doi:10.26443/seismica.v5i2.1775

SEISMICA

Generative model to improve the physical simulation of normalized seismic acceleration

Gottfried Jacquet * [1], **Filippo Gatti** [1], **Didier Clouteau** [1]

[1]Université Paris Saclay, CentraleSupélec, CNRS, ENS Paris-Saclay, Laboratoire de Mécanique Paris-Saclay UMR 9026, Gif-sur-Yvette, France

Author contributions: *Conceptualization*: Gottfried Jacquet. *Investigation*: Gottfried Jacquet. *Writing - Original Draft Preparation*: Gottfried Jacquet. *Visualization*: Gottfried Jacquet, Filippo Gatti. *Data Curation*: Filippo Gatti. *Methodology*: Filippo Gatti. *Validation*: Filippo Gatti. *Writing - Review & Editing*: Filippo Gatti. *Formal Analysis*: Didier Clouteau. *Project Administration*: Didier Clouteau.

Abstract Earthquake physics-based simulations have revealed their accuracy limit in high-fidelity broadband strong ground motion scenario prediction. Earlier attempts leveraging AI tried to fill the gap between the low-frequency high-fidelity numerical simulation and the high-frequency accuracy and realism demanded in earthquake engineering. Our research is built on top of the SeismoALICE solution proposed by Gatti and Clouteau (2020), an AI generative approach that renders broadband (0-20 Hz) single-station accelerograms conditioned by low-frequency physics-based simulation outcomes. First, we developed a novel neural architecture based on Encoder, Decoder, and Discriminators that integrates Conformer's advanced attention techniques (Gulati et al., 2020), stabilising the training and producing realistic output for the generation. This novel architecture is trained according to Adversarial Learning Inference with Conditional Entropy (ALICE, Li et al., 2017). Second, our approach employs a similarity evaluation technique called Hyper-Spherical Loss (HSL) in the time domain and an adaptation of the Focal Frequency Loss (FFL) for time series. Our investigation demonstrates that Conformer architecture outperforms previous approaches in super-resolution of earthquake data, such as the STanford EArthquake Dataset (STEAD) Dataset (Mousavi et al., 2019). We finally showcase the enhanced strong motion synthesizer to predict the seismic response at the Cruas Nuclear Power Plant during the 2019 M_W 4.9 Le Teil earthquake, providing insightful perspective for future large-scale applications as a downstream generative pipeline.

Production Editor:
Andrea Llenos
Handling Editor:
Carmine Galasso
Copy & Layout Editor:
Hannah F. Mark

Received:
July 3, 2025
Accepted:
June 27, 2026
Published:
September 2, 2026

1 Introduction

In the last two decades, increasingly available parallel computing architectures enabled seismologists and earthquake engineers to perform impressive numerical simulations both at regional and global scales (Ichimura et al., 2021), directly including the Soil-Structure Interaction (SSI, see McCallen et al., 2021a,b).

However, in numerical soil-structure interaction studies, in order to fully characterize the earthquake-induced transient and modal response of structures and infrastructures, the realism of the synthetic seismograms adopted as input ground motion must reach 20 Hz for residential buildings and up to 30-40 Hz for critical structures (Paolucci et al., 2014; Castro-Cruz et al., 2021). Despite the use of high-fidelity numerical software, such as SEM3D[1] (Touhami et al., 2022), such simulations are generally accurate in a 0-10 Hz frequency band because of the lack of data on the underground geological features and the seismic context of the region at stake (namely, the characteristics of active faults). The sensitivity of high-fidelity physics-based earthquake simulation engines, such as SEM3D, to geological and seismological characteristics is extremely high. A large uncertainty on such elements can negatively impact the reliability of the predicted ground motion at the surface. Even if the uncertainty about the underground features of the Earth's crust could be reduced by extensive and expensive seismic tomography and inversion studies in the region of interest, a challenge remains. The computational burden of achieving accurate solutions in a large frequency band, such as 0-30 Hz, is considerable. Accurate SEM3D simulations at this scale would require millions of CPU hours. However, 0-30 Hz-accurate earthquake synthetic ground motion is essential for designing critical infrastructure, such as nuclear power plants.

The inherent computational burden and experimental cost of such a brute-force deterministic approach are prohibitive. Therefore, reduced-order and/or surrogate modeling of cumbersome earthquake simulations are needed, with probabilistic modeling of the geological and seismological parameters at stake (see, for instance, Sochala et al., 2020; Pitarka and Mellors, 2021; Rekoske et al., 2023-08; Lehmann et al., 2023). Alternatively, hybrid modeling may employ numerical simulations at low frequency and empirical Green's functions stochastic simulations (Kamae et al., 1998; Pitarka and Mellors, 2021; Smerzini and Villani, 2012; Castro-Cruz et al., 2021), or data-driven approaches (Paolucci et al., 2018; Gatti and Clouteau, 2020; Florez et al., 2020) at high frequency. More recently, diffusion models have been employed for the sake of improving the realism of numerical simulations at high frequencies (Bergmeis-

*Corresponding author: jacquetgottfried@gmail.com
[1]https://github.com/sem3d/sem

ter et al., 2024; Bi et al., 2025; Gabrielidis et al., 2025; Nakata et al., 2025; Esfahani et al., 2023; Aquib and Mai, 2024).

In particular, this work focuses on the last data-driven approach to generate realistic broadband ground motion seismograms from a Gaussian latent manifold. Inspired by the pioneering work made by Gatti and Clouteau (2020), the physics-based low-frequency synthetic ground motion data $\mathbf{x}$ (represented by three-component acceleration time histories, numerically simulated) is encoded into a Gaussian latent space $\mathbf{z}_x = F_x(\mathbf{x})$, which is then mapped into another Gaussian manifold $\mathbf{z}_y = F_z(\mathbf{z}_x)$ (via a deep generative auto-encoder) and then decoded back to the space of broadband acceleration traces $\tilde{\mathbf{y}} = G_y(F_z(F_x(\mathbf{x})))$ that represents a hybrid surrogate of recorded signals. Hereafter, we compare this strategy with two others issued from the well-known image super-resolution task. The latter framework is applied to physics-based numerical simulations (analogous to low-resolution images) to generate realistic yet synthetic hybrid seismograms (the super-resolution counterpart of low-resolution images). The latter results from mapping a low-frequency signal – issued from physics-based simulation and accurate within a 0-1 Hz frequency band – into a broadband realistic signal in the 0-30 Hz frequency range. The deep learning training task is constrained by the inductive bias of the earthquake physics, integrated into the low-frequency input issued from deterministic numerical simulations. The deep learning mapping learns to render hybrid signals that preserve such an *a priori* knowledge of the earthquake scenario. This approach can be rephrased as a signal-to-signal translation problem.

In this paper, we present the methodology employed to tackle the intricacies of signal-to-signal translation for the sake of enhancing the realism of physics-based earthquake simulations to *translate* them into broadband time histories that can be considered as recorded seismographs of real seismic events. The major contributions of this paper are listed as follows:

- Introduction of a novel architecture, Generators and Discriminators, to support time series generation with Adversarial training.
- Use of the more adapted combination of loss functions for time series data.
- Manipulation of the latent bottleneck to better extract common information from broadband and filtered data.
- Definition of a more suitable architecture to make the one-to-many mapping of synthetic broadband data from filtered data, while guaranteeing the one-to-one mapping between physics-based simulation and target broadband.
- Proof of concept made on a nuclear power plant earthquake simulation.

In Section 3, the limitations of the strategy proposed by Gatti and Clouteau (2020) are highlighted to motivate the novel approach proposed in this paper. In Section 4, we outline the proposed improvements in terms of learning strategy. In Section 5 we revise the neural network architecture adopted for the low-to-high frequency signal-to-signal translation. In Section 6, we showcase the improved realism of the generated signals compared to Gatti and Clouteau (2020). Finally, in Section 7, we present the outcome of our new approach when tested on a real test case scenario, i.e., the simulated ground motion response of the 2019 M_W4.9 Le Teil earthquake.

2 Datasets

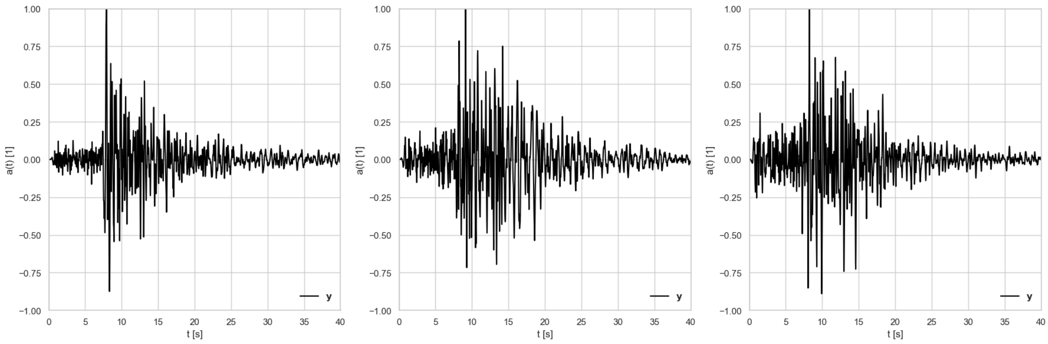
In the case of our study, we used the STEAD dataset (Mousavi et al., 2019), which brings together all the earthquakes that have occurred around the world. STEAD contains 85GB of local acceleration earthquakes and noise, and metadata on peak ground acceleration (PGA), depth, magnitude, P-wave arrival time, and S-wave arrival time. In the case of our research, we will concentrate on acceleration. The accelerations represent a 3D signal, each dimension of which is a direction as such, North-South, East-West and Top-Bottom. A representation of a sample of these dataset is presented in the Figure 1. We concentrate our signal from the arrival of the P-wave and the S-wave and stop at 4096 milliseconds later. In the case of our studies, we will restrict ourselves to normalized signals. This generalization has also been made by other researchers in the field (Gabrielidis et al., 2025).

The training dataset encompasses a broad spectrum of seismic scenarios. As shown in Figures 1, 1c, the M_W magnitude distribution ranges from 3.5 to 7.9, with a mean of 4.0 and a median of 3.8. This range ensures that the model is primarily trained on well-recorded local events, while also incorporating higher-magnitude transients. Source depths are concentrated among crustal events (median depth $\approx$ 15 km), with a distribution tail extending to deep-focus earthquakes (greater than 300 km). The PGA distribution, represented on a $\log_{10}$ scale, exhibits a Gaussian profile centered at approximately $10^{4.5}\mathrm{m/s}^2$. This confirms that the dataset spans several orders of magnitude in ground motion intensity. Such statistical diversity establishes a robust foundation for the generator to learn generalized mappings between low-frequency physical processes and broadband stochasticity.

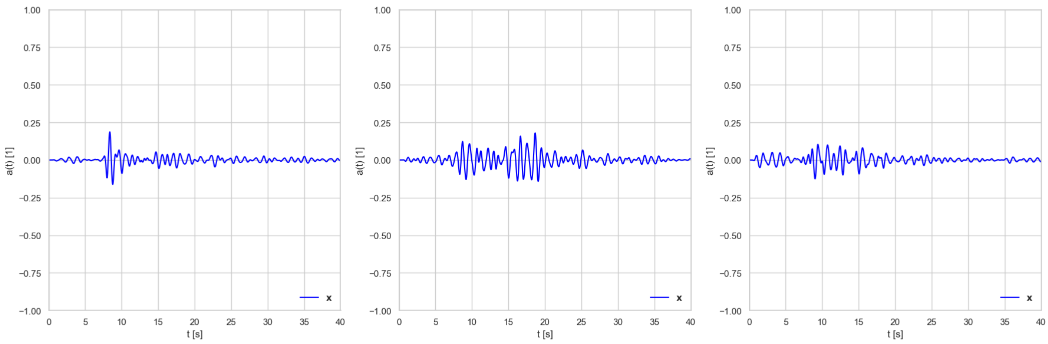
3 Related works

Gatti and Clouteau (2020) introduced the concept of Adversarial Learning Inference with Conditional Entropy in the seismic research field through the network SeismoALICE[2]. As of now, recorded earthquakes are adopted instead (referred as to $\mathbf{y}$, Gatti and Clouteau, 2020). However, some regions worldwide, such as metropolitan France, do not have large earthquake catalogues and associated seismograms over extended regions and do not have sufficient resolution (*i.e.*, a recording station every 100 m). Therefore, Gatti and

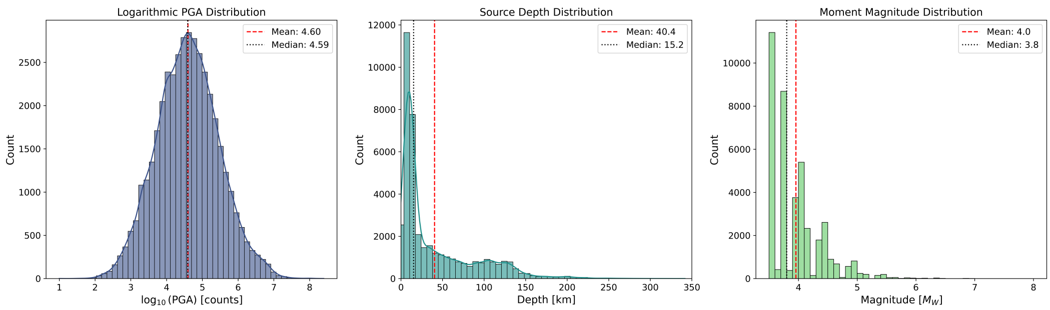
[2] https://github.com/FilLTP89/SeismoALICE



(a) Broadband signal in 3D, $\mathbf{y} \in \mathbb{R}^3 \times \mathbb{R}^{4096}$.



