



HAL
open science

Denoising of geodetic time series using spatiotemporal graph neural networks: application to slow slip event extraction

Giuseppe Costantino, Sophie Giffard-Roisin, Mauro Dalla Mura, Anne Socquet

► To cite this version:

Giuseppe Costantino, Sophie Giffard-Roisin, Mauro Dalla Mura, Anne Socquet. Denoising of geodetic time series using spatiotemporal graph neural networks: application to slow slip event extraction. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2024, 17, pp.17567-17579. 10.1109/jstars.2024.3465270 . hal-04880304

HAL Id: hal-04880304

<https://cnrs.hal.science/hal-04880304v1>

Submitted on 11 Jan 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Denoising of Geodetic Time Series Using Spatiotemporal Graph Neural Networks: Application to Slow Slip Event Extraction

Giuseppe Costantino , Sophie Giffard-Roisin , Mauro Dalla Mura , *Senior Member, IEEE*, and Anne Socquet 

Abstract—Geospatial data have been transformative for the monitoring of the Earth, yet, as in the case of (geo) physical monitoring, the measurements can have variable spatial and temporal sampling and may be associated with a significant level of perturbations degrading the signal quality. Denoising geospatial data is, therefore, essential, yet often challenging because the observations may comprise noise coming from different sources, including both environmental signals and instrumental artifacts, which can be spatially and temporally correlated, thus hard to disentangle. This study addresses the denoising of multivariate time series acquired by irregularly distributed networks of sensors, requiring specific methods to handle the spatiotemporal correlation of the noise and the signal of interest. Specifically, our method focuses on the denoising of geodetic position time series, used to monitor ground displacement worldwide with centimeter-to-millimeter precision. Among the signals affecting global navigation satellite system (GNSS) data, slow slip events (SSEs) are of interest to seismologists. These are transients of deformation that are weakly emerging compared to other signals. Here, we design SSEdenoiser, a multistation spatiotemporal graph-based attentive denoiser that learns latent characteristics of GNSS noise to reveal SSE-related displacement with submillimeter precision. It is based on the key combination of graph recurrent networks and spatiotemporal Transformers. The proposed method is applied to the Cascadia subduction zone, where SSEs occur along with bursts of tectonic tremors, a seismic rumbling identified from independent seismic recordings. The extracted events match the spatiotemporal evolution of tremors. This good space–time correlation of the denoised GNSS signals with the tremors validates the proposed denoising procedure.

Index Terms—Deep learning, denoising, geodesy, geospatial data, global navigation satellite system (GNSS), global positioning system (GPS), graph neural networks, multistation, seismology, slow slip events (SSEs), spatiotemporal, spatiotemporal attention, time-series analysis.

Received 19 April 2024; revised 24 June 2024 and 27 August 2024; accepted 11 September 2024. Date of publication 19 September 2024; date of current version 7 October 2024. This work was supported in part by ERC under Grant CoG 865963 DEEP-trigger, in part by MIAI@Grenoble Alpes under Grant ANR-19-P3IA-0003, in part by HPC resources, and in part by GENCI-IDRIS under Grant 2023-AD011012373R2. (*Corresponding author: Giuseppe Costantino.*)

Giuseppe Costantino, Sophie Giffard-Roisin, and Anne Socquet are with CNRS, IRD, ISTerre, University Grenoble Alpes, 38000 Grenoble, France (e-mail: giuseppe.costantino@univ-grenoble-alpes.fr).

Mauro Dalla Mura is with CNRS, Grenoble INP, GIPSA-Lab, University Grenoble Alpes, 38000 Grenoble, France, and also with the Institut Universitaire de France, 75000 Paris, France.

The codes of SSEdenoiser are publicly available at <https://github.com/gcostantino/sse-denoising-gnss>

Digital Object Identifier 10.1109/JSTARS.2024.3465270

I. INTRODUCTION

GEOSPATIAL data represents a crucial resource for the efficient acquisition of large-scale measurements, providing a detailed view of Earth processes across numerous applications. In the past few decades, the use of geospatial data [e.g., satellite imagery, global navigation satellite system (GNSS), seismic data, climate data] has significantly contributed to remote sensing, both for the monitoring of the environment [1] and natural processes [2], [3], [4], and for urban [5] and infrastructure development [6], [7]. Geospatial data typically incorporates information on the location of target objects with the temporal tracking of some of their characteristics, allowing for spatiotemporal analysis. However, these measurements are strongly affected by various perturbations. These noises can be spatially and temporally correlated, making it challenging to extract useful information, i.e., to separate the characteristics of target objects from those of external perturbations. Depending on the specific application, the raw geospatial measurements can be modeled as a mixture of several sources, e.g., atmospheric interference, uncertainties and artifacts related to the sensor measurements, and different tectonic or topographic signals. Separating these mixed sources can be a difficult task, yet it is key to performing successful signal extraction. Depending on the structural type of geospatial data (e.g., images, time series), different noise components arise, which require dedicated post-processing, such as in the case of image-like data, e.g., satellite images, or time-series-like data, e.g., GNSS data. In this article, we focus on time-series-like data and, more specifically, we address the problem of denoising multivariate time series in a multistation setting. We target our analysis on the denoising of GNSS position time series and we apply our method to the extraction of transients of aseismic deformation. Albeit designed for this specific task, the developed framework is general enough to be applied to any time series denoising problem with multiple measurement sites (array of sensors) and multiple nonindependent components or channels (multivariate), e.g., GNSS data, seismic data, gravimeters, tiltmeters, strainmeters.

GNSS is an essential tool in modern geodesy, providing accurate and precise positional information that has transformed our understanding of Earth's shape, movement, and processes. Using satellite constellations, time series of displacement relative to a GNSS site (antenna) are calculated, which can have high precision, usually of the order of the millimeter. GNSS

time series usually fall into two categories: high-rate GNSS, typically sampled at 5-min intervals, and daily GNSS, where measurements are averaged to provide a daily displacement value. GNSS sites are arranged as a network of antennas, resulting in uneven regional sampling that is sparser than satellite images, for instance.

The use of the GNSS has been transformative for various remote sensing applications, such as the observation and monitoring of soil moisture [8], [9], landslide monitoring [10], and hydrology-related applications [11], opening a large number of perspectives both toward societal and environmental advancements [12], [13]. Also, several studies focus on improving the precision of the GNSS positioning [14] as well as extracting statistics from the noise affecting the data [15]. Recently, advanced signal processing methods (among which machine learning) have been explored for subsurface exploration with promising advances toward better imaging the Earth structure [16]. In this direction, deep learning-based methods have been widely used in remote sensing as a way to boost traditionally employed strategies as well as to extract finer grained knowledge from massive datasets. Diverse studies have used machine and in particular deep learning in geoscience, e.g., for the detection of volcano-seismic signals [17], [18], [19], [20], seismic signal restoration [21], and full-waveform inversion [22]. Some works also use GNSS data, e.g., for probabilistic source characterization [23] or time series regression [24]. However, the number of studies combining deep learning with GNSS data for tectonics is still limited, probably due to its scarcity (in space, but also in time: the acquisition frequency is between an hour and one day) and because proper modeling of the noise in the GNSS positioning remains a challenge.

Position time series derived from GNSS are affected by different signals generally characterized by spatial and temporal correlation. The challenge in GNSS data denoising lies in developing a method able to learn how to decorrelate these signals by separating what we consider as noise and the different signals from each other. Two major families of signals can be identified as follows. 1) Geophysical (e.g., tidal or hydrological loading) and tectonic signals (e.g., earthquakes and related processes, such as aseismic deformation and postseismic relaxation). 2) Uncertainties in the position (e.g., errors on the calculation of the orbits of the satellites, clock synchronization, multipath errors, ionospheric, and tropospheric delays [25], [26], [27]). Most studies use parametric models at each station [28], [29] or blind source separation methods [30] to model signals constituting the GNSS position, yet these methods may not be able to fully reproduce the complexity of the signals.

Recent works have shown promising results in seismology to address the problem of denoising seismic waveforms or high-rate GNSS time series for seismological applications. Saad et al. [31] developed a deep autoencoder to attenuate the random noise affecting time-series seismic data. The model works by first encoding the time series to several levels of abstraction and then decoding the compressed information to reconstruct the noiseless seismic signal. Zhou et al. [32] designed a deep learning-based decomposition/denoising method named DeepDenoiser. The method is a U-Net trained to learn a sparse

representation of the data in the time–frequency domain and nonlinear-mapped masks to separate the signal of interest from the noise. Thomas et al. [33] used a modified version of DeepDenoiser adapted to work with three-component high-rate GNSS data. They test three versions of DeepDenoiser to account for amplitude distortions and phase differences. All the previous methods employ single-station time series as input, to provide clean time series for each station separately. However, the main limitation arising from the use of single-station methods is that the spatial variability of the data is not exploited. Multistation methodologies have been also explored in seismology as an attempt to take into account the spatial coherency of the measurements, for several tasks in which this would translate into an improvement of the performance [34], [35].

In the case of GNSS data, and especially for the identification of slow slip events, exploiting the spatial variability of the GNSS position becomes critical because of the lower signal-to-noise ratio compared to seismic data or high-rate GNSS. Moreover, both the GNSS noise perturbations and the interesting tectonic signals are spatially and temporally correlated and, therefore, the spatial and temporal analysis should not be decoupled. In previous works, we tested several GNSS data representations aiming to explicitly inject spatial information in the context of earthquake characterization [36] and we developed a multistation attentive deep neural network for slow slip event identification [37]. We found that multistation approaches achieve better performance when targeting the ground deformation with daily position GNSS time series and should be generally preferred over single-station methods.

