in Table 1, the visual evidence highlights the ability of PointNet-PSI to model a wide variety of geophysical deformation regimes without any hand-crafted priors.

5 Conclusion

Accurately forecasting ground deformation is pivotal for hazard mitigation, infrastructure management, and urban planning. This paper has shown that PS–InSAR stacks can be treated as 3-D point clouds in which each scatterer carries a rich displacement history. Leveraging this view, we proposed **PointNet–PSI**, a hybrid architecture that augments the time-series foundation model **MOMENT** with a lightweight, permutation-invariant **PointNet** encoder. Experiments on the *European Ground Motion Service Basic 2019-2023* data set demonstrated that the spatially enriched backbone not only improves numerical accuracy but also produces forecasts that remain consistent across space. This predictive method is likely to advance downstream tasks such as anomaly detection.

Despite its benefits, PointNet-PSI incurs additional cost because each forward pass must query a fixed-radius neighbourhood, an operation whose complexity grows with point density. Performance may therefore degrade in tiles where the scatterer spacing varies sharply (for example, from rural to urban zones), and the current design has not been validated on man-made structures such as dams or bridges, where deformation mechanisms differ from those of natural terrain. Three avenues appear particularly promising. First, *scal-*

able indexing: replacing on-the-fly radius searches with hierarchical spatial indices or pre-computed k-NN graphs would dramatically reduce the cost of neighbourhood retrieval and make continental-scale inference routine. Second, *spatial-aware pooling*: incorporating set-abstraction layers in the spirit of PointNet should bolster robustness when the PS–InSAR scattered points are highly non-uniform, as are common near coastlines or urban centres. Third, *task transfer*: the large PS–InSAR dataset leveraged here effectively trains a foundation model for ground motion; fine-tuning this model on sparse, high-rate monitoring networks deployed around dams, bridges and tunnels could unlock reliable early-warning capability for critical infrastructure.

By coupling spatial context with foundation-scale temporal modelling, PointNet-PSI charts a practical path toward continent-wide, physically coherent now-casting and opens the door to a new class of point-cloud-based deformation forecasting systems.

Acknowledgements

This study makes use of displacement time-series data provided by the *European Ground Motion Service* (EGMS), whose open access policy is gratefully acknowledged. We used ABCI 3.0 provided by AIST and AIST Solutions [38].

References

[1] Abdul Fatir Ansari, Lorenzo Stella, Caner Turkmen, Xiyuan Zhang, Pedro Mercado, Huibin Shen, Oleksandr Shchur, Syama Sundar Rangapuram, Sebastian Pineda Arango, Shubham Kapoor, et al. Chronos: Learning the language of time series. *arXiv preprint arXiv:2403.07815*, 2024.

[2] Martin Arjovsky, Amar Shah, and Yoshua Bengio. Unitary evolution recurrent neural networks. In Maria Florina Balcan and Kilian Q. Weinberger, editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 1120–1128, New York, New York, USA, 20–22 Jun 2016. PMLR.

[3] Alex Aussem. Dynamical recurrent neural networks towards prediction and modeling of dynamical systems. *Neurocomputing*, 28(1-3):207–232, 1999.

[4] Coryn A L Bailer-Jones, David J C MacKay, and Philip J Withers. A recurrent neural network for modelling dynamical systems. *Network: Computation in Neural Systems*, 9(4):531–547, 1998. doi: 10.1088/0954-898X_9_4_008. URL https://doi.org/10.1088/0954-898X_9_4_008. PMID: 10221578.

[5] Y. Bengio, P. Simard, and P. Frasconi. Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 5(2):157–166, 1994. doi: 10.1109/72.279181.

[6] Bo Chang, Minmin Chen, Eldad Haber, and Ed H Chi. Antisymmetricrnn: A dynamical system view on recurrent neural networks. *arXiv preprint arXiv:1902.09689*, 2019.

[7] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*, 2014.

[8] Abhimanyu Das, Weihao Kong, Rajat Sen, and Yichen Zhou. A decoder-only foundation model for time-series forecasting. In *Forty-first International Conference on Machine Learning*, 2024.

[9] A. Ferretti, C. Prati, and F. Rocca. Permanent scatterers in sar interferometry. *IEEE Transactions on Geoscience and Remote Sensing*, 39(1):8–20, 2001. doi: 10.1109/36.898661.

[10] Azul Garza and Max Mergenthaler-Canseco. Timegpt-1, 2023.

[11] Ebrahim Ghaderpour, Benedetta Antonielli, Francesca Bozzano, Gabriele Scarascia Mugnozza, and Paolo Mazzanti. A fast and robust method for detecting trend turning points in insar displacement time series. *Computers & Geosciences*, 185:105546, 2024.