(b) Physics-based filtered signal in 3D, $\mathbf{x} \in \mathbb{R}^3 \times \mathbb{R}^{4096}$.



(c) Statistical distributions of the training dataset ($N = 51\,200$). Left: $\log_{10}$ of the maximum Peak Ground Acceleration (PGA) across all channels. Center: Source depth distribution (km), showing a dominance of shallow crustal events. Right: Moment magnitude (M_W) distribution. Dashed lines represent the mean, and dotted lines represent the median of each distribution.

Figure 1 Exploration of the Training Dataset.

Clouteau (2020) attempted to transform simulations 0-5 Hz $\mathbf{x}$ into realistic 0-30 Hz using ALICE. They split the problem into three parts: $\mathcal{P}_x$, to learn the latent features $\mathbf{z}_x$ of $\mathbf{x}$; $\mathcal{P}_y$, to learn the latent features $\mathbf{z}_y$ of $\mathbf{y}$; and $\mathcal{P}_z$, to bridge the gap between physics-based simulations $\mathbf{x}$ and recorded data $\mathbf{y}$, through their latent representations $\mathbf{z}_x$ and $\mathbf{z}_y$.

To apply ALICE to the super-resolution problem in earthquake engineering, Gatti and Clouteau (2020) subdivided the learning problem into three sub-problems. Two hidden variables ($\mathbf{z}_x$ and $\mathbf{z}y$) are introduced, for which the distribution is a mathematically well-known Gaussian or uniform. To transform the seismic signal to a physics-based simulation and vice versa, Gatti and Clouteau (2020) had to find a reduced abstract representation of each data and define a bridge between them, as shown in Figure 2.

The first problem, referred to as to $\mathcal{P}_x$, consists of finding a coupled pair of Decoder $G_x : \mathbb{R}^r \longrightarrow \mathbb{R}^d$ and Encoder $F_x : \mathbb{R}^d \longrightarrow \mathbb{R}^r$ for physics-based numerical time-histories, such a way that the couple $(\mathbf{z}_x, G_x(\mathbf{z}_x))$ and $(F_x(\mathbf{x}), \mathbf{x})$ are identically distributed. As in the ALICE framework, two discriminators were adopted to drive this learning task. The second problem, $\mathcal{P}_y$, consists in finding a coupled pair of Decoder $G_y : \mathbb{R}^r \longrightarrow \mathbb{R}^d$ and Encoder $F_y : \mathbb{R}^r \longrightarrow \mathbb{R}^d$ that match the distribution of $(\mathbf{z}_y, G_y(\mathbf{z}_y))$ and $(F_y(\mathbf{y}), \mathbf{y})$, corresponding to recorded seismograms. Finally, Gatti and Clouteau (2020) introduced the problem $\mathcal{P}_z$, in which we have to find $G_z : \mathbb{R}^r \longrightarrow \mathbb{R}^R$ and $F_z : \mathbb{R}^R \longrightarrow \mathbb{R}^r$ to map the latent variable corresponding to $\mathbf{x}$ into $\mathbf{y}$: $(\mathbf{x}, G_y \circ G_{z_y} \circ F_x(\mathbf{x}))$ and $(G_x \circ G_{z_x} \circ F_y(\mathbf{y}), \mathbf{y})$. So, the composed Encoder and Decoder read $F = G_x \circ G_{z_x} \circ F_x$ and $G = G_y \circ G_{z_y} \circ F_x$ respectively. All G_i and F_i mentioned above were modeled through deep CNN, as detailed in the original publication (Gatti and Clouteau, 2020).

The authors adopted a Wasserstein GAN (WGAN) framework, as described by Arjovsky and Bottou (2017). The idea was to force synthetic data $\mathbf{x}$, i.e. $G_y \circ G_z \circ F_x(\mathbf{x})$, to resemble $\mathbf{y}$, forcing the encoders and decoders to mimic a simple distribution close to the $p(\mathbf{y})$ distribution. In this study we also adopt a similar ap-

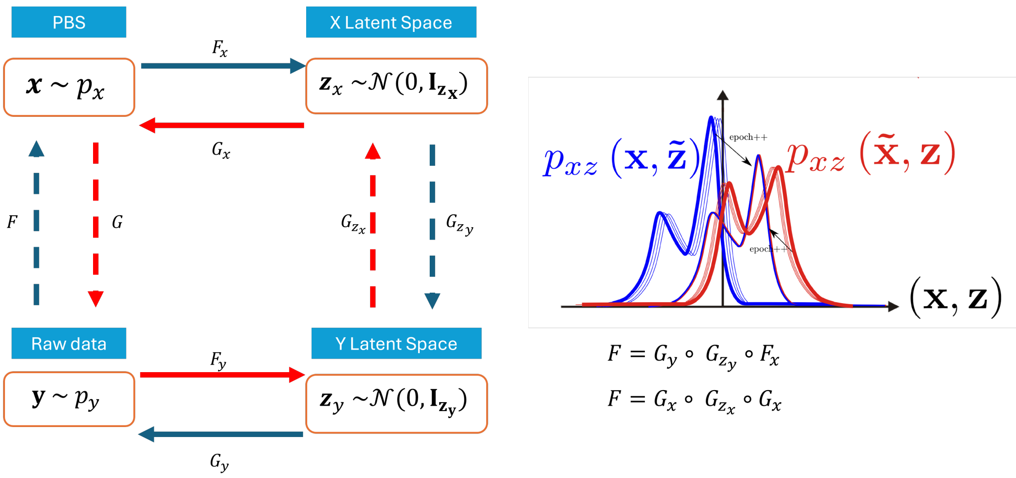


Figure 2 Resume of the SeismoALICE framework to pass from physics-based signal to ground motion.

proach by forcing G_x, F_z, and F_z to learn the mapping between $\mathbf{x}$ and $G_x \circ F_z \circ F_y(\mathbf{y})$. $\mathbf{z}_x$ is the latent space adapted that should capture the relevant features of $\mathbf{x}$, as should $\mathbf{z}_y$ for $\mathbf{y}$, as illustrated in Figure 2.

3.1 Limitation of this strategy

The SeismoALICE method, applied in a tripartite manner, faces limitations due to its sequential nature.

In its original formulation, SeismoALICE duplicates the number of parameters, because of the number of discriminators, namely D_x and D_x^h, D_y and D_x^h, and D_z and D_z^h (with these definitions $D_x : \mathbb{R}^d \to \mathbb{R}$, $D_x^h : \mathbb{R}^d \to \mathbb{R}$, $D_y : \mathbb{R}^d \to \mathbb{R}$, $D_y^h : \mathbb{R}^d \to \mathbb{R}$). Calibration of these discriminators for generators is time-consuming and requires rather intricate training hacks, as discussed by Lee and Seok (2020). The interplay between physics-based and seismic data introduces unnecessary parameters, thereby complicating the pursuit of optimal solution searches. In the context of adversarial learning, training is framed as a zero-sum game between two neural networks (a generator and a discriminator). The objective is to reach a Nash equilibrium among them: a convergence state where the generator learns to produce realistic accelerograms that the discriminator can no longer distinguish from real records (effectively guessing with 50% probability). Although it is clear in theory, achieving such an equilibrium in practice is numerically challenging and unstable, often leading to *vanishing gradient* or mode collapse issues, which complicate the optimization of seismic time series.

(i) **Interpretability of the latent spaces**
A critical drawback is found in the final layer of Generators when CNN is adopted as a neural network (see Gatti and Clouteau, 2020). This layer struggles to capture broad contextual information, diminishing overall model performance. The network's limited ability to project data into a Gaussian distribution preserves data correlations, necessitating exploration of alternative layer structures to overcome these constraints.

(ii) **Multi-objective optimization conflict**
While the loss of adversarial learning inference minimizes the distribution discrepancy between real and generated data, as demonstrated by Li et al. (2017), it is useful but not enough to guarantee the preservation of intricate data structure. Such issues are found when manipulating the cycle consistency loss, which is useful for mapping data. There are two ways to perform cycle consistency: explicitly and implicitly. The explicit cycle consistency corresponds to the ℓ_k norm for reconstruction, whereas the implicit cycle consistency learns reconstruction via adversarial learning. Here is a formulation of the implicit loss providing a reconstruction score for the reconstruction:

$$\mathcal{L}_{\text{imp.}} = \mathbb{E}_{x \sim p_{\text{data}}} [\log(D_{xx}(\mathbf{x}, \mathbf{x})) + \log(1 - D_{xx}(\mathbf{x}, G_x \circ F_x(\mathbf{x})))] \quad (1)$$

with D_{xx} is the Discriminator that evaluates the pair of distributions, $D_{xx} : \mathbb{R}^{2d} \to \mathbb{R}$. A probability is given with a score going [0-1].

The expression of the explicit loss is :

$$\mathcal{L}_{\text{exp.}} = \mathbb{E}_{x \sim p_{\text{data}}} \left[\|\mathbf{x} - G_x \circ F_x(\mathbf{x})\|_1^2 \right] \quad (2)$$

Theoretically, according to Li et al. (2017), implicit consistency is equivalent to explicit consistency. However, cycle consistency loss can lead to conflict with objective optimization: the improvement of one objective worsening the other. This situation does not always preserve consistency in the context of ALICE (Belghazi et al., 2021). Implicitly imposed cycle consistency leads to a deterioration of the data reconstruction, although it improves unconditional generation from white noise. This drawback disappears with explicit cycling through standard ℓ^2 loss (Belghazi et al., 2021). The tuning or arrangement of discriminators is a potential bottleneck in the process.

3.2 Possible Improvements

To circumvent the issues underlined in Section 3.1, we present some possible ideas for improvement:

(i) **Fuse the training step**
A solution to the redundancy of the strategy is to fuse the steps to perform the training of Broadband and PBS simultaneously and force the Generators to perform the hybrid signal-to-signal translation at the same time. As a matter of fact, we have opted for a unique encoder and two decoders (see Section 4).

(ii) **Reduce the discriminators' redundancy**
Recorded time histories and PBS share common features, which suggests that making the discriminator share weights could improve their discriminative performance and force encoders and decoders to better mimic the data distribution.

(iii) **Data Projection into Latent Space**
A linear bottleneck, similar to a linear classifier, is added to the top of the final CNN layer to project data into a latent space. This latent space serves as a compact, flattened representation of the input data (Aggarwal, 2018). Donahue and Simonyan (2019) addressed the challenge of aligning the data with a latent distribution $p(\mathbf{z})$ by reusing a classifier as an encoder. This encoder approximates the latent distribution as a Gaussian $p(\mathbf{z}) \approx \mathcal{N}(\mu, \sigma)$, where the mean $\mu = 0$ and the standard deviation σ is the identity matrix. Unlike Variational Autoencoders which explicit employ the Kullback-Leibler (KL) divergence to regularize the posterior, the ALICE framework achieves marginal Gaussianity through **Adversarial Joint Distribution Matching**.

Mathematically, the discriminator is trained to distinguish between the joint distributions $q(\mathbf{x}) = p(\mathbf{x}|\mathbf{z})p_{\text{prior}}(\mathbf{z})$. According to the principles of adversarial inference (Dumoulin et al., 2017; Donahue et al., 2017), when $q(\mathbf{x}, \mathbf{z}) = p(\mathbf{x}, \mathbf{z})$, the marginal distributions must necessarily match the following the prior, namely:

$$\begin{aligned} \int q(\mathbf{x}, \mathbf{z}) d\mathbf{x} &= \int p(\mathbf{x}, \mathbf{z}) d\mathbf{x} \\ &\implies q(\mathbf{z}) \\ &= p_{\text{prior}}(\mathbf{z}) \end{aligned} \tag{3}$$

By defining $p_{\text{prior}}(\mathbf{z}) = \mathcal{N}(0, \mathbf{I})$, the adversarial process forces the encoder $q(\mathbf{z}|\mathbf{x})$ to project input seismic data into a manifold that is statistically consistent with the Gaussian prior. This ensure that the latent space is not degenerate and allows for the stochastic *one-to-many* mapping required for hybrid synthesis. This approach avoids the *over-smoothing* artifacts, typically observed in VAEs, preserving the sharp transients of waveforms, essential to ensure high-frequency seismic realism.

Such improvements will help disentangle the manifold and have an independent Gaussian representation of each of the different 3 channel signals. This improvement will serve to split latent and share latent space in the next section.

4 Methodology

One primary objective of this research is to solve the problem $\mathcal{P}_y$ by reconstructing the broadband signals (0-30 Hz). The signal displays long-range dependencies, rendering it a complex challenge to analyze, focusing on capturing the distinct P- and S-wave patterns and their long-range dependencies. This distinct seismic signal pattern consistently manifests within our dataset (STEAD, see more explanation in Section 4.4). Therefore, the foremost step in our research involves identifying the most suitable architecture to handle this task effectively. To initiate our approach, we conduct a meticulous investigation to determine the optimal encoder and decoder, ensuring the highest precision in data reconstruction. In the next step, we introduce an adversarial loss to improve the reconstruction quality using a discriminator. Once this step is validated, the $\mathcal{P}_x$ problem is solved as well, and has a less complex dataset (0-5 Hz).