The aforementioned approaches rely on single-station analysis, which represents a limitation for GNSS, being the measured displacement correlated in the spatial and temporal dimensions. In this work, we aim to develop a deep learning-based method for the denoising of raw GNSS data with a multistation approach. One way to deal with sparse GNSS measurements is to arrange them as a matrix, where each row is a time series, sorted for example by latitude (or longitude), and to use 2-D convolutional neural networks (CNNs), specifically developed for image-like data [see the representation in Fig. 1(a)]. With this method, 2-D convolutions (time and station dimension) can model both local- and large-scale spatial relationships between stations, but no explicit information coming from the geometry of the GNSS network is enforced: one out of the two spatial dimensions is lost. To this end, spatiotemporal graph neural network (STGNN) methods [38], which, to the best of our knowledge, have never been applied to remote sensing, can learn these relationships and take full benefit of the spatial information, with the potential of outperforming the 2-D-CNN approach. However, developing such models requires setting up explicit joint processing of the information in the temporal and spatial domains. Classical graph-based approaches rely on message passing to propagate the information between neighboring nodes. Time-varying features can be handled by using a recurrent neural network (RNN) cell as an aggregation function, e.g., long short-term memory (LSTM) or gated recurrent unit, but extending this mechanism to multivariate time series analysis for spatiotemporal methods is not straightforward since GNSS measurements usually have

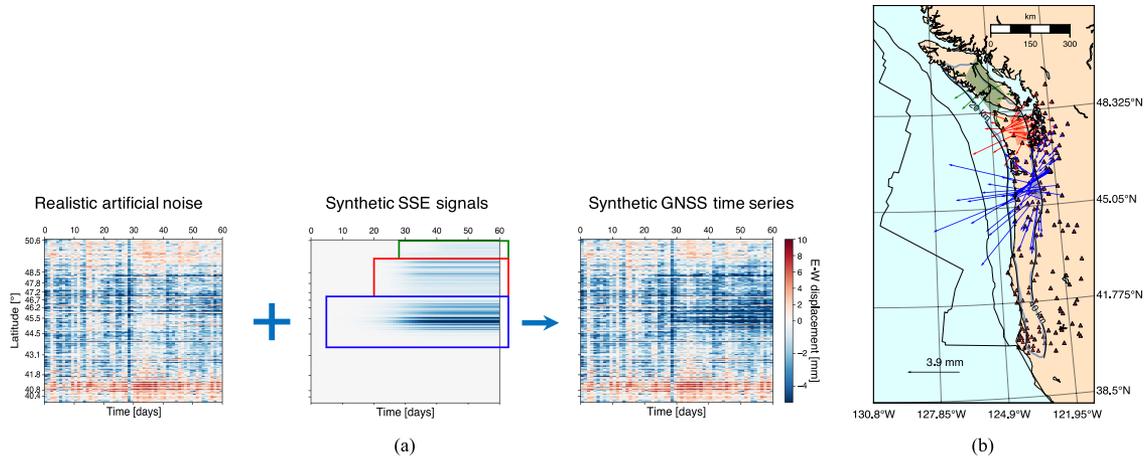


Fig. 1. Overview of the synthetic data generation. (a) Each row of the matrices represents a synthetic detrended GNSS position time series (E-W component) color-coded by the amplitude of displacement, created using the proposed synthetic data generation technique. Starting from the left: artificial noise $n(t)$, modeled SSE signals $d(t)$, synthetic GNSS time series $\xi(t) = n(t) + d(t)$. (b) Red triangles represent the location of the GNSS stations (MAGNET GNSS network) used in this study. The arrows show the static synthetic displacement modeled at each GNSS station. In this example, the SSE signal is modeled through three dislocations, shown with different colors (red, green, blue), slipping in an elastic half-space with a different slip amount, slip initiation time, and slip duration. The synthetic displacement time series associated with each dislocation are identified by rectangles in the $d(t)$ matrix in the (a) panel with the same color. The light blue contour represents the SSE area, e.g., the locations of the generated synthetic slow slip events.

more than one temporal channel (here, we use North–South displacement and East–West displacement components). One solution is to consider the different components as statistically independent and enforce classic message-passing approaches on each component independently. Yet, the assumption of independence between components is too strong since the different GNSS components are generally correlated.

Spatiotemporal approaches have thus been developed to jointly rely on multiple sensor measurements and multiple components. They can be classified in RNN-based and attention-based methods, and are capable of dealing with 3-D data (here, [stations, time, directions]). The first family of approaches uses RNNs to extract temporal features [39], while the second focuses on attention mechanisms in the time and/or space dimension [40]. Also, spatiotemporal methods usually consider that the graph connectivity (adjacency matrix) is available beforehand. Most of the previous works compute the adjacency matrix based on distance metrics (e.g., traffic forecasting) [38] or prior information on the nodes (e.g., skeleton-based action recognition) [40]. However, relying on precomputed connectivity might not be the optimal choice, since additional edges could be learned from the data, and superfluous connections could also be removed. Moreover, graph connectivity is not always available, as in our case, and computing a satisfactory adjacency matrix is challenging since it could require connecting nodes both within a homogeneous mesh and with long-range connections, which are not easy to model. Hence, the solution of learning the adjacency matrix during the training, following the same approach by Bai et al. [39], is often relevant.

GNSS time series are affected by data gaps and missing stations, making it necessary to interpolate the missing information or to set up specific techniques to account for data discontinuities effectively. Here, we use static graphs and set the data gaps to zero (as in our previous work [37]). Dynamic graphs (e.g., [41] and [42]) offer another significant strategy to

define finer and more robust spatiotemporal modeling. However, it is not straightforward to make existing STGNs dynamic, and computationally expensive because a distinct graph needs to be built at each time step. Here, we will focus on static graphs.

In this article, we apply the denoising of GNSS position time series to identify slow slip events. These tectonic events are characterized by slip on faults, as are earthquakes, yet, unlike them, they do not produce ground shaking and they last over days to years. However, most of the occurrences of slow slip transients have low amplitude and remain undetected, since other signals (either geophysical signals or noise) prevail in the GNSS time series. Thus, this represents a relevant case study for the denoising of GNSS data, which can be extended to many other applications dealing with the extraction of low-amplitude signals mixed with other noise components.

Here we present SSEdenoiser, a spatiotemporal graph-based deep neural network designed to extract spatiotemporal features from multistation raw (daily) GNSS position time series and to leverage them to isolate the aseismic slip contribution from the rest of the tectonic and nontectonic signals.

The originality and novelty of this work lie in 1) the coupled processing of the spatial and temporal information through a STGNN with a learned graph connectivity, aimed to uncover hidden relationships between GNSS sites based on the GNSS recordings themselves and 2) the use of a spatiotemporal transformer, aimed to filter the relevant information encoded by the graph neural network both in space and time.

The rest of this article is organized as follows. In Section II, we present the methodology. We first introduce our overall approach, then we describe the data generation strategy and the architecture of the proposed model. In Section III, we present the experimental results, in terms of: results on the synthetic data set (with a comparison with several baseline methods), an ablation study to test the robustness of the proposed methodology, analysis of the learned graph connectivity, results on real GNSS

data in the Cascadia subduction zone and qualitative comparison between the proposed approach and a previous single-station slow slip event detector for the Cascadia subduction zone. Finally, Section IV concludes this article.

II. METHODOLOGY

A. Overall Approach

In this study, we build SSEdenoiser, a multistation deep learning-based method aimed at denoising raw GNSS position time series to extract aseismic transients of deformation. We train SSEdenoiser on synthetic data because of the paucity of previously catalogued slow slip events, by following the same approach previously designed for slow slip event detection [37]. We generate synthetic data by relying on an underlying additive model. The displacement $\xi_i(t) \in \mathbf{R}^{N_c}$ at a given station s_i , with N_c the number of components, is obtained as the sum of a noise term $\mathbf{n}_i(t)$ and a synthetic slow slip signal $\mathbf{d}_i(t)$

$$\xi_i(t) = \mathbf{n}_i(t) + \mathbf{d}_i(t). \quad (1)$$

We generate the noise term (here, everything that is not the displacement of interest) $\mathbf{n}(t)$ from raw GNSS time series using a principal component analysis (PCA) and a Fourier phase randomization and amplitude matching technique, as described in [37]. As a result, the generated noise will have the same spatial covariance and the same spectral and statistical characteristics as the real noise. We generate the synthetic slow slip signal $\mathbf{d}(t)$ by relying on known characteristics of slow slip events and by drawing random events along the subduction interface, with various nuances of signal-to-noise ratio and randomly generated physical parameters, having a depth between 20 and 40 km (with ± 10 km variability). We design the database to contain four different settings: (zero events, one event, two events, three events). The final database includes 25% of each setting. We detail the data generation strategy in the following paragraph.

SSEdenoiser is a deep denoising neural network, made of the combination of a graph-based RNN, extracting spatial and temporal features from the data in a multistation fashion, and a spatiotemporal transformer, designed as a cascade of temporal and spatial self-attention, capable of focusing on precise space–time relationships. We train the method to extract the slow slip event signals in a temporal window with a fixed size of 60 days. We test the method both on synthetic and real data. On synthetic data, we feed 60-day samples to SSEdenoiser to test the performance both on *negative* samples (zero events in the window) and on *positive* samples (up to three events in the window). On real data, we use a sliding window approach, by collecting the denoised data for each window and by aggregating them afterward. The rationale for choosing the window length follows the same motivations as our previous work [37].

B. Synthetic Data Generation

In [37], we introduced SSEgenerator, a method for generating synthetic GNSS noise time series and synthetic slow slip events, based on synthetic surface displacements output by physical models. Here, we improve and adapt this method for a denoising pipeline and we detail the technique in the following paragraphs. We present an overview of the synthetic data generation in Fig. 1.

We use 200 raw GNSS position time series from the Cascadia subduction zone from 2007 to 2022. We choose the stations with the least missing data [37]. As opposed to SSEgenerator, we use the full 15-year sequence to generate the geodetic noise, to have longer sequences and thus a more realistic artificial noise time series. Moreover, we allow each sample to belong to one out of four different settings. 25% of the synthetic samples are considered as *negative* samples, thus they do not contain any slow slip event (SSE) signal. These are achieved by training the model to output a zero-displacement signal, thus focusing on extracting an enhanced representation of the input noise. When addressing denoising, this is key to learning the noise structure at best and can be used to better denoise the *positive* samples.