[12] Mononito Goswami, Konrad Szafer, Arjun Choudhry, Yifu Cai, Shuo Li, and Artur Dubrawski. Moment: A family of open time-series foundation models. In *International Conference on Machine Learning*, 2024.

[13] Samuel Greydanus, Misko Dzamba, and Jason Yosinski. Hamiltonian neural networks. *Advances in neural information processing systems*, 32, 2019.

[14] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[15] Andrew Hooper, Howard Zebker, Paul Segall, and Bert Kampes. A new method for measuring deformation on volcanoes and other natural terrains using insar persistent scatterers. *Geophysical research letters*, 31(23), 2004.

[16] Xiaowei Jia, Jared Willard, Anuj Karpatne, Jordan Read, Jacob Zwart, Michael Steinbach, and Vipin Kumar. *Physics Guided RNNs for Modeling Dynamical Systems: A Case Study in Simulating Lake Temperature Profiles*, pages 558–566. doi: 10.1137/1.9781611975673.63.

[17] Asako Kanezaki, Yasuyuki Matsushita, and Yoshifumi Nishida. Rotationnet: Joint object categorization and pose estimation using multiviews from unsupervised viewpoints. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5010–5019, 2018.

[18] Giancarlo Kerg, Kyle Goyette, Maximilian Puelma Touzel, Gauthier Gidel, Eugene Vorontsov, Yoshua Bengio, and Guillaume Lajoie. Non-normal recurrent neural network (nnrnn): learning long time dependencies while improving expressivity with transient dynamics. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[19] B. O. Koopman. Hamiltonian systems and transformation in Hilbert space. *Proceedings of National Academy of Sciences*, 17:315–318, 1931.

[20] Shiyi Lan, Ruichi Yu, Gang Yu, and Larry S Davis. Modeling local geometric structure of 3d point clouds using geo-cnn. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 998–1008, 2019.

[21] Yangyan Li, Rui Bu, Mingchao Sun, Wei Wu, Xinhan Di, and Baoquan Chen. Pointcnn: Convolution on χ-transformed points. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pages 828–838, 2018.

[22] Jiageng Mao, Xiaogang Wang, and Hongsheng Li. Interpolated convolutional networks for 3d point cloud understanding. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1578–1587, 2019.

[23] Daniel Maturana and Sebastian Scherer. Voxnet: A 3d convolutional neural network for real-time object recognition. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 922–928. IEEE, 2015.

[24] John Miller and Moritz Hardt. Stable recurrent models. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=Hygxb2CqKm.

[25] Yuqi Nie, Nam H. Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. In *International Conference on Learning Representations*, 2023.

[26] Charles R. Qi, Hao Su, Kaichun Mo, and Leonidas J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[27] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 652–660, 2017.

[28] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in Neural Information Processing Systems*, 30, 2017.

[29] Copernicus Land Monitoring Service. European ground motion service: Basic 2019-2023 (vector), europe, yearly, oct. 2024, 2024. URL https://doi.org/10.2909/7eb207d6-0a62-4280-b1ca-f4ad1d9f91c3.

[30] Takayuki Shinohara and Hidetaka Saomoto. Ground-displacement forecasting from satellite image time series via a koopman-prior autoencoder. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 2025.

[31] Takayuki Shinohara and Hidetaka Saomoto. Vit-koop: Vision-transformerkoopman operators for efficient time-series forecasting of earth-observation data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 2025.

[32] Takayuki Shinohara, Haoyi Xiu, and Masashi Matsuoka. Fwnet: Semantic segmentation for full-waveform lidar data using deep learning. *Sensors*, 20(12), 2020. ISSN 1424-8220. doi: 10.3390/s20123568. URL https://www.mdpi.com/1424-8220/20/12/3568.

[33] Takayuki Shinohara, Haoyi Xiu, and Masashi Matsuoka. Semantic segmentation for full-waveform lidar data using local and hierarchical global feature extraction. In *Proceedings of the 28th International Conference on Advances in Geographic Information Systems*, SIGSPATIAL '20, page 640650, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450380195. doi: 10.1145/3397536.3422209. URL https://doi.org/10.1145/3397536.3422209.

[34] Takayuki Shinohara, Haoyi Xiu, and Masashi Matsuoka. Point2wave: 3-d point cloud to waveform translation using a conditional generative adversarial network with dual discriminators. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:11630–11642, 2021. doi: 10.1109/JSTARS.2021.3124610.

[35] Martin Simonovsky and Nikos Komodakis. Dynamic edge-conditioned filters in convolutional neural networks on graphs. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3693–3702, 2017.

[36] Rupika Soni, Mohammad Soyeb Alam, and Gajendra K Vishwakarma. Prediction of insar deformation time-series using improved lstm deep learning model. *Scientific Reports*, 15(1):5333, 2025.