4.1 Strategy to tackle hybrid prediction

The motivation behind our approach is to enhance the versatility of our latent space representation. Therefore, we have introduced this factorization of the latent space, allowing for a more nuanced and expanded one-to-many mapping. In Equations 5 and 4, we expose how we intend to achieve our goal. We describe in Figure 3 how we proceeded to reconstruct broadband recordings and physics-based simulations, while considering their latent representations.

$$\mathbf{x} \xrightarrow[F_{xy}]{} \mathbf{z}_x = (F_{xy}(\mathbf{x})^{xy}, F_{xy}(\mathbf{x})^{xx}) \xrightarrow[\text{trunc}]{} \hat{\mathbf{z}}_{xy} \xrightarrow[G_x]{} \tilde{\mathbf{x}} \tag{4}$$

$$\mathbf{y} \xrightarrow[F_{xy}]{} \mathbf{z}_y = (F_{xy}(\mathbf{y})^{xy}, F_{xy}(\mathbf{y})^{yy}) \xrightarrow[\text{cat}]{} \hat{\mathbf{z}}_y \xrightarrow[G_y]{} \tilde{\mathbf{y}} \tag{5}$$

With the definition of F_{xy}, F_{xx} and F_{yy}:

$$\begin{aligned} F_{xy}(\bullet)^{xy} &: \mathbb{R}^d \to \mathbb{R}^{d_{xy}}, \quad \mathbf{x} \to \mathbf{z}_{xy} \\ F_{yy}(\bullet)^{yy} &: \mathbb{R}^d \to \mathbb{R}^{d_{yy}}, \quad \mathbf{y} \to \mathbf{z}_{yy} \\ F_{xx}(\bullet)^{xx} &: \mathbb{R}^d \to \mathbb{R}^{d_{xx}}, \quad \mathbf{x} \to \mathbf{z}_{xx} \end{aligned} \tag{6}$$

In the previous expressions, the symbol *cat* is for the concatenation of the two components of the latent space, while the *trunc* refers to the extraction of the standard sub-vectors within the latent space. At the end of the training, the generators should consider the value of the latent space from $\mathbf{x}$ and the Gaussian noise and then transform it into a signal from the same database. When only the common part is preserved and the specific part is replaced with Gaussian noise, the decoder interprets the signal as having a common part with a frequency between [0–1 Hz], and the specific part should provide complementary information, such as a signal for which the frequency field is between 1 and 30 Hz.

$$\begin{aligned} (\mathbf{z}_{xy}, \mathbf{z}_{xx}) &= F_{xy}(\mathbf{x}) \\ (\mathbf{z}_{yx}, \mathbf{z}_{yy}) &= F_{xy}(\mathbf{y}) \end{aligned} \tag{7}$$

As shown in Equation 4, we have to pay attention to $\mathbf{z}_{xy}$ to train the adversarial discriminator for physics-based simulations. The following equation is the expression of the loss:

$$\begin{aligned}\mathcal{L}^{\mathrm{x}}_{\mathrm{ALI}} &= \mathbb{E}[\log D_{xz}(\mathbf{x}, F_{xy}(\mathbf{x})^{xy})] \\ &+ \mathbb{E}[\log(1 - D_{xz}(G_x(\mathbf{z}_x), \mathbf{z}_x))]\end{aligned} \tag{8}$$

For the broadband signal, we consider the complete latent space with the following loss:

$$\begin{aligned}\mathcal{L}^{\mathrm{y}}_{\mathrm{ALI}} &= \mathbb{E}[\log D_{yz}(\mathbf{y}, \mathrm{cat}(F_{xy}(\mathbf{x})^{xx}, F_{xy}(\mathbf{y})^{yy}))] \\ &+ \mathbb{E}[\log(1 - D_{xz}(G_{\mathbf{y}}(\mathbf{z}_d), \mathbf{z}_d))]\end{aligned} \tag{9}$$

For the hybrid discrimination, we choose to use a loss that observes the pair of $(\mathbf{x}, \hat{\mathbf{y}})$ and the pair $(\hat{\mathbf{x}}, \mathbf{y})$ as we have done in the previous strategy. The expression of $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ is expressed as below:

$$\mathbf{x} \xrightarrow[F_{xy}]{} \mathrm{cat}(\mathbf{z}_{xx}, \mathcal{N}(0, I)) \xrightarrow[G_y]{} \hat{\mathbf{y}} \tag{10}$$

$$\mathbf{y} \xrightarrow[F_{xy}]{} \mathbf{z}_d \xrightarrow[\text{troncat}]{} \mathbf{z}_{xx} \xrightarrow[G_x]{} \hat{\mathbf{x}} \tag{11}$$

Then, the hybrid adversarial loss is presented before:

$$\mathcal{L}^{\mathrm{xy}}_{\mathrm{Hybrid}} = \mathbb{E}[\log D_{xy}(\boldsymbol{x}, \hat{\boldsymbol{y}})] + \mathbb{E}[\log(1 - D_{xy}(\hat{\boldsymbol{x}}, \boldsymbol{y}))] \tag{12}$$

4.2 Unique encoder and new joint discriminators design

To improve super-resolution mapping, encoders F_y and F_x were merged into dual-branch encoder F_{xy}. However, the decoders G_y and G_x remained separate. A new discriminator D_{xy} was introduced. The particularity of D_{xy} is now the fact that it shares weights with D_{yz} and D_{xz} discriminators, as illustrated in Figure 4b. Reusing the same part to form D_{xy} stabilizes the training of the generators and reduces the memory footprint placed in the GPU, allowing faster training. The equation modeling is presented below:

$$D_{xy}(\mathbf{x}, \mathbf{y}) = D_{xy}(D_{s_x}(\mathbf{x}), D_{s_y}(\mathbf{y})) \tag{13}$$

To focus on discriminating hybrid mappings in the unified domain, we use Equation 12. Since we manipulate time series data, we are supposed to observe the quality of the reconstruction both in the time domain and the frequency domain. To force the quality reconstruction, we adapted focal frequency loss (FFL) (view Equation 14) and the Hyper Spherical Loss (HSL) for this purpose (view Equation 15). The equivalent of FFL for a 1D signal would be:

$$\mathrm{FFL} = \frac{1}{N}\sum_{u=0}^{N} w(u)\,|F_r(u) - F_f(u)| \tag{14}$$

in which t is the instant in the time signal, u is the value from the spectrum, $x(t)$ is the value of the signal at time t, its Fourier amplitudes have the value $F(u) = \sum_{i=1}^{N-1} x(t)e^{i2\pi\left(\frac{u\cdot t}{N}\right)}$, the matrix weight element $w(u) = |F_r(u) - F_f(u)|^{\alpha}$, the factor α is a scaling factor for flexibility (trade off $\alpha = 1$), $F_r(u) = a_r + b_r\mathrm{i}$, and $F_f(u) = a_f + b_f\mathrm{i}$. The equation for the HSL rescales to prevent vanishing gradient, addressing the significant difference in amplitude between physics-based simulations and broadband signals:

$$\begin{aligned}\mathrm{HSL}(\mathbf{y}, \hat{\mathbf{y}}) &= \sum_{i=1}^{N}(1 - \cos(\theta_{\mathbf{y}_i, \hat{\mathbf{y}}_i}))^2 \\ &= \sum_{i=1}^{N}\left(1 - \frac{\|\mathbf{y}_i \cdot \hat{\mathbf{y}}_i\|}{\|\mathbf{y}_i\|\cdot\|\hat{\mathbf{y}}_i\|}\right)^2\end{aligned} \tag{15}$$

4.3 Phase Alignment

Potential misalignment artifacts encountered during the initial optimization process are detailed in Appendix A.7. This unexpected behaviour occurs when standard minimization techniques lack sufficient temporal constraints to preserve phase arrival consistency. Such alteration of the natural alignment of the S-wave and P-wave is observed in the hybrid data. To mitigate this, we incorporate a pre-trained neural network designed for phase picking as an additional constraint in the generative process. Specifically, we adopted the EQ-Transformer architecture proposed by Mousavi et al. (2020) as feature extractor. This architecture is designed for signals of 6000 time steps; since our dataset features signals of 4096 time steps, we zero-padded them accordingly. The resulting penalty loss, $\mathcal{L}_{EQT}$, is defined as the sum of discrepancies across the three outputs probability traces:

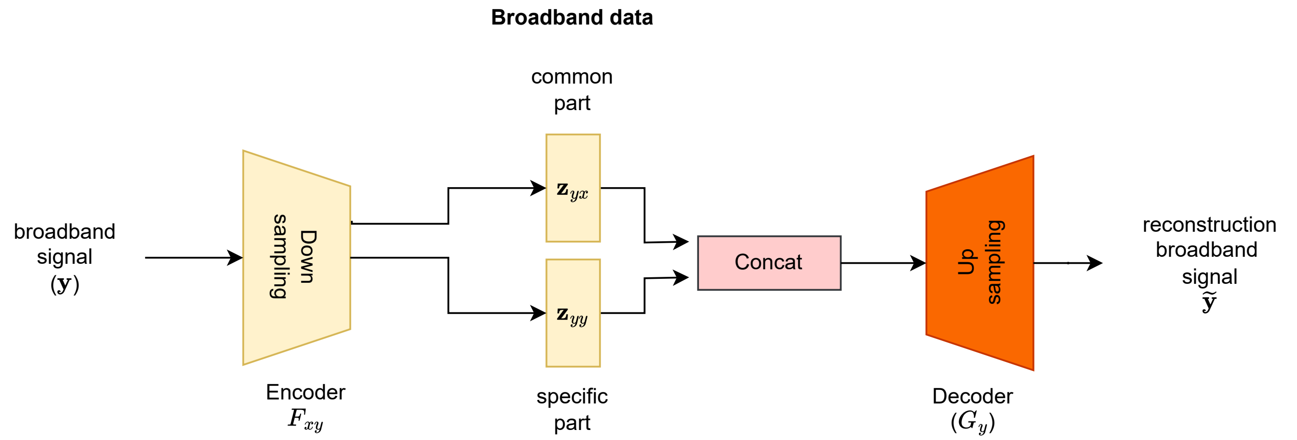
$$\begin{aligned}&\mathcal{L}_{EQT} \\ &= \|EQ_P(\mathbf{y}) - EQ_P(\hat{\mathbf{y}})\|_2 \\ &\quad + \|EQ_S(\mathbf{y}) - EQ_S(\hat{\mathbf{y}})\|_2 + \|EQ_D(\mathbf{y} - EQ_D(\hat{\mathbf{y}})\|_2\end{aligned} \tag{16}$$

where $EQ_P(\mathbf{y}) \in [0,1]^T$, $EQ_S(\mathbf{y}) \in [0,1]^T$ and $EQ_D(\mathbf{y}) \in [0,1]^T$ are the probability distributions of P- and S-wave arrival times and temporal probability traces for global phase detection of the real signal, $\mathbf{y}$, whereas $EQ_P(\hat{\mathbf{y}})$, $EQ_S(\hat{\mathbf{y}})$, and $EQ_D(\hat{\mathbf{y}})$ are respectively the corresponding predictions obtained with EQ-Transformer. T is the number of the time steps. Unlike traditional phase picking which requires a discrete threshold to identify a single arrival time, EQTransformer uses the full probability envelope. A penalty factor λ_{EQ} (trade-off coefficient) is adopted. Adding this penalty loss corrects the misalignment of P- and S-waves of hybrid generated data from physics-based inputs $\mathbf{x}$.

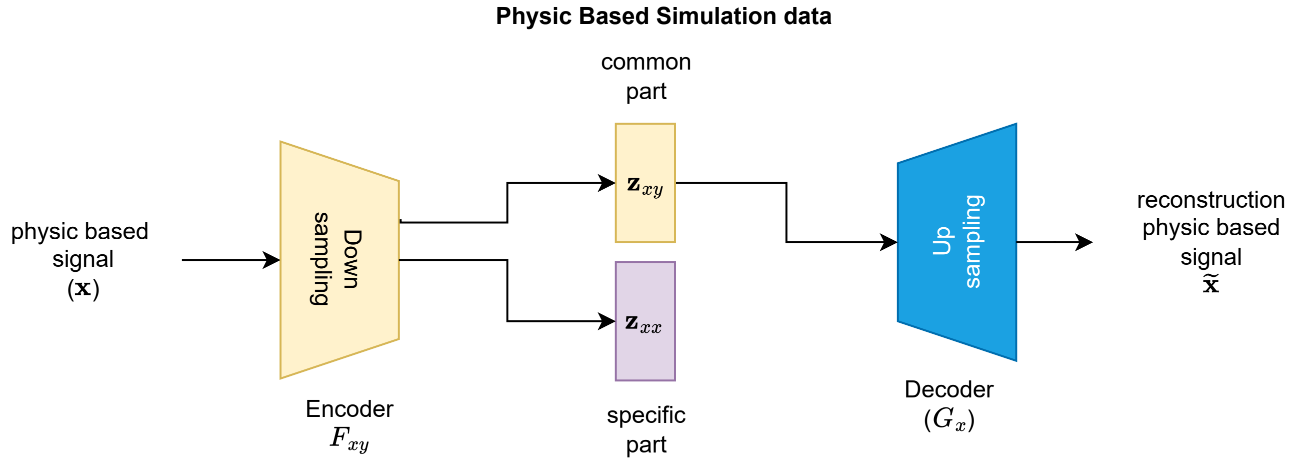
By penalizing the discrepancy across the entire probability envelope for phase detection of P-wave and S-wave onsets, we ensure that the generator reconstructs the transient characteristics that a state-of-the-art phase picker recognizes as arrival signals. This approach maintains full differentiability during the back propagation, avoiding the numerical instabilities associated with non-differentiable threshold or $\arg\max$ operations.

Therefore, we want to satisfy:

(I) Reconstruction of ground motion, $G_y(F^{yx}_{xy}(\mathbf{y}), F^{yy}_{xy}(\mathbf{y}))$;



(a) Inference of Broadband. The latent space is split into two parts: the $\mathbf{z}_{xy}$ and the $\mathbf{z}_{yy}$. $\mathbf{z}_{xy}$ represent the common part, which is the low-frequency signal. The specific part is the $\mathbf{z}_{yy}$, which corresponds to the high-frequency part.