The challenge in constructing the slow slip signal $\mathbf{d}_i(t)$ lies in employing a synthetic displacement model that generates surface displacements that mimic well the features expected for slow slip events. To do so, we use Okada [43] equations, that link the slip on a dislocation buried in an elastic medium to surface displacements. This allows us to generate synthetic surface displacements that can resemble the expected physical properties of slow slip events. Because SSEs can have a complex source, including variable slip or lateral propagation of the slip pulse at depth, we impose, for each 60-day window, a variable number of synthetic dislocations per window between 0 and 3 (25% each), with independent source parameters

We uniformly generate the SSE sources within a band along the subduction between 20 and 40 km depth [see Fig. 1(b)], with uniformly generated source parameters (magnitudes uniformly drawn between 6 and 7, strike and dip angles enforced by the slab geometry with ± 10 km variability on the depth and rake angle variable from 80° to 100° to produce a thrust focal mechanism). For more details, we refer the reader to our previous work [37].

The temporal signature of the SSE signals is assumed to be a logistic function [37]. We allow the SSEs to last between 10 and 30 days. We model the SSE temporal evolution (one component is shown for simplicity) as follows:

$$d_i(t) = \frac{D}{1 + e^{-\beta(t-t_0)}} \quad (2)$$

where D is the static displacement output by the dislocation model [43], β is associated with the growth rate of the curve and t_0 is the time corresponding to the inflection point of the logistic function. The value of D modulates the amplitude of the surface displacement and depends on the source parameters (e.g., magnitude, depth). By generating surface displacements corresponding to randomly generated source parameters, we explore different nuances of signal-to-noise ratio to train the denoising method. We generate t_0 randomly between 0 and 60 days. We obtain β as a function of the SSE duration T , which can be rewritten as $T = t_{\max} - t_{\min}$, where t_{\max} is the time corresponding to the post-SSE displacement (i.e., D), and t_{\min} is the pre-SSE displacement (i.e., 0). We set a threshold γ , such that t_{\max} and t_{\min} are associated with $d_s(D - \gamma D)$ and $d_s(\gamma D)$, respectively, and $\gamma = 0.01$. We finally obtain

$$\beta = \frac{2}{T} \ln \left(\frac{1}{\gamma} - 1 \right). \quad (3)$$

We generate the SSE duration T as a uniform random variable from 10 to 30 days. We further extract 60-day windows of noise

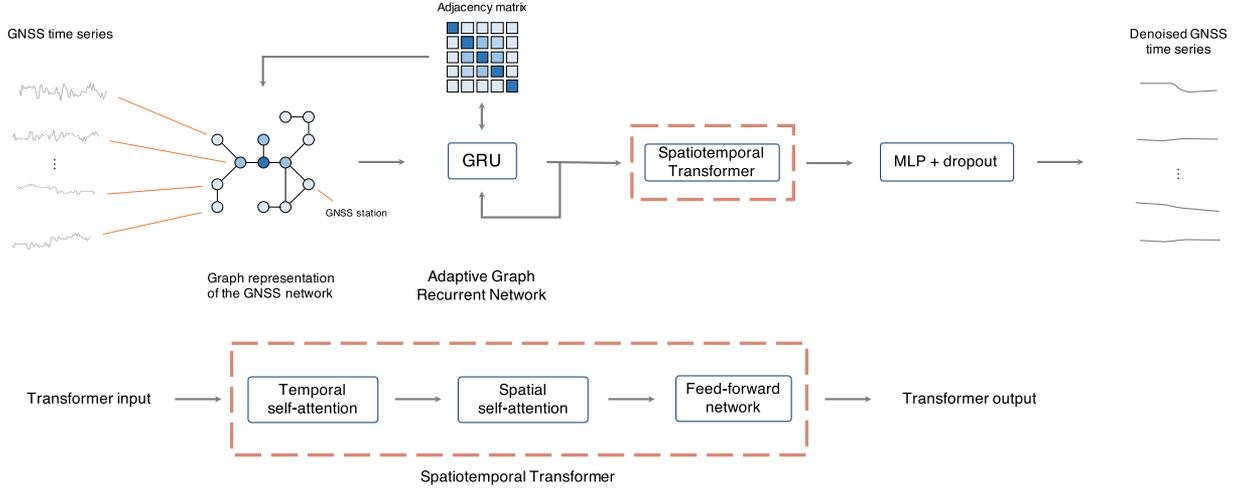


Fig. 2. High-level architecture of SSEdenoiser. GNSS time series are first processed by a graph-based RNN, where temporal features are extracted and spatial relationships are inferred by learning the adjacency matrix. A spatiotemporal transformer is then used, where temporal and spatial self-attentions attend to the learned temporal features and the spatial relationships. At the end of the pipeline, a fully connected network reprojects the transformer’s output to the input dimension to produce the denoised GNSS time series.

$n_i(t)$ to be added to the slow slip signals $d_i(t)$. We also account for missing data by relying on the data gap distribution of the real data, i.e., we use the data gap distribution of real data, and we shuffle it, meaning that a given station will be assigned the data gap distribution of another station (more details in our previous work [37]). In this way, the model learns to account for realistic data gaps with the same number of missing stations as in real data.

C. Architecture of SSEdenoiser

SSEdenoiser is a deep neural network consisting of two main modules: a graph-based RNN and a spatiotemporal transformer. A high-level architecture is provided in Fig. 2.

The graph-based RNN module has been taken from the work by Bai et al. [39]. The method relies on a learnable adjacency matrix A , computed as

$$A = \text{softmax}(\text{ReLu}(\mathbf{E}\mathbf{E}^T)) \quad (4)$$

where $\mathbf{E} \in \mathbb{R}^{N \times d_N}$ are learnable node embeddings. N is the number of nodes in the graph and d_N is the node embedding dimension. We set $N = 200$, corresponding to the number of GNSS stations, and $d_N = 32$. The adjacency matrix is learned to be symmetric, as in the case of undirected graphs. The adjacency matrix is used by the graph convolutional recurrent unit, having the following constitutive equations [39]:

$$\mathbf{z}(t) = \sigma(A[\mathbf{X}(t), \mathbf{h}(t-1)]\mathbf{E}\mathbf{W}_z + \mathbf{E}\mathbf{b}_z) \quad (5)$$

$$\mathbf{r}(t) = \sigma(A[\mathbf{X}(t), \mathbf{h}(t-1)]\mathbf{E}\mathbf{W}_r + \mathbf{E}\mathbf{b}_r) \quad (6)$$

$$\hat{\mathbf{h}}(t) = \tanh(A[\mathbf{X}(t), \mathbf{r} \odot \mathbf{h}(t-1)]\mathbf{E}\mathbf{W}_{\hat{\mathbf{h}}} + \mathbf{E}\mathbf{b}_{\hat{\mathbf{h}}}) \quad (7)$$

$$\mathbf{h}(t) = \mathbf{z} \odot \mathbf{h}(t-1) + (1 - \mathbf{z}) \odot \hat{\mathbf{h}}(t) \quad (8)$$

where $\mathbf{X}(t)$ and $\mathbf{h}(t)$ are the input and output at time t , $[\cdot]$ is the concatenation operation, $\sigma(\cdot)$ the sigmoid function and \odot the Hadamard (elementwise) product. \mathbf{W}_z , \mathbf{W}_r , $\mathbf{W}_{\hat{\mathbf{h}}}$, \mathbf{b}_z , \mathbf{b}_r , and $\mathbf{b}_{\hat{\mathbf{h}}}$ are learnable parameters. We set the hidden size of the recurrent unit to 128 and we use one-hop graph convolution.

The output of the first module, i.e., the output of the last hidden layer $\mathbf{h}(t)$ is then used as input of a Transformer neural network. Usually, transformers are used to attend either to the temporal or the spatial dimension. Here, we use a novel approach, consisting of a cascade of two self-attention mechanisms, one for the temporal and one for the spatial dimension. The proposed strategy first attends to the temporal dimension. The temporal attention output is then fed to the spatial self-attention. As a result, the stacked attentions attend both time and space axes. Mathematically, we define the temporal α_t and spatial α_s self-attention layers [44] as

$$\alpha_t(\mathbf{x}) = \text{softmax}\left(\frac{(\mathbf{W}_{q,t}\mathbf{x})(\mathbf{W}_{k,t}\mathbf{x})^T}{\sqrt{d_k}}\right) \quad (9)$$

$$\alpha_s(\mathbf{x}) = \text{softmax}\left(\frac{(\mathbf{W}_{q,s}\mathbf{x})(\mathbf{W}_{k,s}\mathbf{x})^T}{\sqrt{d_k}}\right) \quad (10)$$

where \mathbf{x} denotes the generic input to the attention layer. $\mathbf{W}_{q,t}$, $\mathbf{W}_{k,t}$ and $\mathbf{W}_{q,s}$, $\mathbf{W}_{k,s}$ are learnable projection matrices for query and key for the temporal and spatial self-attentions, respectively. The output of the temporal self-attention layer is then computed as

$$\mathbf{o}_t = \alpha_t(\mathbf{x})\mathbf{W}_{v,t}\mathbf{x} \quad (11)$$

and the output of the stacked self-attentions is then computed as

$$\mathbf{o}_{t,s} = \alpha_s(\mathbf{o}_t)\mathbf{W}_{v,s}\mathbf{o}_t \quad (12)$$

where $\mathbf{W}_{v,t}$, $\mathbf{W}_{v,s}$ are learnable projection matrices for the value. We set the output and embedding size of the transformer to 128. After the self-attention, we use a dropout layer to avoid overfitting (dropout rate $\rho = 0.1$). We do not derive here the equations of the feedforward network following the self-attention, nor the normalization layers used inside the self-attention: we refer the reader to the general transformer formulation [44]. The output of the transformer is then linearly projected using a linear transformation \mathbf{W}_o (fully connected layer) to output a 2-D (N-S, E-W) denoised time series, obtained from a weighted average of the previous 128 feature maps, modulated, during training, by a dropout layer (dropout rate $\rho = 0.5$). Please

note that here we use the term “feedforward” instead of “MLP” by following the original formulation of Vaswani et al.’s [44] work, indicating a modified MLP that includes skip connections, layer normalization, and dropout.

The proposed method has a total of 3 396 226 trainable parameters, with 3 213 568 parameters associated with the graph-based RNN module and 182 400 with the spatiotemporal transformer. The last linear projection layer has 258 trainable parameters.