[37] Hang Su, Subhransu Maji, Evangelos Kalogerakis, and Erik Learned-Miller. Multi-view convolutional neural networks for 3d shape recognition. In *Proceedings of the IEEE international conference on computer vision*, pages 945–953, 2015.

[38] Ryousei Takano, Shinichiro Takizawa, Yusuke Tanimura, Hidemoto Nakada, and Hirotaka Ogawa. Abci 3.0: Evolution of the leading ai infrastructure in japan, 2024. URL https://arxiv.org/abs/2411.09134.

[39] Sean J Taylor and Benjamin Letham. Forecasting at scale. *The American Statistician*, 72(1):37–45, 2018.

[40] Hugues Thomas, Charles R Qi, Jean-Emmanuel Deschaud, Beatriz Marcotegui, François Goulette, and Leonidas J Guibas. Kpconv: Flexible and deformable convolution for point clouds. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6411–6420, 2019.

[41] Lei Wang, Yuchun Huang, Yaolin Hou, Shenman Zhang, and Jie Shan. Graph attention convolution for point cloud semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10296–10305, 2019.

[42] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E Sarma, Michael M Bronstein, and Justin M Solomon. Dynamic graph cnn for learning on point clouds. *Acm Transactions On Graphics (tog)*, 38(5):1–12, 2019.

[43] Gerald Woo, Chenghao Liu, Akshat Kumar, Caiming Xiong, Silvio Savarese, and Doyen Sahoo. Unified training of universal time series forecasting transformers. In *Forty-first International Conference on Machine Learning*, 2024.

[44] Wenxuan Wu, Zhongang Qi, and Li Fuxin. Pointconv: Deep convolutional networks on 3d point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9621–9630, 2019.

[45] Haoyi Xiu, Takayuki Shinohara, and Masashi Matsuoka. Dynamic-scale graph convolutional network for semantic segmentation of 3d point cloud. In *2019 IEEE International Symposium on Multimedia (ISM)*, pages 271–2717. IEEE, 2019.

[46] Haoyi Xiu, Xin Liu, Weiming Wang, Kyoung-Sook Kim, Takayuki Shinohara, Qiong Chang, and Masashi Matsuoka. Enhancing local feature learning for 3d point cloud processing using unary-pairwise attention. *arXiv preprint arXiv:2203.00172*, 2022.

[47] Haoyi Xiu, Xin Liu, Weimin Wang, Kyoung-Sook Kim, Takayuki Shinohara, Qiong Chang, and Masashi Matsuoka. Diffusion unit: Interpretable edge enhancement and suppression learning for 3d point cloud segmentation. *Neurocomputing*, 559:126780, 2023.

[48] Haoyi Xiu, Xin Liu, Weimin Wang, Kyoung-Sook Kim, Takayuki Shinohara, Qiong Chang, and Masashi Matsuoka. Ds-net: A dedicated approach for collapsed building detection from post-event airborne point clouds. *International Journal of Applied Earth Observation and Geoinformation*, 116:103150, 2023.

[49] Haoyi Xiu, Xin Liu, Weimin Wang, Kyoung-Sook Kim, Takayuki Shinohara, Qiong Chang, and Masashi Matsuoka. Optimizing local feature representations of 3d point clouds with anisotropic edge modeling. In *International Conference on Multimedia Modeling*, pages 269–281. Springer, 2023.

[50] Yong-An Xue, You-Feng Zou, Hai-Ying Li, and Wen-Zhi Zhang. Regional subsidence monitoring and prediction along high-speed railways based on ps-insar and lstm. *Scientific Reports*, 14(1):24622, 2024.

[51] Nur Yagmur, G Taskin, N Musaoglu, and E Erten. Forecasting surface movements based on psi time series using machine learning algorithms. *International Journal of Remote Sensing*, 45(7):2462–2485, 2024.

[52] Jiayi Zhang, Jian Gao, and Fanzong Gao. Time series land subsidence monitoring and prediction based on sbas-insar and geotemporal transformer model. *Earth Science Informatics*, 17(6):5899–5911, 2024.

[53] Zhiyuan Zhang, Binh-Son Hua, and Sai-Kit Yeung. Shellnet: Efficient point cloud convolutional neural networks using concentric shells statistics. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1607–1616, 2019.

[54] Hengshuang Zhao, Li Jiang, Chi-Wing Fu, and Jiaya Jia. Pointweb: Enhancing local neighborhood features for point cloud processing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5565–5573, 2019.

[55] Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *The Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Virtual Conference*, volume 35, pages 11106–11115. AAAI Press, 2021.

[56] Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun, and Rong Jin. FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *Proc. 39th International Conference on Machine Learning (ICML 2022)*, 2022.

[57] Yin Zhou and Oncel Tuzel. Voxelnet: End-to-end learning for point cloud based 3d object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4490–4499, 2018.