(b) Inference of physics-based Simulation. We illustrated how the latent part is split into two parts. The $\mathbf{z}_{xx}$ is not used for the reconstruction of physics-based simulations

Figure 3 Workflow of encoding and decoding of the Physic Based and Raw seismic data

(II) Reconstruction of physic based simulation, $G_x(F_{xy}^{xy}(\mathbf{x}))$;

(III) Force the coherence of common part, $F_{xy}^{xy}(\mathbf{x}) = \mathbf{z}_{xy} \sim F_{xy}^{xy}(\mathbf{y}) = \mathbf{z}_{yx}$;

(IV) Force each part to follow Gaussian Distributions;

(V) Generation of $G_y(\mathbf{z}_y)$;

(VI) Generation of $G_x(\mathbf{z}_x)$;

(VII) Generation of many hybrid output $G_y(\mathbf{x}, \mathcal{N}(0, \mathbf{I}))$;

(VIII) Force the phase alignments for generation one to many.

4.4 Database Size

The complexity of the training requires, in addition, an appropriate architecture with sufficient data to train. Therefore, endeavours are made to extract the entire database of STEAD. Such training has necessitated significant improvement. We have extracted 51,200 three-component broadband signals and and 51,200 three-component filtered signals. This latter represents the amount of 128,000 of three-component signals. The filtered waveforms used in this research are the direct low-pass filtered versions of the corresponding broadband signals, ensuring a strictly paired dataset for the generative mapping. For each three-component seismic record, the preprocessing involves several steps: (i) linear detrending, (ii) the application of a 5% Tukey window to mitigate edge effects, and (iii) normalization by horizontal PGA. The filtering is performed using a **fourth-order Butterworth low-pass filter** implemented via Second Order Sections (SOS) to ensure numerical stability. The cut-off frequency is set to 1.0 Hz. Regarding the dataset selection criteria, we extracted "*local*" earthquake traces from the STEAD database (Mousavi et al., 2019) with a minimum source magnitude of $M_W \geq 3.5$. While no explicit site class or epicentral distance (R) filtering was applied during extraction, the random sampling strategy ensures a globally representative distribution of these parameters. We had yet to tackle the problem of data-model inflation, as expressed in the *Chinchilla* paper (Hoffmann et al., 2022), but a series of successive research and development have led us to this final approach. This has necessitated parallel extraction. Such an explanation is beyond the scope of this paper.

4.5 Total training loss

The holistic training process integrates three key components: the adversarial loss, the reconstruction loss, and the EQTransformer loss, serving as feature extractors. Each component contributes to the overall training loss; specific weighting parameters regulate their revealed importance. In our case we use as parameters $\lambda_{\mathrm{adv}} = 1$, $\lambda_{\mathrm{rec}} = 10$, and $\lambda_{\mathrm{EQ}} = 0.1$. These values are a trade-off that stabilizes the influence of the different parts of the whole training process. We comprehensively understand how each component contributes by visualizing the entire spectrum of losses and metrics. Because all the elements have been separately studied before being combined, we can assess that each could not be removed without endangering the stability of the training process.

$$\begin{aligned}\mathcal{L} &= \lambda_{\mathrm{adv}}(\mathcal{L}^{\mathrm{x}}_{\mathrm{ALI}} + \mathcal{L}^{\mathrm{y}}_{\mathrm{Hybrid}} + \mathcal{L}^{\mathrm{xy}}_{\mathrm{ALI}}) \\ &+ \lambda_{\mathrm{rec}}(\mathcal{L}^{\mathrm{x}}_{\mathrm{rec}} + \mathcal{L}^{\mathrm{y}}_{\mathrm{rec}} + \mathcal{L}^{\mathrm{x}}_{\mathrm{rechyb}}) \\ &+ \lambda_{\mathrm{EQ}}\mathcal{L}^{\mathrm{EQ}}\end{aligned} \tag{17}$$

with

$$\begin{aligned}\mathcal{L}^{\mathrm{x}}_{\mathrm{rec}} &= \mathrm{HSL}(\mathbf{x}, G_x(F^{xy}_{xy}(\mathbf{x})) + \mathrm{FFL}(\mathbf{x}, G_x(F^{xy}_{xy}(\mathbf{x})) \\ \mathcal{L}^{\mathrm{y}}_{\mathrm{rec}} &= \mathrm{HSL}(\mathbf{y}, G_y(F^{yx}_{xy}(\mathbf{y}), F^{yy}_{xy}(\mathbf{y}))) + \\ &\quad \mathrm{FFL}(\mathbf{y}, G_y(F^{yx}_{xy}(\mathbf{y}), F^{yy}_{xy}(\mathbf{y}))) \\ \mathcal{L}^{\mathrm{x}}_{\mathrm{rechyb}} &= \mathrm{HSL}(\mathbf{x}, G_x(F^{xy}_{xy}(\mathbf{y})) + \mathrm{FFL}(\mathbf{x}, G_x(F^{xy}_{xy}(\mathbf{y}))\end{aligned} \tag{18}$$

5 Improving SeismoALICE

5.1 Generators

(i) **Auto-Encoder structure (AE)**
The original SeismoALICE adopted encoders and decoders, conceived as stacks of 1D convolutional and transposed convolutional layers. In order to preserve long-range correlation, convolutional residual blocks (ConvResBlock and ConvBlockTranspose, respectively), with skip-connection, were adopted (see the Figure 4a).

The ConvResBlock comprises two branches: one for normal feature extraction and the other for skip-connection. This architecture is inspired by SAGAN (Self-Attention Generative Adversarial Networks, Zhang et al., 2019). Moreover, we employ an attention mechanism inspired by the Conformer architecture (Gulati et al., 2020), which we use in both encoder and decoder spectral normalization (Miyato et al., 2018). This was adopted as a regularization technique at all levels of the architecture, to improve Generator stability and reinforcing the Discriminator Lipschitz continuity (Brock et al., 2019; Zhang et al., 2019). Special Normalization normalizes the weight with the respect of their norm:

$$\begin{aligned}W_{\mathrm{SN}}(\mathbf{W}) &:= \frac{\mathbf{W}}{\sigma(\mathbf{W})} \\ \sigma(\mathbf{W}) &:= \max_{\mathbf{h}:\mathbf{h}\neq\mathbf{0}} \frac{\|\mathbf{W}\mathbf{h}\|_2}{\|\mathbf{h}\|_2}\end{aligned} \tag{19}$$

In addition, to normalize the layers, we perform the normalization of the output, which is replaced by the Instance Norm.

(ii) **Choice of bottleneck layer**
For Gaussian mapping, dense layers are used at the end of the encoder and the beginning of the decoder. The same logic is employed in the discriminators.

(iii) **Architecture variant proposed**
Following an extensive ablation study, we determined that the Conformer architecture — which integrates self-attention with convolutional modules — is the most effective for capturing multi-scale temporal dependencies and generating realistic seismic waveforms from a Gaussian latent manifold. We benchmarked the reconstruction performance of several hybrid configurations, as illustrated by the Goodness-of-Fit (G.O.F.) distributions in Figure A.5. While the hybrid Conformer + UNet architecture achieves high-fidelity reconstruction, it does not yield a statistically significant improvement over the standalone Conformer generator. Furthermore, comparative analyses (detailed in Appendix Figure A.6) reveal that the inclusion of UNet skip-connections or Adaptive Instance Normalization (AdaIN) tends to over-smooth the output, effectively suppressing the high-frequency ambient noise components that are essential for realistic seismic textures. Given that the standalone Conformer already achieves a *Good* to *Excellent* rating according to the quantitative criteria in Table A.1, we opted for this more parsimonious architecture. This choice ensures computational efficiency and mitigates the risk of over-fitting associated with redundant skip-connections or style-transfer layers, while better preserving the stochastic characteristics of the ground motion.

(iv) **Post-processing**
To avoid overestimation of the spectra we have introduced a Global Spectra Calibration. We pre-compute a frequency-dependent transfer function, $H_{calib}(f)$, defined as the ratio between the mean magnitude spectra of the target records and the mean magnitude spectra of the generated ensemble over the calibration set. This normalization ensures that the generated signals respect the empirical energy distribution across the entire 0.1–50 Hz bandwidth, reducing the spectral bias to within ± 0.2 log-units as required for engineering-grade applications.

(v) **Comparison with previous architecture**
Compared to Gatti and Clouteau (2020), our architecture reduces the number of hidden variables

to translate from broadband to physics-based and vice versa by applying regularization techniques (refer to Section A.1 for better understanding). Then, a better understanding of the interpretability of the latent values is provided. Our architecture can perform generation through the adequate choice of projection technique to latent space.

5.2 Discriminators

The formulation of Discriminator posed another significant challenge, particularly due to its pivotal role in the comprehensive training process. In the complete training, the generators (encoder and decoder) undergo updates based on the evaluation provided by the discriminator. So, the design of the discriminator should be coherent and efficient, capturing and not destroying the information, even though the main task in the complete training is to fool the discriminator (Goodfellow et al., 2014; Gulrajani et al., 2017; Kumar et al., 2019). To this end, the original SeismoALICE discriminators were condensed into a unified architecture with three branches. This occurred for both joint discriminators, for $(\mathbf{x}, \mathbf{z})$ and $(\mathbf{y}, \mathbf{z}')$. For the sake of clarity, we described the three branches only for the pair $(\mathbf{y}, \mathbf{z}')$, but the same architecture was adopted for $(\mathbf{x}, \mathbf{z})$. The first one, named D_{sy}, has the task to extract the relevant signature of the signal $\mathbf{y}$ or $G_y(\mathbf{z}_y)$. We adopted the same architecture of the encoder to design D_{sy}. This proposed design was advised by Brock et al. (2019). The second branch, D_{szb}, has the task to discriminate between $\mathbf{z}_y$ and $F_y(\mathbf{z}_y)$. D_{sy} and D_{szb} can be considered as logits values or intermediate feature maps that are concatenated and fed into the discriminating branch $D_{yz}(D_{sy}(\mathbf{y}), D_{szb}(\mathbf{z}_y))$. D_{yz} is naturally featured by a final sigmoid activation function, $\sigma(\bullet)$, that should output a value between 0 and 1 (a probability of the sample to be true or false). By experience, the BCEWithlogit[3](Binary Cross Entropy With Logits) is numerically more stable than the classic BCE[4] (Binary Cross Entropy). This same output is suitable for Hinge Loss, proposed in GAN by (Brock et al., 2019), which is treated in Section 6, and this is our choice. Figure 4b depicts the scheme of the unified discriminator architecture. Therefore, the retained ALI loss has the following expression for the broadband:

$$\begin{aligned}\mathcal{L}_{\text{ALI}} = &\mathbb{E}_{\mathbf{y}\sim p(\mathbf{y})}\left[\log(\sigma(D_{yz}(\mathbf{y}, F_y(\mathbf{y}))))\right] \\ &+ \mathbb{E}_{\mathbf{z}'\sim p(\mathbf{z}')}\left[\log(1-\sigma(D_{yz}(G_y(\mathbf{z}'), \mathbf{z}')))\right]\end{aligned} \tag{20}$$

Additional information on the architecture can be found in Tables A.4 A.5. The architecture of D_{s_y} is the same as Encoder F_y (explained above).

[3]The mathematical expression of cross entropy between two variable x and y is defined by: $\ell(\mathbf{x}, \mathbf{y}) = L = \{l_1, \dots, l_N\}^\top$, $l_n = -w_n\left[y_n \cdot \log\sigma(x_n) + (1-y_n)\cdot\log(1-\sigma(x_n))\right]$. In this formulation we have $\mathbf{x} = x_1, x_2, \cdots, x_n$ is the set and $\mathbf{y} = y_1, y_2, \cdots, y_n$ are the labels, usually $\mathbf{y} = \mathbf{1}$ or $\mathbf{0}$. σ correspond to the sigmoid. In the case of machine learning, this corresponds to the fact that this activation is not present in the last discriminator layer. The numerical stability of this expression has been advocated. The w_n is the weight of importance of the loss

[4]The expression of BCE is defined by: $\ell(x, y) = L = \{l_1, \dots, l_N\}^\top$, $l_n = -w_n\left[y_n \cdot \log x_n + (1-y_n)\cdot\log(1-x_n)\right]$

In addition, to fool the discriminator easily and to foster generator training, we carefully calibrated the noise level (Aggarwal, 2018) by clamping a white noise $\mathbf{n}$ to $\mathbf{y}$, *i.e.*, $\hat{\mathbf{y}} = \mathbf{y} + \mathbf{n}$, with values in the range -0.5 and 0.5. It is noteworthy, that $\mathbf{y}$ is normalized by default within -1 and +1. The choice of the value of the clamp is a trade-off, and it might change depending on the dataset. The clamp reduces the amplitude of the white noise, polluting the original data $\mathbf{y}$, particularly for the high frequencies. A quick overview of the frequency domain of synthetic data will showcase that high-frequency data is different from broad-band data. This pathology was already present in Gatti and Clouteau (2020). Therefore, we must make it difficult for the discriminator to penalize high frequency regarding numerical limitations, and the introduction of that noise stabilizes the training.