D. Training Details

The input of SSEdenoiser is given by the time series $\xi(t)$. We train the model via mini-batch training (batch size of 128) by minimizing the mean squared error (MSE) between the target (clean) $\mathbf{d}(t)$ and the output (denoised) $\hat{\mathbf{d}}(t)$ time series

$$\begin{aligned} \text{MSE}(\mathbf{d}(t), \hat{\mathbf{d}}(t)) &= \frac{1}{n \cdot N_s \cdot N_t \cdot N_d} \sum_{i=1}^n \sum_{j=1}^{N_s} \sum_{t=1}^{N_t} \sum_{k=1}^{N_d} \left(d_j^{i,k}(t) - \hat{d}_j^{i,k}(t) \right)^2 \end{aligned} \quad (13)$$

where n is the number of samples in the minibatch, N_s is the number of stations, N_t is the window length (60 days) and N_d is the number of components (2, N-S, and E-W). We choose the MSE as error loss since it is a standard regression error loss and it is effective in penalizing large denoising errors during training. The training database consists of 60 000 samples, which are split into training (80%), validation (10%), and test (10%) sets.

We use the validation set to perform the hyperparameter selection, notably of the node embedding dimension d_N , the size of the hidden state of the recurrent unit h_R , and the embedding dimension of the spatiotemporal transformer h_T . We tested the following parameters for d_N, h_R , and h_T : [8, 16, 32, 64], [32, 64, 128, 256], [32, 64, 128, 256], respectively. The best parameters were selected as those corresponding to the lowest validation loss.

III. RESULTS

A. Overall Results on Synthetic Data

We compare SSEdenoiser against both traditional signal processing techniques and deep learning strategies. We adopt moving mean and median filtering as a baseline comparison, with different kernel sizes, here {3, 7, 15} days. We will refer to these models as **moving_avg_k** and **moving_med_k**, where k indicates the kernel size. Furthermore, we test the following two deep learning-based methods, which we will refer to as **single_station_RNN** and **2D_conv_u_net**.

- **single_station_RNN**. This model has been adapted from Xue and Freymueller (2023) [45] and it was originally intended as a detector of transient deformation signals in GNSS time series. We modify the last layer to match the size of the input for our denoising application. The model applies an RNN with a bidirectional LSTM (BiLSTM) cell to each station separately and it is followed by two fully connected layers. We keep the same parameters as

the original model, notably 64 for the BiLSTM embedding dimension and 16 for the fully connected layer.

- **2-D_conv_u_net**. This model works on GNSS time series stacked to form a matrix representation of dimensions (N_s, N_t) for each component, where the time series are sorted by latitude. The idea is to create a multistation method working on image-like data, since the rows represent stations that are close in space. To this end, we develop a standard U-Net architecture following its original formulation [46] and using four blocks (we will refer to this as a convolutional block) with 2-D convolution, ReLU activation, and batch normalization with [32, 64, 128, 256] feature maps for the contraction path. We interleave the convolutional blocks with 2-D max pooling layers. We replace the 2-D convolution and the 2-D max pooling with 2-D transposed convolution and upsampling layers for the expanding path, respectively. Other approaches could be used, such as image time series [36], possibly easing the processing in the spatial domain (e.g., by spatial convolutions), although dependent on the interpolation method, yet way more expensive in terms of training time and memory consumption.

We use the MSE and the mean absolute error (MAE) as error metrics, which we report along with their standard deviation. We use the MSE definition already presented in (13). We first define the squared error relative to the i th sample as

$$\text{SE}_i(\mathbf{d}(t), \hat{\mathbf{d}}(t)) = \sum_{j=1}^{N_s} \sum_{t=1}^{N_t} \sum_{k=1}^{N_d} \left(d_j^{i,k}(t) - \hat{d}_j^{i,k}(t) \right)^2. \quad (14)$$

The corresponding mean and standard deviation are computed as

$$\text{MSE}(\mathbf{d}(t), \hat{\mathbf{d}}(t)) = \frac{1}{n} \sum_{i=1}^n \text{SE}_i(\mathbf{d}(t), \hat{\mathbf{d}}(t)) \quad (15)$$

$$\sigma_{\text{SE}} = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\text{SE}_i(\mathbf{d}(t), \hat{\mathbf{d}}(t)) - \text{MSE}(\mathbf{d}(t), \hat{\mathbf{d}}(t)) \right)^2}. \quad (16)$$

The absolute error is defined similarly

$$\text{AE}_i(\mathbf{d}(t), \hat{\mathbf{d}}(t)) = \sum_{j=1}^{N_s} \sum_{t=1}^{N_t} \sum_{k=1}^{N_d} |d_j^{i,k}(t) - \hat{d}_j^{i,k}(t)| \quad (17)$$

$$\text{MAE}(\mathbf{d}(t), \hat{\mathbf{d}}(t)) = \frac{1}{n} \sum_{i=1}^n \text{AE}_i(\mathbf{d}(t), \hat{\mathbf{d}}(t)) \quad (18)$$

$$\sigma_{\text{AE}} = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\text{AE}_i(\mathbf{d}(t), \hat{\mathbf{d}}(t)) - \text{MAE}(\mathbf{d}(t), \hat{\mathbf{d}}(t)) \right)^2}. \quad (19)$$

We evaluate the denoising performance of the aforementioned models on a synthetic test set composed of 6000 samples and we present the numerical results in Table I. Traditional methods (mean and median filtering) are associated with a much higher error than deep learning-based methods. The error decreases as the kernel size increases, yet not significantly, probably because

TABLE I
DENOISING ERROR OF THE TESTED METHODS (MEAN \pm STANDARD DEVIATION) ON THE SYNTHETIC TEST SET

Model	MSE $\pm \sigma_{SE}$	MAE $\pm \sigma_{AE}$
moving_avg_3	5.7418 \pm 1.6430	1.7702 \pm 0.2903
moving_med_3	5.8487 \pm 1.6394	1.7800 \pm 0.2856
moving_avg_7	5.2399 \pm 1.6270	1.6870 \pm 0.3039
moving_med_7	5.1542 \pm 1.5988	1.6498 \pm 0.2982
moving_avg_15	4.8934 \pm 1.6120	1.6290 \pm 0.3141
moving_med_15	4.5036 \pm 1.5267	1.5080 \pm 0.3045
single_station_RNN	0.2382 \pm 0.3383	0.2136 \pm 0.1455
2D_conv_u_net	0.1724 \pm 0.4217	0.1313 \pm 0.1172
SSEdenoiser	0.0304 \pm 0.0558	0.0630 \pm 0.0565

The bold values indicate best performance.

the GNSS noise cannot be eliminated through simple mean or median filtering. This is particularly true for signals having a low signal-to-noise ratio, which are masked by the noise, where a mean or median filter, especially with a relatively large kernel, would not be able to extract the signal from the noise. We also observe that, when the kernel size increases, the error associated with the median filtering is lower than the mean one, probably because the mean filter has an intrinsic smoothing that increases with the kernel size, failing to preserve the amplitude of the denoised signal. Conversely, deep learning-based methods are associated with a much lower error, both absolute and MSE. Among the deep learning-based methods, the single-station approach outperforms the traditional methods, yet it is less precise than the multistation ones (2-D_conv_u_net and SSEdenoiser), probably because it cannot distinguish between signals that are registered by multiple stations as well as properly accounting for the spatial variability of the noise. As to the multistation approaches, SSEdenoiser achieves superior performance, with an average error ten times lower than 2-D_conv_u_net, thanks to the graph representation that helps in leveraging the network geometry as well as intra-station properties emerging from the data. In the 2-D_conv_u_net the convolution is applied blindly to stations that are close in latitude and that may not be necessarily more informative than a farther station. Hence, a graph turns out to be more accurate and flexible.

We also discuss the sensitivity of the models as a function of the signal-to-noise ratio of the input time series. We focus this analysis on the deep learning models only. First, we define the (average) signal-to-noise ratio for a multistation setting as the average signal-to-noise ratio over the stations that recorded a displacement for all components

$$\overline{\text{SNR}} = \frac{1}{|S'| \cdot N_c} \sum_{j \in S'} \sum_{k=1}^{N_c} 10 \log_{10} \left(\frac{\sum_{t=1}^{N_t} |c_j^k(t)|^2}{\sum_{t=1}^{N_t} |n_j^k(t)|^2} \right) \quad (20)$$

$$S' = \{j : d_j^k(t) \neq 0\},$$

$$\forall k \in (0, N_c), \forall t \in (0, N_t), \forall j \in (0, N_s)$$

where S' indicates a set of indices relative to the stations where a nonzero displacement occurred. Hence, stations that did not record any displacement are not taken into account. To compare the denoising performance, we define the denoising error as the MAE relative to the maximum amplitude displacement for the

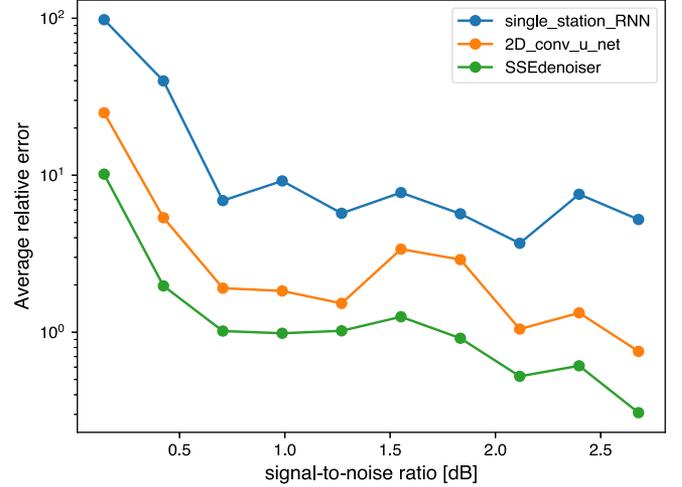


Fig. 3. Evaluation of the denoising power as a function of the signal-to-noise ratio for the tested deep learning models, namely single_station_RNN, 2-D_conv_u_net, and SSEdenoiser. The average absolute error [see (21) and (22)] for a given SNR bin is plotted.