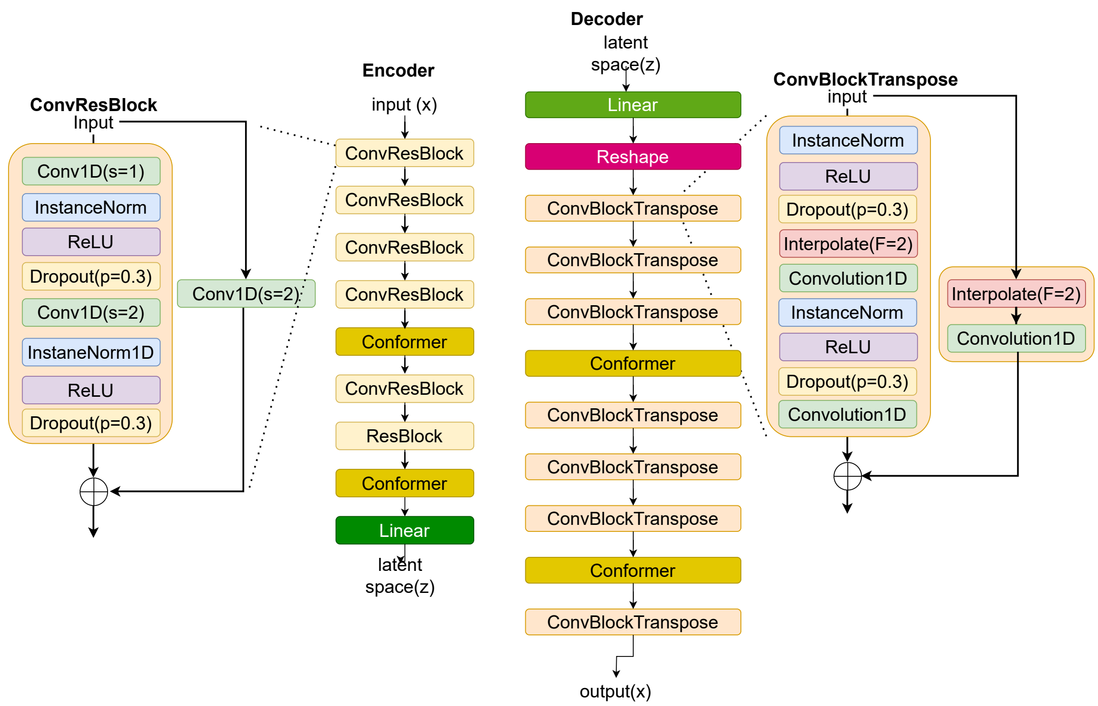
Furthermore, our investigation into discriminator stability revealed that the inclusion of residual blocks and Spectral Normalization (SN) is essential for robust training (see Figure 4b). Stable training—defined as the avoidance of common GAN pathologies such as mode collapse or gradient oscillations—is characterized by a consistent convergence toward a Nash equilibrium without significant deviation from the optimization objective (Becker et al., 2022). We apply SN to both convolutional and fully-connected layers to enforce Lipschitz continuity by bounding the spectral norm of the weight matrices. The final layer of the discriminator maps the extracted features to a scalar score rather than a probability distribution. Consequently, we utilize the `BCEWithLogitsLoss` function; by combining a sigmoid layer with the binary cross-entropy in a single class, this approach offers superior numerical stability over the standard `BCE` loss, particularly when dealing with the high dynamic range of seismic amplitudes.

6 Results

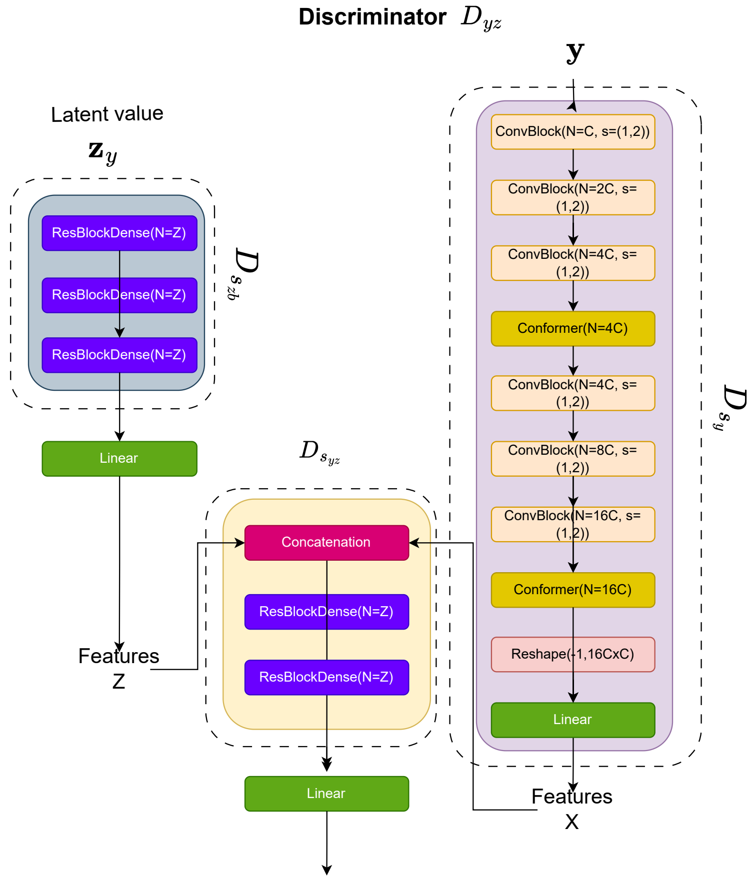
6.1 Comparison with Previous method

The novel architecture surpasses the limitation of the previous strategy constraint in the one-to-one mapping. Such architecture could map one-to-many, as illustrated in Figure 5. For the result we use an alternative architecture more suitable to handle the factorized latent space which is a conformer supporting adaptive instance norm (AdaIN).

The architecture validates our objectives' specific criteria. Objectives I, II, IV, V, VI, and VII are satisfied. Please refer to these aforementioned objectives in Section 4.3. Figure A.5 provides a detailed overview of the model's performance. In the figure, we have made a histogram to view which architecture is more adapted for hybrid mapping. The comparison is between $G_y(F_{xy}(\mathbf{x}), F_{xy}(\mathbf{y}))$ and the targeted broadband $\mathbf{y}$. Among many architectures the included Conformer is more adapted for the hybrid reconstruction task (see Figure A.6). This breakthrough in mapping flexibility and validation of crucial criteria signifies an improvement of our architecture, showcasing its effectiveness in handling complex data representations and generations.



(a) Generators with conformer architectures. The ConvResblock is composed of a residual Block. The ConvTransposeBlock does not use ConvTranspose1d, but the upsampling is done using interpolate with factor 2. A description of ConvResBlock and ConvTransposeBlock architecture can be found in section §5.



(b) Architecture of the joint discriminator of D_{yz}. The Design includes a Residual Block. $C = 64$, N represents the number of Channels, s represents the strides. The number of parameters of the discriminator is 10 million parameters. This architecture comprises two branches, D_{s_x} and $D_{s_{z_y}}$.

Figure 4 Definition of the Generators and Discriminator including the Conformers adapted for the training of the time series data.

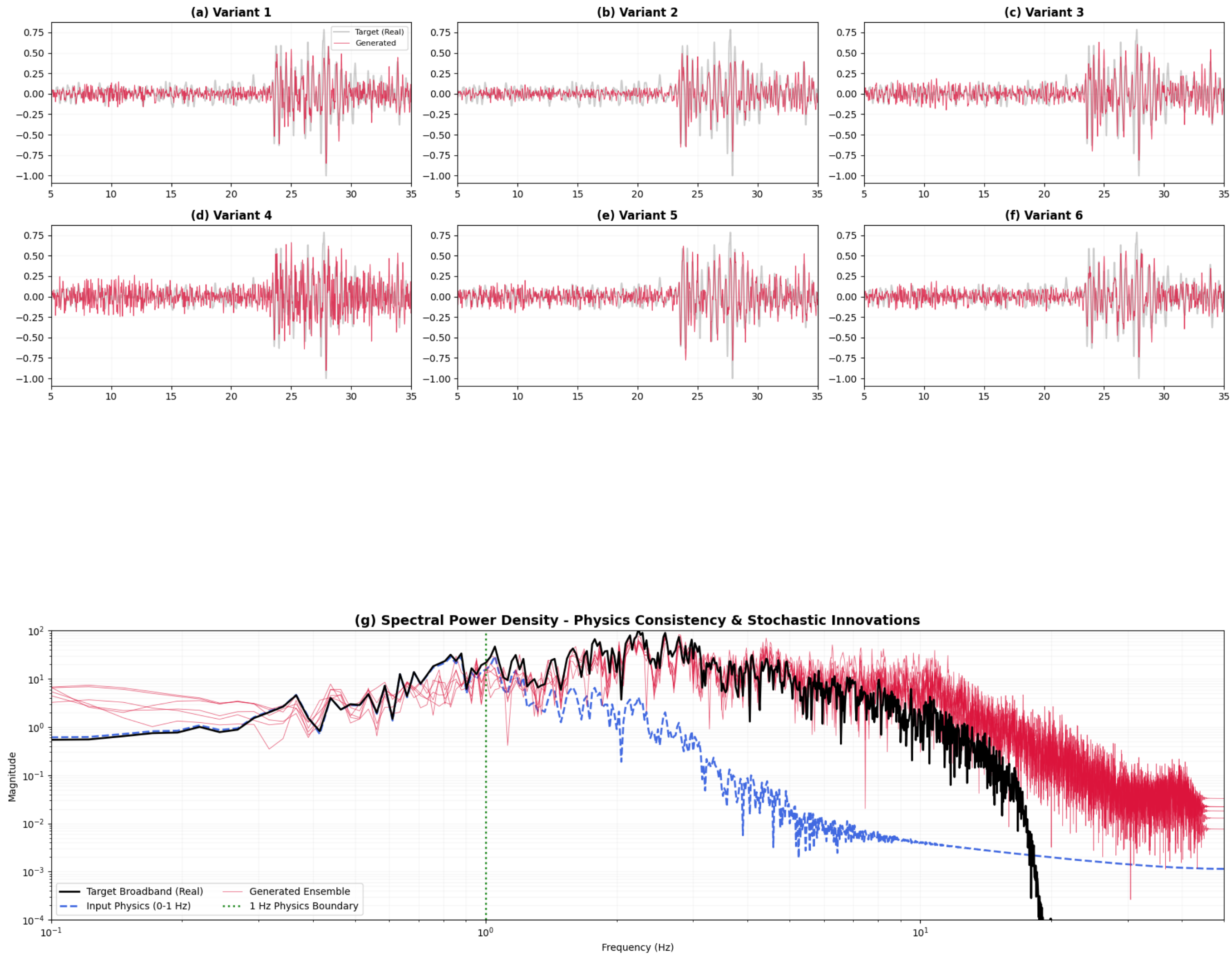


Figure 5 Here is an illustration of the multi-modal generation for the seismic data. The input $\mathbf{x}$, the high-frequency encoding aspect is replaced by a Gaussian distribution, $\mathcal{N}(0, I)$. The generated output should produce an infinity value likelihood to the targeted broadband data. $G_y(\mathrm{cat}(F_{xy}^{yx}(\mathbf{x}), \mathcal{N}(0, I)) \sim \{\mathbf{y}_0, \mathbf{y}_1, ..., \mathbf{y}_\infty\}$.

6.2 Model performance

The quantitative performance of the different architectures is summarized in Table A.6. We evaluate the models using the Envelope (EG) and Phase (PG) Goodness-of-fit, which provide a rigorous measure of misfit in the time-frequency domain. As shown, the standalone Conformer architecture has a PG score of (8.32 ± 0.46), indicating high temporal coherence in the phase reconstruction. While the more complex Conformer + AdaIN + UNet variant shows a marginal increase in Broadband EG (6.62 ± 0.80), the overlapping confidence intervals suggest that this improvement is not statistically significant. Furthermore, when considering the Hybrid EG, the stand-alone Conformer (5.80 ± 1.07) remains competitive with more parametrized hybrid variants and realistic results. This quantitative evidence supports our selection of parsimonious Conformer architecture. By maintaining a high PG score while utilizing fewer parameters (219.44 M vs 231.00 M), the model satisfies the demand for computational efficiency without compromising the phase accuracy essential for seismic engineering. These metrics confirm that the generator effectively maps the low-frequency physical constraints into a broadband stochastic manifold that preserves the non-stationary envelope characteristics of the original records.

Beyond average error metrics, the model's ability to capture the statistical distribution of seismic intensity measures was verified through a one-to-many coverage analysis. As detailed in Appendix A.10 and Figure A.8, the generator reproduces the empirical distribution of Arias Intensity and significant duration with high fidelity ($W_1 = 0.185$). Although the PGA distribution exhibits a high Kolmogorov-Smirnov distance due to the unit-normal of the STEAD dataset, the Wasserstein distance confirms that the stochastic variability remains within a physically consistent envelope. Thus, the one-to-many mapping addresses the mode-dropping issues often encountered in adversarial training.

Focusing on the quality of the reconstruction of the stand-alone conformer architecture, we verify the statistical robustness of our approach. We conduct a global performance assessment on a test subset comprising 10% of the STEAD dataset ($N = 12,800$ three-component signals, per type of accelerograms). Figure 6 illustrates the variate distribution of PG and EG for the different types of accelerograms (broadband, filtered, and hybrid). The scatter plot demonstrates the high concentration of points in the upper-right quadrant, especially within the [6–10] numerical range. According to the verbal classification in Table A.1, this corresponds to a good to excellent reconstruction quality across the vast majority of the test population. Crucially, the marginal histograms reveal a minimal performance gap between the direct Broadband reconstruction (red) and the generated hybrid signal (green).

This finding is significant as it demonstrates that the generator is capable of synthetizing high-frequency details from low-frequency physics-based inputs without a substantial loss in statistical coherence compared to a standard auto-encoding task. This aforementioned result for the different types of accelerograms provides empirical evidence that the model has successfully generalized the mapping between latent manifolds, mitigating the risk of overfitting that might otherwise be expected in a high-parameter architecture.

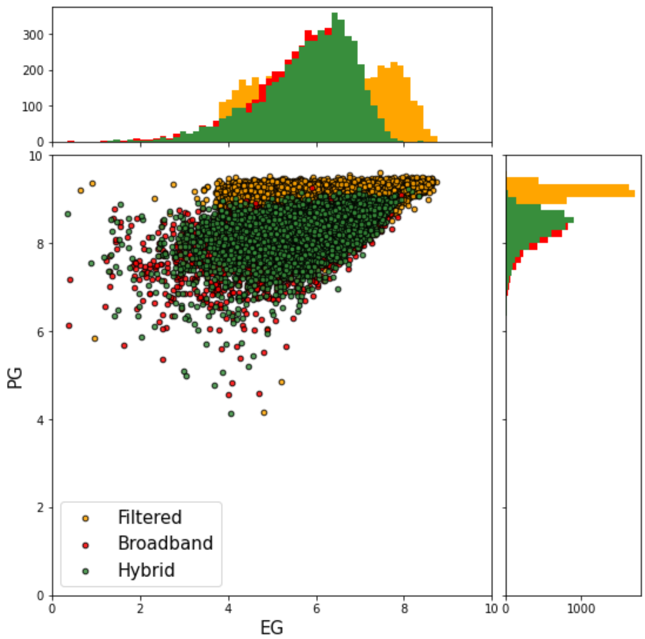


Figure 6 Plot of GOF of the architecture for the unique training. Representation of the quality of the reconstruction for Broadband, Filtered, and Hybrid. Tests have been made on 10% of the STEAD database. The dataset is 128,000 three-component signals for Broadband and Filtered data.

Finally, the temporal accuracy of the generated outputs was evaluated using the Median Absolute Error (MedAE) between the real P and S arrival times and the generated arrival times. The SeimoALICE-v2 architecture (with mask and $\mathcal{L}_{EQT}$) achieves MedAE of 0.19 s for P-wave arrivals (see Table A.7 in Appendix A.9)

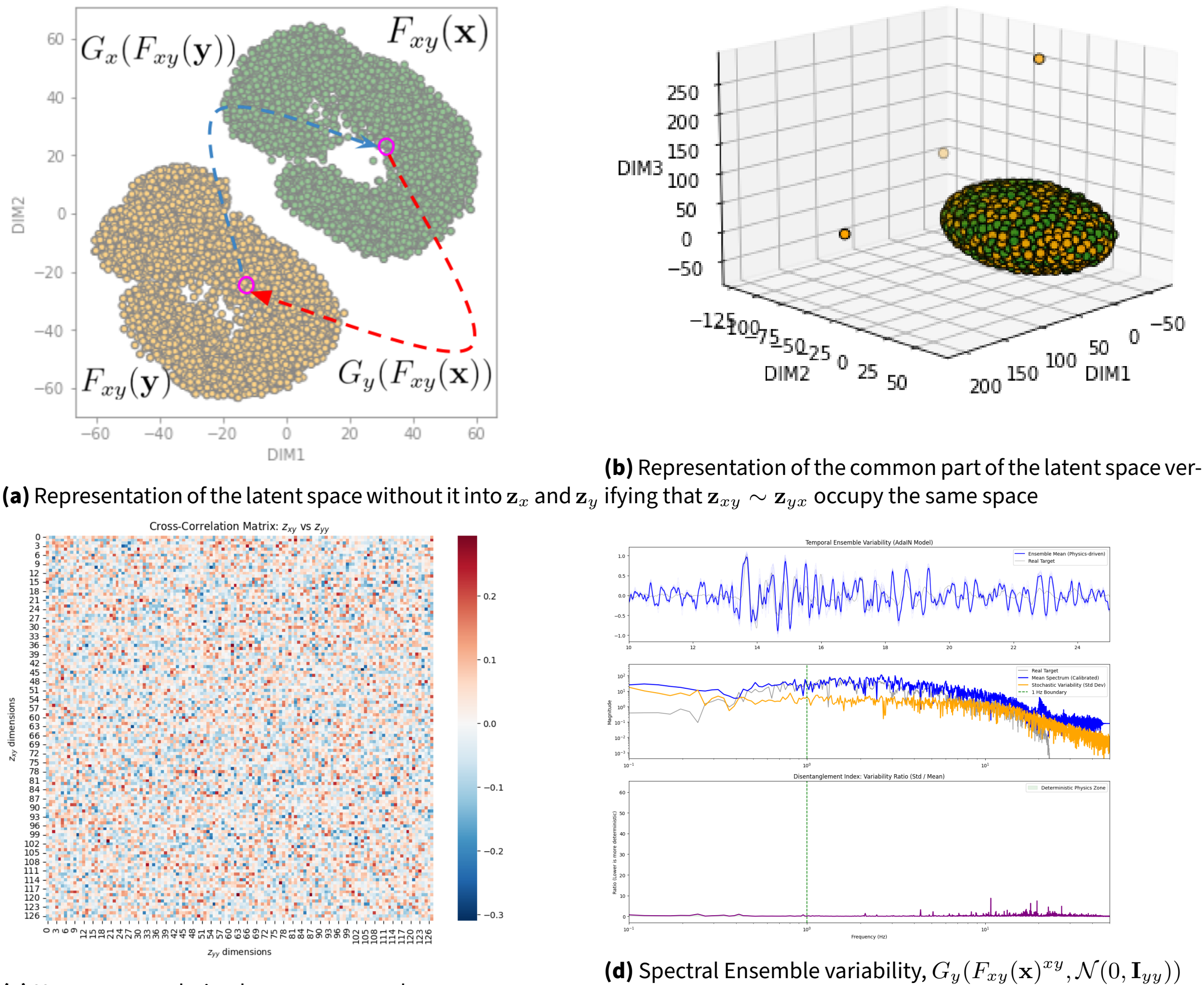
6.3 Interpretability of the latent space

The consistency of the low-frequency information in the common latent space, whether from physics-based simulation or ground motion, has been observed using a t-SNE (*t-distributed stochastic neighbor embedding*). Figure 7b presents the results of this analysis using dimensionality reduction methods. While t-SNE visualizations offer qualitative insights, we formally verify the disentanglement between the common subspace $\mathbf{z}_{xy}$ and the broadband-specific subspace $\mathbf{z}_{yy}$ through a dual structural and functional analysis. To provide a quantitative bound on the information leakage between the subspaces, we investigate the use of Mutual Information Neural Estimation (MINE). However, estimating MI in high-dimensional spaces (128D) presents a significant challenges regarding sample complexity and gradient variance. In our case, the estimator failed to reach a stable numerical convergence despite the dataset size. Consequently, we computed the cross-correlation matrix between these 128-dimensional subspaces over the test set ($N = 5,120$ 3D signals). The resulting matrix exhibits no systematic patterns, with a mean absolute correlation of **0.0659** (see Figure 7c). This near-zero correlation indicates a lack of linear coupling between the two subspaces, suggesting that the latent factorization effectively separates common low-frequency features from stochastic high-frequency innovations. Secondly, a functional ablation study was conducted by fixing $\mathbf{z}_{xy}$ and sampling multiple realizations of $\mathbf{z}_{yy}$. As shown in Figure 7d, the disentanglement index (variability ratio) remains negligible (< 0.1) within the deterministic physics bandwidth (0.1–1 Hz). Consequently, the stochastic output variability is strictly confined to frequencies above 1 Hz, where the standard deviation (orange) is at least an order of magnitude lower than the ensemble mean (blue) in the low-frequency band. This confirms that $\mathbf{z}_{yy}$ has only a marginal influence on the 0–1 Hz range, effectively ruling out consequential leakage.

Additionally, the visual representation in Figures 7a and 7b demonstrates how *a priori* separated latent representation $F_{xy}(\mathbf{x})$ and $F_{xy}(\mathbf{y})$ could be refined, in such a way that the common information becomes a single clustering structure. We observe that the projection of any point of two data types is nearby. The points are from $F_{xy}^{xy}(\mathbf{x})$ (physics-based simulations) and from ground motion data ($F_{xy}^{xy}(\mathbf{y})$). Consequently, our initial assumption is unequivocally satisfied: our architecture is adept at maximizing dependencies between $\mathbf{z}_{xy}$ and $\mathbf{z}_{yx}$.

To supplement the linear correlation analysis, a functional independence study was performed by fixing and sampling. The results (Figure 7d) show that variability is strictly confined to the >1 Hz regime, providing physical evidence that no high-frequency information leaks into the common subspace.

To evaluate the generative robustness of the Conformer-based architecture, we demonstrate its capability for unconditional synthesis by sampling a latent vector $\mathbf{z}$ from a standard multivariate Gaussian distribution $\mathcal{N}(0, \mathbf{I})$. We test whether the Decoder G_y has effectively mapped the high-dimensional seismic signal space into a continuous, low-dimensional manifold, for broadband and filtered signals (for filtered signals with fewer features this capacity is easy for our model to acheive). Figures A.4 and A.3 showcase several independent realizations generated from random latent seeds. The variety in the produced



(a) Representation of the latent space without it into $\mathbf{z}_x$ and $\mathbf{z}_y$

(b) Representation of the common part of the latent space verifying that $\mathbf{z}_{xy} \sim \mathbf{z}_{yx}$ occupy the same space

(c) Heatmap correlation between $\mathbf{z}_{xy}$ and $\mathbf{z}_{yy}$

(d) Spectral Ensemble variability, $G_y(F_{xy}(\mathbf{x})^{xy}, \mathcal{N}(0, \mathbf{I}_{yy}))$

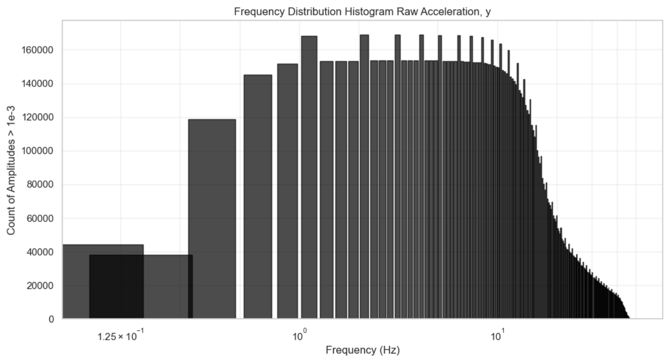
Figure 7 Comparison with two different representations of the latent space. This shows how one extracts the subspace representing the common information between broadband and filtered data.

waveforms—ranging from distinct impulse onsets to complex coda decay—demonstrates that the model has avoided “mode collapse”, a common GAN pathology where the generator produces a limited set of repetitive outputs. Statistically, this capability confirms that the generators have learned the implicit temporal patterns of an earthquake (the sequential transition from P-wave to S-wave and subsequent scattering). From a machine learning perspective, the stability of these unconditional samples is attributed to the spectral normalization applied across the Conformer blocks and the residual network supporting self-attention mechanisms. Such characteristics enforce Lipschitz continuity and prevent gradients from exploding during the mapping from the latent space to the physical domain. Consequently, the model does not merely act as a deterministic filter but as a true stochastic synthesizer capable of exploring the infinite variety of seismic manifolds while maintaining strict physical consistency.

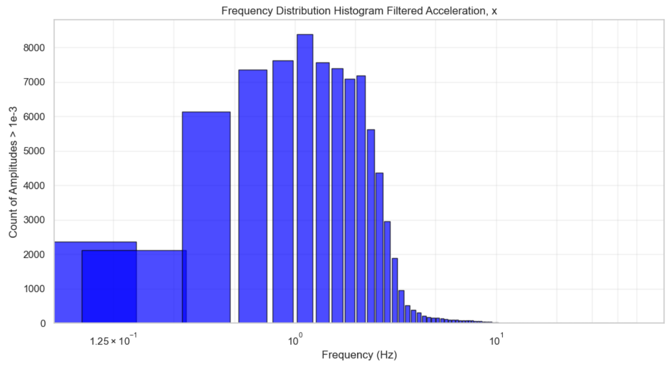
6.4 Discrepancy

A persistent challenge is the alignment of the generated signal frequency with the very low-frequency range of the PBS. This difficulty arises from the inherent characteristics of the signal distribution. Examination of the distributions of both the raw signals ($\mathbf{y}$) and the filtered signals ($\mathbf{y}$), which typically exceed 0.5 Hz, highlights this issue. Figure 8 presents a spectral transformation of the 3D signals using a discrete Fourier transform. Applying a simple detection threshold may help clarify why the discrepancy remains unresolved.