i th sample

$$\mathcal{E}(i) = \frac{1}{N_t \cdot N_s' \cdot N_c} \sum_{j \in S'} \sum_{k=1}^{N_c} \frac{\sum_{t=1}^{N_t} |d_j^{i,k}(t) - \hat{d}_j^{i,k}(t)|}{\max_t |d_j^{i,k}(t)|} \quad (21)$$

and we compute the average denoising error as

$$\bar{\mathcal{E}} = \frac{1}{n} \sum_{i=1}^n \mathcal{E}(i). \quad (22)$$

This allows us to compare the denoising power of the models regardless of the displacement amplitude, so that statistics can be made. Fig. 3 shows the average denoising error as a function of the (binned) signal-to-noise ratio for the three deep learning-based models. In all models, the average error naturally decreases as the signal-to-noise ratio increases. Globally, SSEdenoiser performs the best among the three models, as also seen in Table I. For low values of signal-to-noise ratio (e.g., $\overline{\text{SNR}} < 1$), the three models behave differently and the performance is much more degraded for the single_station_RNN model, with an average error on the displacement which is 1.3 times higher than the one obtained for $\overline{\text{SNR}} > 1$. The other two models, namely 2-D_conv_u_net and SSEdenoiser, are more stable. For $\overline{\text{SNR}} < 1$, their average error is 0.4 and 0.1 times larger than for $\overline{\text{SNR}} > 1$, respectively. SSEdenoiser exhibits better performance and stability for all values of $\overline{\text{SNR}}$, with the average error for $\overline{\text{SNR}} = 0$ being four times lower than 2D_conv_u_net and 12 times lower than single_station_RNN.

B. Ablation Study

The novelty of SSEdenoiser lies in the combination of a graph-based RNN [39] with a spatiotemporal transformer module. We thus perform an ablation study to test the robustness of the proposed spatiotemporal transformer. We test SSEdenoiser against three ablated models, obtained by completely removing the transformer, by using temporal and spatial attention mechanisms separately, namely **no_transformer**, **spatial_attention_only**,

TABLE II
DENOISING ERROR (MEAN \pm STANDARD DEVIATION) ON THE SYNTHETIC
TEST SET FOR THE ABLATION STUDY

Model	MSE $\pm \sigma_{SE}$	MAE $\pm \sigma_{AE}$
no_transformer	0.0496 \pm 0.0738	0.0976 \pm 0.0626
spatial_attention_only	0.0429 \pm 0.0706	0.0790 \pm 0.0647
temporal_attention_only	0.0302 \pm 0.0560	0.0670 \pm 0.0554
SSEdenoiser	0.0304 \pm 0.0558	0.0630 \pm 0.0565

The bold values indicate best performance.

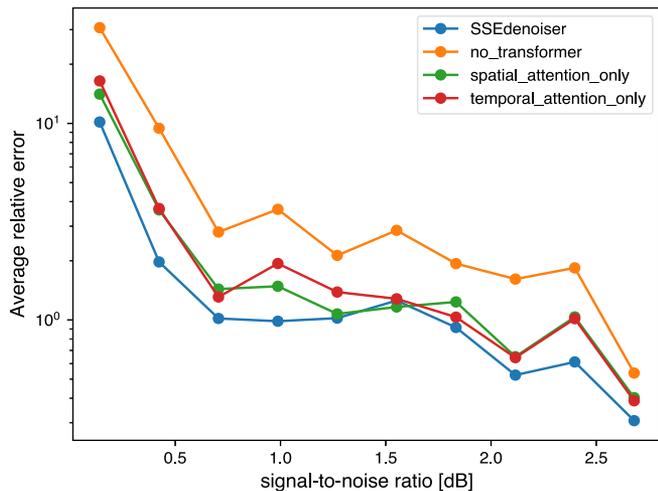


Fig. 4. Evaluation of the denoising power as a function of the signal-to-noise ratio for the ablated models, namely no_transformer, spatial_attention_only and temporal_attention_only, against SSEdenoiser. The average absolute error for a given SNR bin is plotted.

and **temporal_attention_only**. We train and test each ablated model with the same configuration as SSEdenoiser and we report the results in Table II and Fig. 4. Here, we do not test SSEdenoiser ablated from the graph-based RNN. The removal of this component must be supported by the addition of a spatiotemporal embedding, which is particularly challenging and, thus, is kept as future work.

Removing the transformer results in a higher global error, almost two times higher than that of SSEdenoiser. When adding the spatial attention, the global error decreases, with the model incorporating the temporal attention corresponding to the lowest misfit among the ablated models. Compared to SSEdenoiser, the temporal attention configuration exhibits the same squared error, yet a higher absolute error.

We further compare the four models by evaluating the denoising power (similarly to what is done in Section III-A), by computing the average relative error as a function of the signal-to-noise ratio, which can better highlight the superior denoising power of SSEdenoiser. The **no_transformer** model exhibits lower performance for any value of SNR, further motivating the need for incorporating an attention mechanism. The **spatial_attention_only** and **temporal_attention_only** have almost equivalent performance. The combination of both spatial and temporal attention (SSEdenoiser) results in higher performance at almost every scale of SNR, showing that, regardless of the noise level, the proposed model can have superior denoising power by coupling attention on the spatial and temporal axis,

which cannot be decoupled without resulting in a loss of resolution.

C. Results on Synthetic Data: Learned Station Connectivity

Our multistation approach also allows access to spatial and temporal relationships between GNSS stations for further interpretation. During training, SSEdenoiser learns which stations (nodes of the graph) should be connected and how strong their connections should be: this information is synthesized in the adjacency matrix, that provides the edge strength of the graph. In Fig. 5(a), we show the learned adjacency matrix, with nodes sorted by latitude (each pixel represents a relationship between two stations). The number of edges in the graph is $N + \frac{N(N-1)}{2} = 20100$ including self-loops. We see that SSEdenoiser has learned to connect nodes that are mostly spatially close to each other (near the diagonal), yet it also allows for weaker long-range connections (e.g., they are assigned a lower edge strength). This suggests that the method could generally rely on information available within a neighborhood and then compare information coming from different subnetworks. Moreover, this information is entirely learned from the data, since the model is not fed with any explicit information about the station location. Also, it is difficult to model station connections a priori, especially to find a meaningful balance between short- and long-range connections, because we need reference stations that are not moving to obtain the “absolute” value of the displacement associated with a given SSE. We tested several choices for the prior edge distribution, such as k-nearest-neighbors, k-neighbors in a given distance range, k-neighbors in an optimal azimuthal coverage. None of these gave satisfactory results, probably because precomputing long-range connections is not straightforward.

To visualize the backbone of the graph structure, we further filter the adjacency matrix by selecting edge strength values higher than 0.008, corresponding to the connections with strong edge weights. We found 878 connections (excluding the self-connections on the diagonal) to which we will refer in the following as “strong connections.” These connections represent 2.2% of the total connections in the graph (20 000, excluding the self-connections). We show their spatial distribution between GNSS stations in Fig. 5(b). We first see that these connections are such that the azimuthal coverage is as high as possible between neighboring stations. Also, the method has learned how to produce a mesh connecting all the stations that are located on top of the slow slip area used in the training phase [between 20 and 40 km depth, see Fig. 1(b)]. Stations located further inland (longitude $< 122^\circ$ W) do not have strong edge weights, suggesting that they are not very informative for slow slip detection given the location of the SSE area. Also, the stronger edges connect stations that are in areas where the station coverage is sparse. When the network is dense, the high weights are indeed less useful because the information can already be included in the signal coming from the numerous nearby stations.

Fig. 5(c) shows the 200 GNSS stations color-coded by the value of the adjacency matrix diagonal term. It indicates the strength of the self-loop connections for each station, which can be thought of as a measure of the learned station’s self-relevance

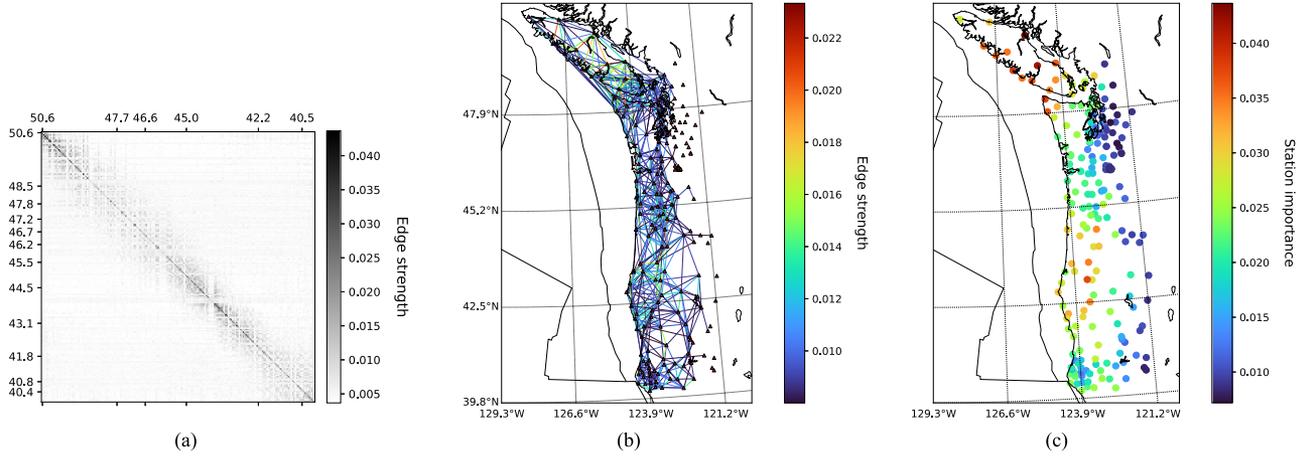


Fig. 5. Graph connectivity learned by SSEdenoiser. (a) Learned adjacency matrix. Nodes are sorted by latitude and color-coded by the learned edge strength. (b) Geographic representation of the strongest connections [edge strength values in the range (0.008, 0.0234) color-coded by edge strength]. (c) GNSS stations used in this study, color-coded by the learned station importance, i.e., the value of the diagonal of the adjacency matrix for each node.