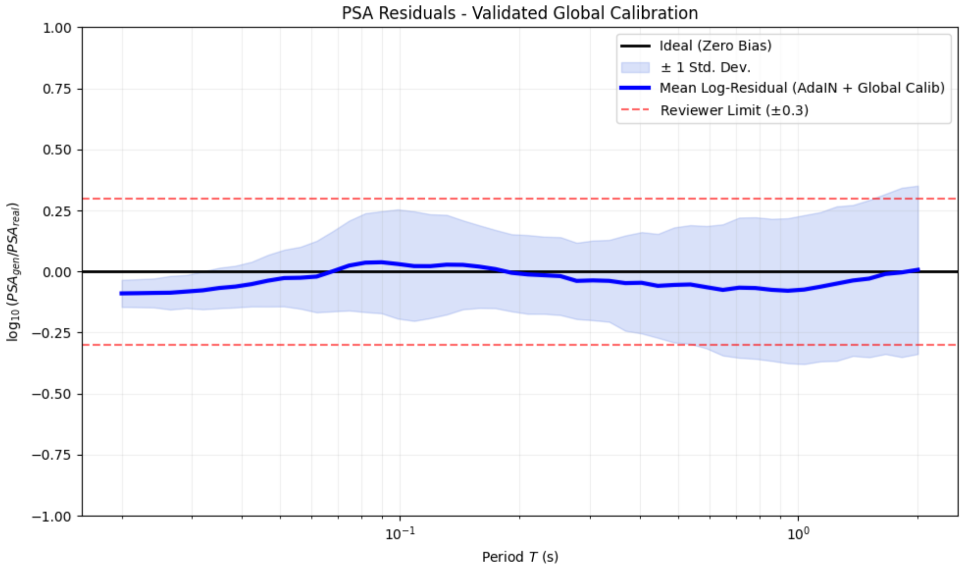
By marking each $\delta f = 0.0125$ Hz interval that exceeds a threshold of 10^{-3} as 1, it is possible to construct a histogram of non-zero frequency ranges. As a result, the difficulty in representing frequencies below 0.5 Hz is intrinsic to the data distribution. This numerical limitation restricts the Encoder and Decoder models, which, when trained on representative data, may not adequately capture very low-frequency components. This observation is supported by multiple analyses, where training demonstrates effectiveness primarily in higher frequency regions. In practice, the



(a) Frequency repartition for the broadband, $\mathbf{y}$.



(b) Frequency representation for the filtered, $\mathbf{x}$.



(c) Quantitative validation of spectral demand via Pseudo-Spectral Acceleration (PSA) residuals, after global spectral calibration. The blue shaded area represents ± 1 standard deviation.

Figure 8 Observation of the distribution of the number of frequency for values in every $\delta f = 0.125$ Hz over the threshold 10^{-3}. We also present an analysis of residual power spectral acceleration (PSA) in Figure 8c.

low-frequency component of the physical simulation is also available. Therefore, we combine the low-frequency segment of $\mathbf{x}$ ([0–1] Hz) with the calibrated high-frequency segment of the simulated signal ([1–50] Hz). As shown in Figure 8c, this integrated approach eliminates the spectral bias.

Quantitative spectral validation via pseudo-spectral acceleration (PSA) is presented in Figure 8c. To meet the stringent requirements for critical infrastructure, a global spectral calibration was integrated into the generative pipeline. This procedure, defined by the frequency-dependent transfer function $H_{calib}(f)$ in Equation 21, is designed to eliminate the systematic high-frequency overestimation ($T < 0.1$ s) identified in earlier iterations.

Figure 8c illustrates the statistical distribution of the log-misfit ($\log_{10}[\mathrm{PSA_{gen}}/\mathrm{PSA_{obs}}]$) for 5% damped spectra across the test ensemble. The solid blue line represents the **mean log-residual**, while the shaded region indicates the ± 1 **standard deviation**, capturing the aleatory variability inherent in the one-to-many mapping. The results demonstrate a robust dual-regime performance: at long periods ($T > 1.0$ s), the near-zero bias confirms that the generator successfully honors the deterministic physics-based backbone. In the high-frequency stochastic regime ($T < 0.1$ s), the calibration ensures that the generated signals strictly follow the empirical energy distribution. The mean log-residual remains centered near zero across the entire 0.1–50 Hz bandwidth, with a bias strictly within the ± 0.3 log-unit threshold. This level of accuracy renders the model suitable for engineering-grade applications, specifically the seismic response analysis of nuclear power plants.

$$H_{calib}(f) = \frac{\sum_{i=1}^{N} |Y_{real,i}(f)| + \epsilon}{\sum_{i=1}^{N} |Y_{gen,i}(f)| + \epsilon} \tag{21}$$

7 Application to the 2019 M_W 4.9 Le Teil earthquake

Lehmann et al. (2022) simulated the 2019 M_W4.9 Le Teil earthquake (France). This is an earthquake characterized by a moderate magnitude, but at shallow depth (≈1 km). Given the scarity near-field recording stations and geophysical measurements at that time, numerical simulations were essential to describe the earthquake. The simulations were conducted using a high-fidelity wave propagation code called SEM3D (Touhami et al., 2022), which has been widely utilized for simulating past earthquakes and evaluating the seismic response of nuclear sites and urban areas (Castro-Cruz et al., 2021, 2022).

The computational domain covered 80 km × 92 km × 79 km and was discretized with 18.3 million elements, enabling wave propagation up to 5 Hz with a minimum S-wave velocity of 2180 m/s. The simulations required approximately 61,440 CPU hours on the TGCC (France) supercomputing facilities. Following the regional framework of Lehmann et al. (2022), our analysis focuses on the three reference stations located outside the primary sediment basin: **CRU1, OGCB, OGDF** (see Figure 9). The selection of these bedrock sites is critical, as they provide the incident wave field required for modeling complex basin effects without the interference of local soil non-linearities.

The primary objective of this application is to evaluate the generalization capability of our trained model, SeismoALICE-v2, on a real-word, unseen dataset. Crucially, the model was applied in a **zero-shot configuration**, meaning no fine-tuning was performed on Le Teil records. The validation was extended to the full three-station network to ensure statistical robustness and avoid site-specific bias.

Table 1 summarizes the Goodness-of-Fit (GOF) scores across the network. The results show a high level of consistency, with network-wide averages of **3.40 ± 0.70 for the envelope (EG)** and **4.30 ± 0.60 for the phase (PG)**, falling within the "Fair" to "Good" categories. While the performance remains stable across all components and sites, station **CRU1** (located at the Cruas Nuclear Power Plant) is selected in Figure 10h to provide a detailed visual inspection of the one-to-many generative variants and their respective frequency content.

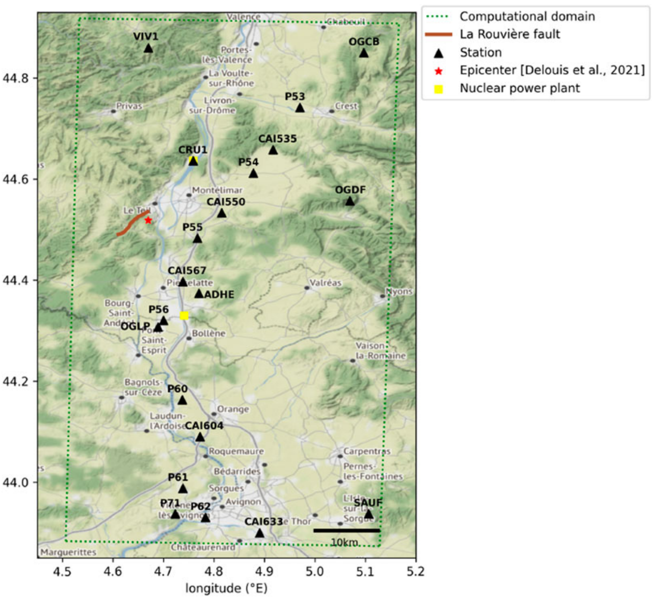


Figure 9 Map depicting the area impacted by the 2019 Le Teil earthquake in southeastern France. The computational domain analyzed in this study is outlined with a dotted box. Black triangles indicate the locations of velocimeters and accelerometers (Lehmann et al., 2022). Figure reproduced from Lehmann et al. (2022), licensed under CC-BY-4.0

Station	Component	Envelope GOF	Phase GOF
CRU1	Accel E	4.309391	3.988873
CRU1	Accel N	4.160553	3.907819
CRU1	Accel Z	4.084372	3.979842
OGCB	Accel E	3.212495	5.789821
OGCB	Accel N	3.237251	4.058073
OGCB	Accel Z	2.985186	4.515649
OGCD	Accel E	3.549475	4.193515
OGCD	Accel N	2.233536	4.377183
OGCD	Accel Z	2.782802	3.911926

Moyenne Globale:
Envelope GOF (EG) = 3.40 ± 0.70
Phase GOF (PG) = 4.30 ± 0.60

Table 1 GOF Final (0.1 – 10 Hz)

We thoroughly investigated the efficacy of both one-to-one mappings and one-to-many mappings by applying them to the seismogram CRU1. The one-to-many mapping is illustrated in Figure 10. We observe a comprehensive representation of the targeted signal. We observe the number of variants of the same signal in the frequency fields. Notice that the phase alignment is respected.

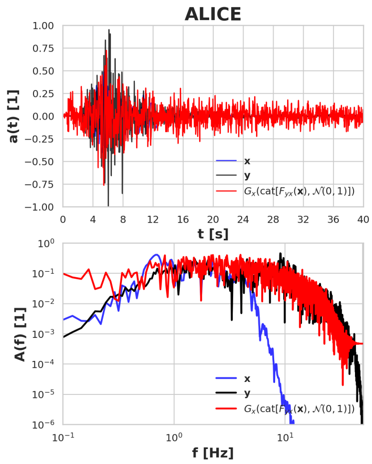
8 Discussion

8.1 Addressing Spectral Leakage and Filtering Bias

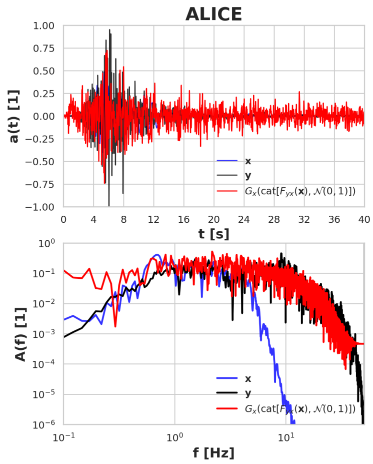
A critical challenge in low-frequency signal translation is potential leakage during the filtering process. As noted by Aquib and Mai (2024), standard causal or zero-phase filters (such as the Butterworth filter used in this study) do not possess an infinitely sharp "brick-wall" cutoff. Consequently, high frequency residuals may persist within the low-pass filtered input, x, potentially providing a "shortcut" for the generator during the super-resolution task.

While Aquib and Mai (2024) proposes a more aggressive Fourier-domain masking to remove all amplitude information beyond the cutoff, our approach relies on the **latent space factorization** and **Focal Frequency Loss (FFL)** as a clever regularizer. By penalizing the model in the frequency domain, FFL forces the generator to reconstruct the broadband spectrum from the abstract common latent manifold $\mathbf{z}_{xy}$ rather than simply amplifying leakage residuals. However, we acknowledge that the filtering bias remains a factor. Future iteration of the SeismoALICE framework could investigate the integration of a non-overlapping spectral masking layer to ensure that the generated high-frequency innovations are derived solely from the stochastic latent components rather than pre-existing residuals in the low-frequency backbone.

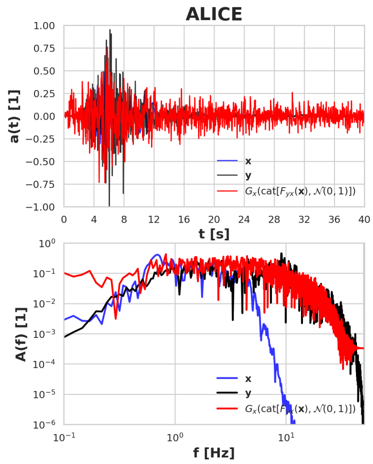
Recent work by Wright and Fayaz (2026) has introduced latent variables within hyperbolic geometry to achieve principal disentanglement in earthquake response spectrum modeling. While their approach of-



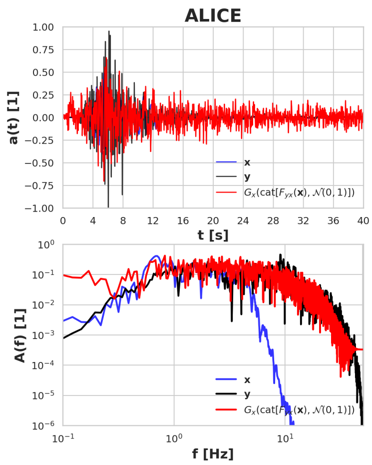
(a) CRU1 (Up-Down), Variant 1 of $G_y(F_{xy}^{xy}(\mathbf{x}), \mathcal{N}(0, I))$



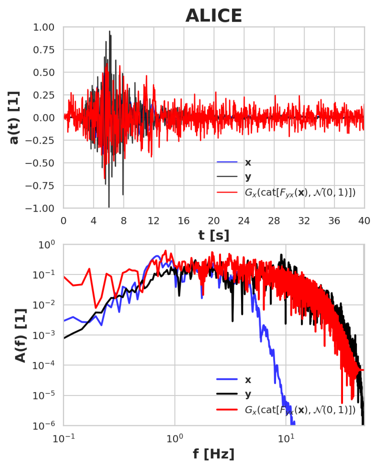
(b) CRU1 (Up-Down),Variant 2 of $G_y(F_{xy}^{xy}(\mathbf{x}), \mathcal{N}(0, I))$



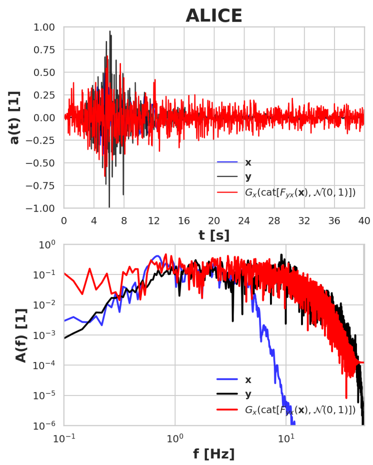
(c) CRU1 (Up-Down), Variant 3 of $G_y(F_{xy}^{xy}(\mathbf{x}), \mathcal{N}(0, I))$