(or importance). We can see that the importance of further inland stations is low compared to stations that are located above the SSE area. The highest importance values are assigned to stations located in Vancouver Island and at latitudes between 42.5°N and 45.2°N , probably linked to a learned tradeoff between slow slip occurrence and coverage of GNSS stations. This can be seen as a proxy of the distance to the SSE source area combined with the local density of stations.

It should be noted that the analysis of the edge weights should not be overinterpreted, since the “strong connections” are further modulated by the spatiotemporal transformer and the MLP modules, thus multiple nodes playing different roles in the graph could still express aggregated features that are not necessarily represented by the specific nodes. Also, the stations’ relevance should be taken as a qualitative result, since it does not directly reflect the station’s influence on the whole graph.

D. Denoising of Real Nonpostprocessed GNSS Time Series in Cascadia in 2007–2022

We test SSEdenoiser on real nonpostprocessed GNSS time series from 2007 to 2022. We take 60-day windows of data and apply SSEdenoiser on each of them, by sliding the window with a stride of one day and collecting the resulting denoised GNSS time series. In practice, nonoverlapping sliding windows can be used (stride of 60 days). Here, we prefer relying on multiple denoising outputs (stride of one day) to provide a better estimate of the displacement by averaging the contributions coming from all the possible sliding windows. We compute the temporal derivative of the estimated denoised time series in each window to obtain the displacement rate and we keep the 20 days in the middle of the window to exclude potential border effects. For each time step, we take the mean of all 20-day overlapping windows (20 windows). With this procedure, we obtain the daily average displacement rate.

Fig. 6(a) shows the obtained denoised GNSS time series of daily (E-W) displacement rates at all stations in matrix form, over the period 2007–2022. For simplicity, only displacement

rates larger than 0.01 mm/day are shown. The retrieved displacement rate has a coherent spatiotemporal distribution: it occurs by bursts that are clustered in latitude and time, consistent with previous studies [30], [49]. The amplitude is nearly always negative (meaning a displacement in the west direction), that is the direction of expected motion. The largest SSEs are associated with large displacement rates, such as the May 2011 or the January 2016 SSE [see zoom in Fig. 6(c)]. We see that SSEdenoiser is also able to constrain slow slip occurring in South Cascadia, which is more difficult than in the northern area because of the known lower signal-to-noise ratio.

The denoised displacement rates have a good correlation with the spatiotemporal distribution of tremors (shown in black in the figures). Tremors are low-frequency and low-amplitude seismic events that have been observed to accompany SSEs in Cascadia, e.g., [50]. SSEdenoiser is blindly trained on GNSS time series, without incorporating any information from tremors in the model: this means that our method is validated thanks to this independent information. The displacement distribution follows the tremor propagation in space and time, both for large and smaller SSEs as well as for propagating events, as can be seen in the zoom in Fig. 6(c). Our method can effectively retrieve slip migration, such as in the case of the 2016 or 2018 SSEs as well as recognizing events that are close in time and space, such as the Feb.–Apr. 2017 or June–Aug. 2018 events. This suggests that SSEdenoiser has effectively learned what the noise structure looks like to retrieve the concealed deformation associated with SSEs at a wide range of space and time scales, especially by comparing the denoised displacement with the raw data that was fed in input to the model, as we see in Fig. 6(b).

E. Comparison Between SSEdenoiser and the Single-Station Method by Xue and Freymueller (2023)

We qualitatively compare the denoised displacements computed from the output of SSEdenoiser with the results obtained by Xue and Freymueller (2023) [45]. They developed a single-station deep learning method [cf., Section III-A] whose output

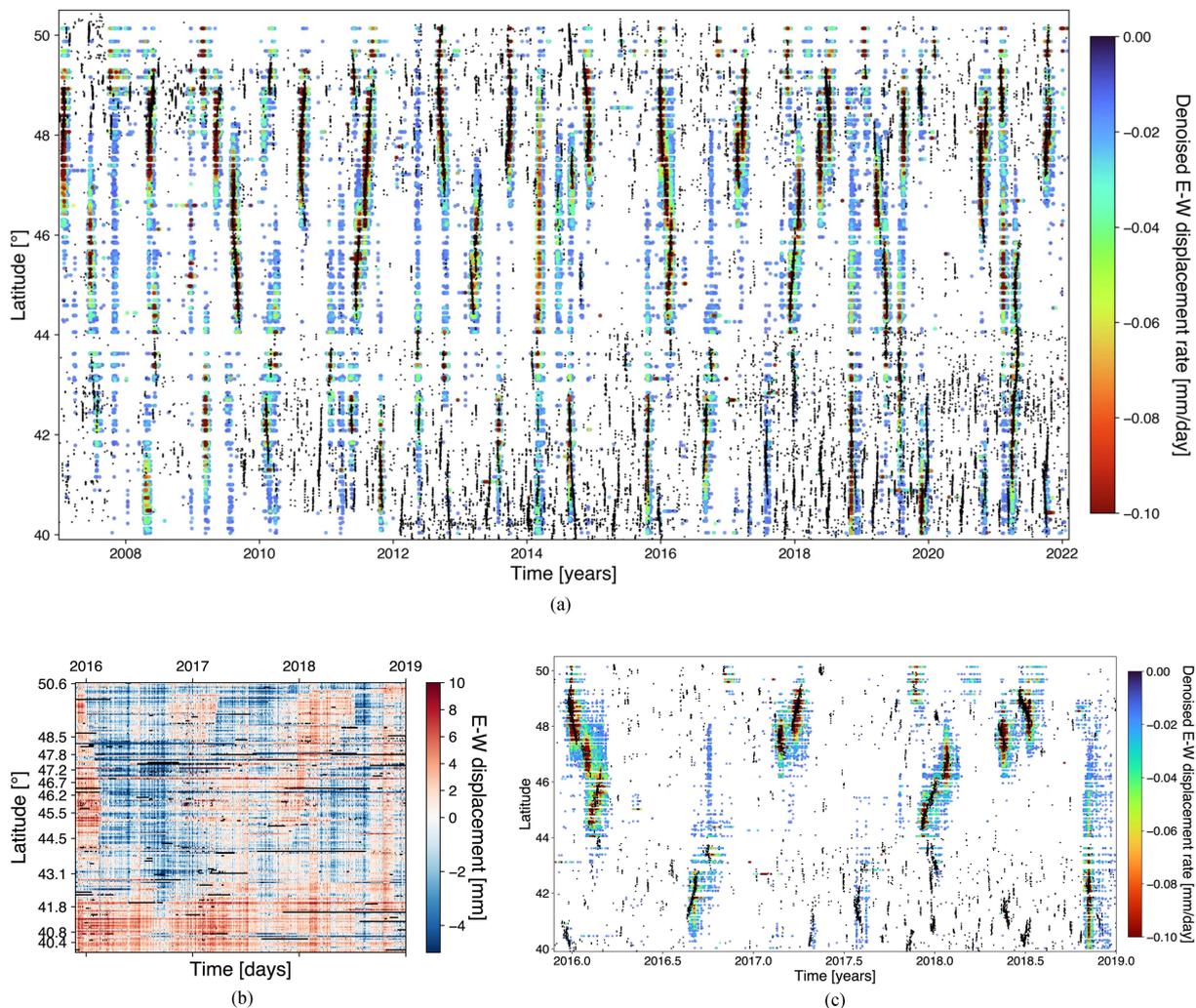


Fig. 6. Denoising of real GNSS time series in Cascadia from 2007 to 2022. (a) Denoised displacement rate (E-W component) as a function of time. The displacement rate computed from the output of SSEdenoiser is shown for each GNSS station, sorted as a function of the latitude. Tremor epicenters are also shown as black points (we use the catalogue from Ide, 2012 [47] until 5 August 2009, and the tremor catalogue from the Pacific Northwest seismic network in the following period [48]). (b) Example of raw GNSS data for the period 2016–2019. Each row of the matrix represents a detrended GNSS time series, color-coded by the amount of displacement in the E-W component. Stations are sorted by latitude. (c) Denoised displacement field computed from the output of SSEdenoiser from the same input time series as in the (b) panel. The displacement rate is shown by the colored points. The black points represent tremor epicenters. Denoised displacements are very well correlated with the occurrence of tremors, which is an expected feature that provides independent validation of the results.

is the probability of occurrence of an SSE for each separate GNSS time series. Fig. 7(a) shows the SSEdenoiser denoised displacements along with this SSE probability. There is, globally, good accordance between our denoised time series and the SSE probability obtained by Xue and Freymueller, which can be thought of as a proxy of the SSE-driven displacement.

Xue and Freymueller's prediction probability (in red) has a good correlation with tremor epicenters, yet it exhibits some level of noise and inconsistencies along the stations. For instance, the horizontal cluster of points at latitude 47°N seems to suggest that, in the case of perturbed time series (which seems likely the case), the ability to discern coherence in the signal is not properly constrained due to the absence of corroborative data from other stations. Also, another limitation inherent in single-station analysis is that these models tend to produce isolated points, as we see in Fig. 7. These isolated points are associated with a high probability, yet there is no agreement

from neighboring stations. Hence, these can be likely considered false positives.

Conversely, SSEdenoiser gains in accuracy thanks to the multistation approach. The denoised displacement output by SSEdenoiser follows the tremor distribution in space and time, suggesting that our approach has the potential to provide finer scale observations of slow slip by leveraging the spatial correlations between the GNSS time series as well as the GNSS geometry. Fig. 7(b) shows a zoom in the period ranging from August 2009 to December 2011. Our method seems to show a better ability to constrain the spatial consistency of the SSE-driven displacement, such as in the case of big events, e.g., the September 2009 SSE, where the probability of the Xue and Freymueller's approach does not exceed the detection threshold for all the stations at latitudes between 47°N and 48°N , as opposed to the displacement retrieved by SSEdenoiser, consistent with tremor recordings. The same seems to happen also for smaller

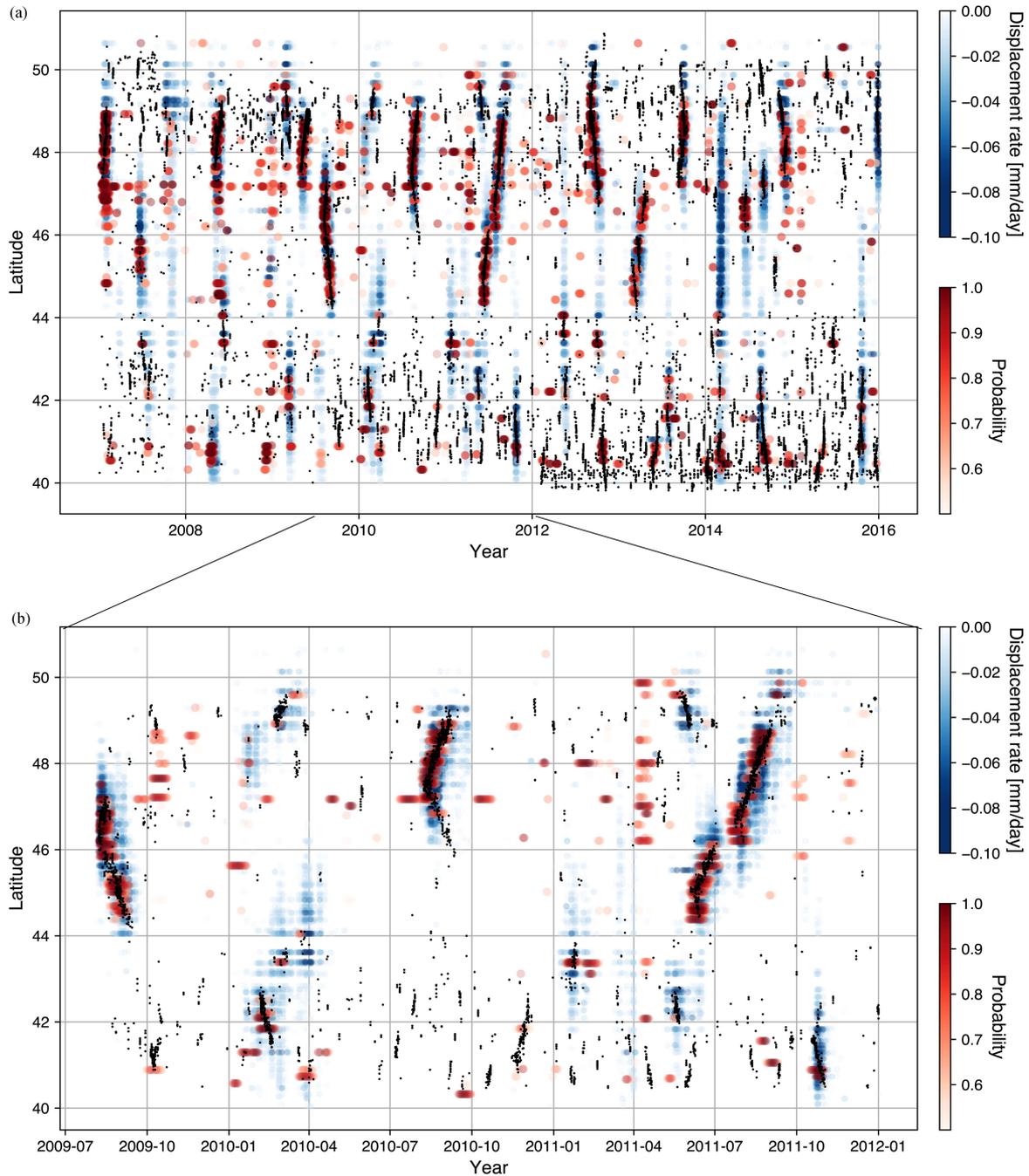


Fig. 7. (a) Qualitative comparison between SSEdenoiser and the single-station method by Xue and Freymueller (2023) [45] in Cascadia from 2007 to 2016. Black points represent tremor epicenters, sorted by latitude. Red points are associated with the probability of SSE occurrence by Xue and Freymueller [45] and blue points represent the denoised displacement field obtained by running SSEdetector on the raw data in Cascadia from 2007 to 2016. (b) Zoom in the period ranging from Aug. 2009 to Dec. 2011.

events, e.g., the November 2011 SSE (average latitude of 41°N) and the June 2011 event (latitude 47°N), where our method shows the potential of revealing slow slip with improved detail thanks to a multistation approach. Yet, SSEdenoiser also has some limitations, such as the latitudinal clusters [e.g., early 2014 or late 2018 in Fig. 6(c)], which can be ascribed to the fact that the synthetic training set may not reflect all the complexity of real data. These narrow latitudinal clusters are most probably the signature of GNSS common mode errors in the real data,

which might not be perfectly modeled by our noise generation strategy and need further improvement.

IV. CONCLUSION

We develop SSEdenoiser, a graph-based deep learning method for denoising nonpostprocessed GNSS position time series. We build a synthetic database, consisting of realistic noise and synthetic SSE signals. We first compare SSEdenoiser to

traditional denoising methods as well as single- and multistation-based deep learning models. We find that our method has superior accuracy and stability also for very low values of signal-to-noise ratio. We also analyze the characteristics of SSEdenoiser by looking at the learned adjacency matrix. We find that SSEdenoiser learns to connect stations based on the region where SSEs are located in the training samples and to the density of GNSS stations.

When tested on real data, SSEdenoiser proves effective in isolating the displacement related to SSEs, with remarkable spatial and temporal correlation with tremors, which are not given as input to the method. By using tremors as independent validation, we conclude that our method can denoise GNSS data and retrieve concealed transient deformation across a wide range of scales of both duration and signal amplitude.

We finally perform a qualitative test between SSEdenoiser and the single-station method by Xue and Freymueller (2023) [45] and we find that SSEdenoiser exhibits superior performance in terms of spatiotemporal distribution of the denoised displacement as well as consistency with tremor episodes in space and time.

Our approach also has the advantage that denoised time series could be used as input for traditional slip inversion techniques as well as to better constrain the source parameters. Since the level of noise in GNSS time series is such that regularization constraints must be used to guarantee temporal smoothness, our denoised time series could be also used for inverting the slip on the subduction interface more easily than the nondenoised data.

ACKNOWLEDGMENT

The author would like to thank the Editor, the Associate Editor, and the two anonymous reviewers for contributing to the peer review of this work.

REFERENCES

- [1] F. Luo, T. Zhou, J. Liu, T. Guo, X. Gong, and J. Ren, "MultiScale diff-changed feature fusion network for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5502713.
- [2] Y. Hong, R. F. Adler, and G. Huffman, "An experimental global prediction system for rainfall-triggered landslides using satellite remote sensing and geospatial datasets," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 6, pp. 1671–1680, Jun. 2007.
- [3] Y. Yi and W. Zhang, "A new deep-learning-based approach for earthquake-triggered landslide detection from single-temporal rapideye satellite imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 6166–6176, 2020.
- [4] M. A. Wagner, S. W. Myint, and R. S. Cerveny, "Geospatial assessment of recovery rates following a tornado disaster," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4313–4322, Nov. 2012.
- [5] L. Ding, H. Guo, S. Liu, L. Mou, J. Zhang, and L. Bruzzone, "Bi-temporal semantic reasoning for the semantic change detection in HR remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5620014.
- [6] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, Mar. 2005.
- [7] P. Ghamisi, J. A. Benediktsson, and S. Phinn, "Fusion of hyperspectral and lidar data in classification of urban areas," in *Proc. IEEE Geosci. Remote Sens. Symp.*, 2014, pp. 181–184.
- [8] A. Camps et al., "Sensitivity of GNSS-R spaceborne observations to soil moisture and vegetation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 10, pp. 4730–4742, Oct. 2016.
- [9] N. Pierdicca et al., "The potential of spaceborne GNSS reflectometry for soil moisture, biomass, and freeze–thaw monitoring: Summary of a European space agency-funded study," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 2, pp. 8–38, Jun. 2022.
- [10] K. Dai et al., "Entering the era of Earth observation-based landslide warning systems: A novel and exciting framework," *IEEE Geosci. Remote Sens. Mag.*, vol. 8, no. 1, pp. 136–153, Mar. 2020.
- [11] M. J. Unwin et al., "An introduction to the hydroGNSS GNSS reflectometry remote sensing mission," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 6987–6999, 2021.
- [12] J. L. Garrison and E. Cardellach, "The IEEE GNSS and signals of opportunity working group [Technical Committees]," *IEEE Geosci. Remote Sens. Mag.*, vol. 2, no. 4, pp. 54–58, Dec. 2014.
- [13] V. U. Zavorotny, S. Gleason, E. Cardellach, and A. Camps, "Tutorial on remote sensing using GNSS bistatic radar of opportunity," *IEEE Geosci. Remote Sens. Mag.*, vol. 2, no. 4, pp. 8–45, Dec. 2014.
- [14] X. Li et al., "High-rate GPS seismology using real-time precise point positioning with ambiguity resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6165–6180, Oct. 2014.
- [15] J.-P. Montillet, P. Tregoning, S. McClusky, and K. Yu, "Extracting white noise statistics in GPS coordinate time series," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 3, pp. 563–567, May 2013.
- [16] G. AlRegib, S. Fomel, and R. Lopes, "Subsurface exploration: Recent advances in Geo-signal processing, interpretation, and learning [from the guest editors]," *IEEE Signal Process. Mag.*, vol. 35, no. 2, pp. 16–18, Mar. 2018.
- [17] M. Malfante, M. Dalla Mura, J.-P. Métaxian, J. I. Mars, O. Macedo, and A. Inza, "Machine learning for volcano-seismic signals: Challenges and perspectives," *IEEE Signal Process. Mag.*, vol. 35, no. 2, pp. 20–30, Mar. 2018.
- [18] A. A. T. Peixoto et al., "Tensor-based learning framework for automatic multichannel volcano-seismic classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 4517–4529, 2021.
- [19] P. E. E. Lara et al., "Automatic multichannel volcano-seismic classification using machine learning and EMD," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 1322–1331, 2020.
- [20] M. López-Pérez, L. García, C. Benítez, and R. Molina, "A contribution to deep learning approaches for automatic classification of volcano-seismic events: Deep Gaussian processes," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 3875–3890, May 2021.
- [21] X. Chai, G. Tang, S. Wang, K. Lin, and R. Peng, "Deep learning for irregularly and regularly missing 3-D data reconstruction," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 6244–6265, Jul. 2021.
- [22] W. Zhang and J. Gao, "Deep-learning full-waveform inversion using seismic migration images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5901818.
- [23] X. Lin, C. Xu, G. Jiang, and J. Zang, "A hybrid deep learning model for rapid probabilistic earthquake source parameter estimation with displacement waveforms from a flexible set of seismic or HR-GNSS stations," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5923416.
- [24] M. K. Shahvandi and B. Soja, "Modified deep transformers for GNSS time series prediction," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 8313–8316.
- [25] P. K. Enge, "The global positioning system: Signals, measurements, and performance," *Int. J. Wireless Inf. Netw.*, vol. 1, pp. 83–105, 1994.
- [26] S. Wdowinski, Y. Bock, J. Zhang, P. Fang, and J. Genrich, "Southern California permanent GPS geodetic array: Spatial filtering of daily positions for estimating coseismic and postseismic displacements induced by the 1992 landers earthquake," *J. Geophys. Res.: Solid Earth*, vol. 102, no. B8, pp. 18057–18070, 1997.
- [27] S. D. Williams et al., "Error analysis of continuous GPS position time series," *J. Geophys. Res.: Solid Earth*, vol. 109, no. B3, 2004.
- [28] L. Marill, D. Marsan, A. Socquet, M. Radiguet, N. Cotte, and B. Rousset, "Fourteen-year acceleration along the Japan trench," *J. Geophys. Res.: Solid Earth*, vol. 126, no. 11, 2021, Art. no. e2020JB021226.
- [29] J. Bedford and M. Bevis, "Greedy automatic signal decomposition and its application to daily GPS time series," *J. Geophys. Res.: Solid Earth*, vol. 123, no. 8, pp. 6992–7003, 2018.
- [30] S. Michel, A. Gualandi, and J.-P. Avouac, "Interseismic coupling and slow slip events on the Cascadia megathrust," *Pure Appl. Geophys.*, vol. 176, no. 9, pp. 3867–3891, 2019.

- [31] O. M. Saad and Y. Chen, "Deep denoising autoencoder for seismic random noise attenuation," *Geophysics*, vol. 85, no. 4, pp. V367–V376, 2020.
- [32] W. Zhu, S. M. Mousavi, and G. C. Beroza, "Seismic signal denoising and decomposition using deep neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 9476–9488, Nov. 2019.
- [33] A. Thomas, D. Melgar, S. N. Dybing, and J. R. Searcy, "Deep learning for denoising high-rate global navigation satellite system data," *Seismica*, vol. 2, no. 1, 2023, Art. no. 240.
- [34] J. Münchmeyer, D. Bindi, U. Leser, and F. Tilmann, "Earthquake magnitude and location estimation from real time seismic waveforms with a transformer network," *Geophys. J. Int.*, vol. 226, no. 2, pp. 1086–1104, 2021.
- [35] M. P. van den Ende and J.-P. Ampuero, "Automated seismic source characterization using deep graph neural networks," *Geophys. Res. Lett.*, vol. 47, no. 17, 2020, Art. no. e2020GL088690.
- [36] G. Costantino et al., "Seismic source characterization from GNSS data using deep learning," *J. Geophys. Res.: Solid Earth*, vol. 128, no. 4, 2023, Art. no. e2022JB024930. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2022JB024930>
- [37] G. Costantino, S. Giffard-Roisin, M. Radiguet, M. Dalla Mura, D. Marsan, and A. Socquet, "Multi-station deep learning on geodetic time series detects slow slip events in Cascadia," *Commun. Earth Environ.*, vol. 4, no. 1, 2023, Art. no. 435.
- [38] B. Yu et al., "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," in *Proc. 27th Int. Joint Conf. Artif. Intell. Org.*, 2018.
- [39] L. Bai et al., "Adaptive graph convolutional recurrent network for traffic forecasting," in *Proc. Int. Conf. Adv. Neural Inf. Process. Syst.*, 2020, pp. 17804–17815.
- [40] L. Shi et al., "Two-stream adaptive graph convolutional networks for skeleton-based action recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 12018–12027.
- [41] E. Rossi, B. Chamberlain, F. Frasca, D. Eynard, F. Monti, and M. Bronstein, "Temporal graph networks for deep learning on dynamic graphs," 2020, *arXiv:2006.10637*.
- [42] A. Casteigts, P. Flocchini, W. Quattrociocchi, and N. Santoro, "Time-varying graphs and dynamic networks," *Int. J. Parallel, Emergent Distrib. Syst.*, vol. 27, no. 5, pp. 387–408, 2012.
- [43] Y. Okada, "Surface deformation due to shear and tensile faults in a half-space," *Bull. Seismological Soc. America*, vol. 75, no. 4, pp. 1135–1154, 1985.
- [44] A. Vaswani et al., "Attention is all you need," in *Proc. 31st Int. Conf. Adv. Neural Inf. Process. Syst.*, 2017, vol. 30, pp. 6000–6010.
- [45] X. Xue and J. T. Freymueller, "Machine learning for single-station detection of transient deformation in GPS time series with a case study of Cascadia slow slip," *J. Geophys. Res.: Solid Earth*, vol. 128, no. 2, 2023, Art. no. e2022JB024859.
- [46] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, Munich, Germany, Oct. 5–9, 2015, pp. 234–241.
- [47] S. Ide, "Variety and spatial heterogeneity of tectonic tremor worldwide," *J. Geophys. Res.: Solid Earth*, vol. 117, no. B3, 2012.
- [48] A. G. Wech and K. C. Creager, "Automated detection and location of Cascadia tremor," *Geophys. Res. Lett.*, vol. 35, no. 20, 2008.
- [49] D. Schmidt and H. Gao, "Source parameters and time-dependent slip distributions of slow slip events on the Cascadia subduction zone from 1998 to 2008," *J. Geophys. Res.: Solid Earth*, vol. 115, no. B4, 2010.
- [50] G. Rogers and H. Dragert, "Episodic tremor and slip on the Cascadia subduction zone: The chatter of silent slip," *Science*, vol. 300, no. 5627, pp. 1942–1943, 2003.



Giuseppe Costantino received the B.Sc. and M.Sc. degrees in information engineering and applied mathematics from the University of Salerno, Fisciano, Italy, in 2017 and 2020, respectively, and the Ph.D. degree in geophysics from Université Grenoble Alpes, Grenoble, France, in 2023.

He is currently a Postdoctoral Researcher with Université Grenoble Alpes. His research focuses on studying the spatiotemporal complexity of the seismic cycle through the lens of artificial intelligence.

Sophie Giffard-Roisin received the Ph.D. degree in computer science from Université Côte d'Azur in Inria Sophia Antipolis, Nice, France, in 2017.

He was as a Postdoctoral Researcher with the University of Colorado, Boulder, CO, USA. He was an IRD Permanent Researcher with ISTERre laboratory, Université Grenoble Alpes, Grenoble, France. His research focuses on the development of machine learning models for natural hazards and remote sensing applications.



Mauro Dalla Mura (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees in telecommunication engineering from the University of Trento, Trento, Italy, in 2005 and 2007, respectively, and the joint Ph.D. degree in information and communication technologies (telecommunications area) from the University of Trento, and in electrical and computer engineering from the University of Iceland, Reykjavik, Iceland, in 2011.

In 2011, he was a Research Fellow with Fondazione Bruno Kessler, Trento, Italy, conducting research on computer vision. Since 2012, he has been an Associate Professor with the Grenoble Institute of Technology (Grenoble INP), Grenoble, France. He is conducting his research at the Grenoble Images Speech Signals and Automatics Laboratory (GIPSA-Lab). From 2019 to 2022, he was a Specially Appointed Associate Professor with the School of Computing, Tokyo Institute of Technology, Meguro, Japan. From 2021 to 2026, he is a Junior Member with the Institut Universitaire de France, Paris, France. His research interests include the fields of remote sensing, computational imaging, image and signal processing.

Dr. Mura was the recipient of the IEEE GRSS Second Prize in the Student Paper Competition of the 2011 IEEE IGARSS 2011 and co-recipient of the Best Paper Award of the International Journal of Image and Data Fusion for the year 2012–2013 and the Symposium Paper Award for IEEE IGARSS 2014. He was the IEEE GRSS Chapter's Committee Chair for 2020–2021. He was President of the IEEE GRSS French Chapter 2016–2020 (he previously served as Secretary 2013–2016). In 2017, he was the recipient of the IEEE GRSS Chapter Award and the "Chapter of the year 2017" from the IEEE French Section. He is on the Editorial Board of IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (J-STARS) since 2016.



Anne Socquet received the B.S. degree in mathematics from the University Blaise Pascal, Clermont Ferrand, France, in 1997, the M.S. degree from Ecole Normale Supérieure, Paris, France, in 2000, and the Ph.D. degree from Paris University, Paris, France, in 2003, both in geophysics.

She was a Postdoctoral Researcher with TU Delft, Delft, The Netherlands, and with the University of California Los Angeles, Los Angeles, CA, USA. In 2006, she was an Associate Professor with Institut de Physique du Globe de Paris, Paris, France. In 2010, she was a Visiting Researcher with the University of Chile, Santiago, Chile. In 2011, she was with the University of Grenoble Alpes, Grenoble, France, where she is currently a Full Professor. In 2019, she was a Fulbright Fellow from the Franco-American Commission and an Invited Miller Professor with the University of California, Berkeley, Berkeley, CA, USA. She has authored 75 publications. She has been the Deputy Director of the Institute of Earth Sciences, Grenoble, France. Her research focuses on a better understanding of fault processes through the seismic cycle. She uses space geodesy to monitor the ground deformation and the space–time variations of the Earth's surface deformation.

Dr. Socquet was a recipient of the Consolidator grant from European Research Council in 2019. She is an Associate Editor for the *Journal of Geophysical Research-Solid Earth and Geochemistry*, *Geophysics*, *Geosystems (G-Cubed)* from the American Geophysical Union (AGU).