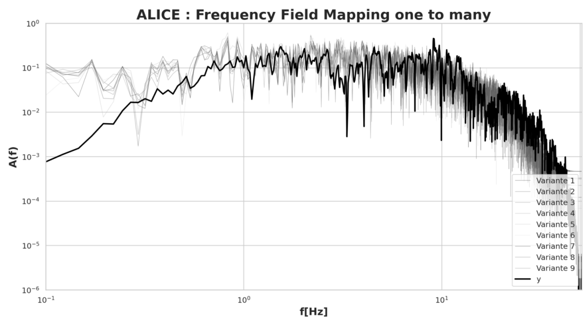
(d) CRU1 (Up-Down), Variant 4 of $G_y(F_{xy}^{xy}(\mathbf{x}), \mathcal{N}(0, I))$



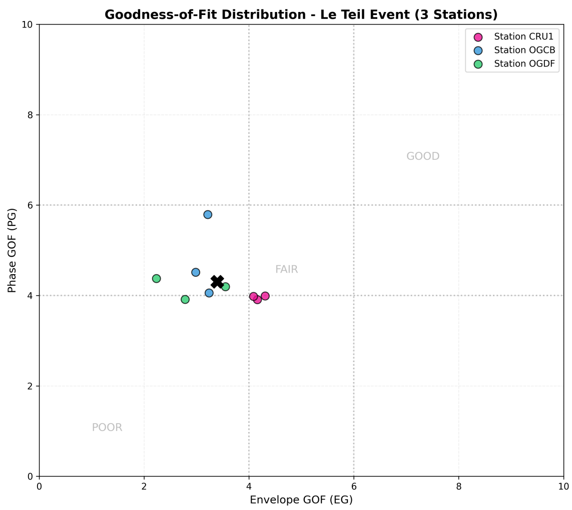
(e) CRU1 (Up-Down), Variant 5 of $G_y(F_{xy}^{xy}(\mathbf{x}), \mathcal{N}(0, I))$



(f) CRU1 (Up-Down), Variant 6 of $G_y(F_{xy}^{xy}(\mathbf{x}), \mathcal{N}(0, I))$



(g) Frequency field of the different variants from Splitted latent space of ALICE



(h) GOF for the 3 Different stations (CRU1, OGCB, OGCD).

Figure 10 Performance validation for the 2019 Le Teil earthquake. (a–f) Ensemble of six stochastic variants generated for station CRU1 (Up-Down component) compared to the real record (black). (g) Frequency-domain representation of the generated variants, highlighting the coverage of high-frequency innovations. (h) Distribution of Goodness-of-Fit (GOF) scores for the three reference stations (CRU1, OGCB, OGDF), demonstrating network-wide consistency.

fers a geometrically rigorous way to separate inter-event and intra-event variability, SeismoALICE-v2 addresses the more complex challenge of generating full-bandwidth accelerometer time histories. Such an approach requires capturing non-stationary phase information that spectrum-based models do not explicitly handle. However, the integration of such geometric priors into waveform generative models represents a promising direction for future research to further strengthen the empirical latent factorization observed in this study.

Furthermore, empirical observations during the training of the dual-branch encoder revealed a significant phenomenon concerning the specific latent subspace, $\mathbf{z}_{xx}$, associated with the physics-based input $\mathbf{x}$. When the optimization objective did not explicitly enforce the extraction of unique features from this branch, the model naturally converged toward a state of latent nullification, in which no relevant information was encoded into $\mathbf{z}_{xx}$. This outcome aligns with the spectral properties of the input data: since $\mathbf{x}$ is a band-limited numerical simulation (0–1 Hz), it lacks the high-frequency stochastic residual that would require a specific representation. This inherent architectural bias was strategically exploited to enforce a more robust factorization of the latent space. Allowing the model to minimize the information content in $\mathbf{z}_{xx}$ ensured that the common manifold, $\mathbf{z}_{xy}$, captured the entirety of the deterministic wavefield. This approach prevented the latent space from encoding redundant or non-physical noise, thereby ensuring that high-frequency innovations were derived exclusively from the stochastic broadband branch, $\mathbf{z}_{yy}$. This self-disentangling behavior provides empirical evidence supporting the validity of the latent space split, even in the absence of explicit mutual information penalties.

8.2 Robustness of our model

The robustness of the Conformer-based ALICE framework is validated through its application to the 2019 Le Teil earthquake. The successful synthesis of high-frequency content at station CRU1– an out-of-distribution geological environment relative to the training set – confirms that the model has captured generalized physical features of seismic wave propagation, effectively mitigating the risk of scaling-induced overfitting.

8.3 Transition from Normalized to Physical Amplitudes

In its current implementation, the model operates on normalized acceleration traces scaled by horizontal PGA. This normalization technique is standard in machine learning to ensure numerical stability and prevent gradient explosion driven by the extreme dynamic range of seismic amplitudes. However, denormalization is essential for practical earthquake engineering. While a simple multiplication by the predicted PGA—as stored in our metadata—provides a baseline, future developments will focus on a **conditioned two-stage synthesis**. In this pipeline, the Generator could be conditioned on scalar values such as moment magnitude M_W and epicentral distance R. This approach will preserve the frequency content of the acceleration signal while predicting the intensity required for site-specific structural analysis.

9 Conclusions and Perspectives

In conclusion, our investigation into data representation and generation methodology has been conducted with several noteworthy advanced techniques to understand and circumvent different bottlenecks and limitations. Initially, the SeismoALICE was designed to map any physics based on broadband data and, conversely, any recorded acceleration to simulated outputs of acceleration. Due to the limitations of this previous strategy, notably the duplicate resources and the contradictory loss, we have investigated and built a novel and more suitable architecture for times series, capable of viewing long-range dependencies to address the challenge of reconstruction and generation of ground motion acceleration and also for physically based data. A better understanding of each part of the problem enables and provides the keys to addressing the hybrid mapping so that each abstract representation caught in this latent value is mathematically interpretable. The successful application to the Le Teil earthquake demonstrates the potential of the SeismoALICE-v2 pipeline as a downstream generative tool for seismic hazard assessment, although further validation on larger-magnitude, truly out-of-distribution events remains necessary to fully establish its domain generalization limits.

Our first investigation to overcome previous limitations focused on developing a surrogate strategy around a shared latent space, providing valuable insights into performance. Based on the advancements made in our architecture design for the two groups of subsets, we designed a unified strategy that satisfies the limitations mentioned above, except for the one-to-many mapping.

To address the latter objective, we reinterpreted the input data and investigated a factorized latent space to overcome this limitation. A common part encodes information in the low-frequency portion of the ground motion (0–1 Hz), while the other part encodes the remaining frequency portion (1–30 Hz). Factorization into common and specific components enhances the model's interpretability and adaptability, improving performance. In addition, removing the hybrid loss from the learning process streamlined the approach, focusing on the most important components and eliminating contradictory or redundant corrections. Meanwhile, introducing the Gaussian-specific part added variability to the results while maintaining consistency, thus enriching the generative capabilities of the model. Despite the challenges encountered in the one-to-many mapping and some performance issues, the iterative development and adjustments have resulted in a more sophisticated architecture better aligned with the requirements of earthquake engineering tasks and probabilistic seismic hazard assessment.

Acknowledgments

Presentation of acknowledgment to Jeremy Fix, Fanny Lehmann, PhD, Gerardo Emanuel Granados, PhD, Hugo

Gabrieldis, and Boris Kratz.

During the preparation of this work the author used Grammarly in order to rephrase some sentences. After using this tool/service, the author reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data and Code Availability

The source code is available on Zenodo (Jacquet, 2025). The STEAD dataset of Mousavi et al. (2019) was used.

References

Aggarwal, C. C. *Neural Networks and Deep Learning: A Textbook*. Springer International Publishing, 2018. doi: 10.1007/978-3-319-94463-0.

Aquib, T. A. and Mai, P. M. Broadband Ground-Motion Simulations with Machine-Learning-Based High-Frequency Waves from Fourier Neural Operators. *Bulletin of the Seismological Society of America*, 114(6):2846–2868, 2024. doi: 10.1785/0120240027.

Arjovsky, M. and Bottou, L. Towards Principled Methods for Training Generative Adversarial Networks. *arXiv:1701.04862 [cs, stat]*, 2017. http://arxiv.org/abs/1701.04862.

Becker, E., Pandit, P., Rangan, S., and Fletcher, A. K. Instability and Local Minima in GAN Training with Kernel Discriminators, 2022. http://arxiv.org/abs/2208.09938.

Belghazi, M. I., Baratin, A., Rajeswar, S., Ozair, S., Bengio, Y., Courville, A., and Hjelm, R. D. MINE: Mutual Information Neural Estimation, 2021. doi: 10.48550/arXiv.1801.04062.

Bergmeister, A., Palgunadi, K. H., Bosisio, A., Ermert, L., Koroni, M., Perraudin, N., Dirmeier, S., and Meier, M.-A. High Resolution Seismic Waveform Generation using Denoising Diffusion. Unpublished arXiv:2410.19343, Oct. 2024. http://arxiv.org/abs/2410.19343.

Bi, Z., Nakata, N., Nakata, R., Ren, P., Wu, X., and Mahoney, M. W. Advancing data-driven broadband seismic wavefield simulation with multi-conditional diffusion model. Unpublished arXiv:2501.14348, Jan. 2025. doi: 10.48550/arXiv.2501.14348.

Brock, A., Donahue, J., and Simonyan, K. Large Scale GAN Training for High Fidelity Natural Image Synthesis. *arXiv:1809.11096 [cs, stat]*, 2019. http://arxiv.org/abs/1809.11096.

Castro-Cruz, D., Gatti, F., and Lopez-Caballero, F. High-fidelity broadband prediction of regional seismic response: a hybrid coupling of physics-based synthetic simulation and empirical Green functions. *Natural Hazards*, 108(2):1997–2031, 2021. doi: 10.1007/s11069-021-04766-x.

Castro-Cruz, D., Gatti, F., Lopez-Caballero, F., Hollender, F., El Haber, E., and Causse, M. Blind broad-band (0–10 Hz) numerical prediction of the 3-D near field seismic response of an *M* W6.0 extended fault scenario: application to the nuclear site of Cadarache (France). *Geophysical Journal International*, 232(1): 581–600, 2022. doi: 10.1093/gji/ggac346.

Donahue, J. and Simonyan, K. Large Scale Adversarial Representation Learning, 2019. doi: 10.48550/arXiv.1907.02544. Number: arXiv:1907.02544.

Donahue, J., Krähenbühl, P., and Darrell, T. Adversarial Feature Learning. Unpublished arXiv:1605.09782, 2017. doi: 10.48550/arXiv.1605.09782.

Dumoulin, V., Belghazi, I., Poole, B., Mastropietro, O., Lamb, A., Arjovsky, M., and Courville, A. Adversarially Learned Inference, 2017. doi: 10.48550/arXiv.1606.00704.

Esfahani, R. D. D., Cotton, F., Ohrnberger, M., and Scherbaum, F. TFCGAN: Nonstationary Ground-Motion Simulation in the Time–Frequency Domain Using Conditional Generative Adversarial Network (CGAN) and Phase Retrieval Methods. *Bulletin of the Seismological Society of America*, 113(1):453–467, 2023. doi: 10.1785/0120220068.

Florez, M. A., Caporale, M., Buabthong, P., Ross, Z. E., Asimaki, D., and Meier, M.-A. Data-driven Accelerogram Synthesis using Deep Generative Models. *arXiv:2011.09038 [physics, stat]*, 2020. http://arxiv.org/abs/2011.09038.

Gabrielidis, H., Gatti, F., and Vialle, S. Physics-based super-resolved simulation of 3D elastic wave propagation adopting scalable Diffusion Transformer. *Computer Physics Communications*, page 109930, 2025. doi: 10.1016/j.cpc.2025.109930.

Gatti, F. and Clouteau, D. Towards blending Physics-Based numerical simulations and seismic databases using Generative Adversarial Network. *Computer Methods in Applied Mechanics and Engineering*, 372:113421, 2020. doi: 10.1016/j.cma.2020.113421.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. Generative Adversarial Networks. *arXiv:1406.2661 [cs, stat]*, 2014. http://arxiv.org/abs/1406.2661.

Gulati, A., Qin, J., Chiu, C.-C., Parmar, N., Zhang, Y., Yu, J., Han, W., Wang, S., Zhang, Z., Wu, Y., and Pang, R. Conformer: Convolution-augmented Transformer for Speech Recognition, 2020. http://arxiv.org/abs/2005.08100.

Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A. Improved Training of Wasserstein GANs. *arXiv:1704.00028 [cs, stat]*, 2017. http://arxiv.org/abs/1704.00028.

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., and Hochreiter, S. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. Unpublished arXiv:1706.08500, 2018. doi: 10.48550/arXiv.1706.08500.

Hoffmann, J., Borgeaud, S., Mensch, A., Buchatskaya, E., Cai, T., Rutherford, E., Casas, D. d. L., Hendricks, L. A., Welbl, J., Clark, A., Hennigan, T., Noland, E., Millican, K., Driessche, G. v. d., Damoc, B., Guy, A., Osindero, S., Simonyan, K., Elsen, E., Rae, J. W., Vinyals, O., and Sifre, L. Training Compute-Optimal Large Language Models, 2022. doi: 10.48550/arXiv.2203.15556.

Huang, L., Liu, X., Lang, B., Yu, A. W., Wang, Y., and Li, B. Orthogonal Weight Normalization: Solution to Optimization over Multiple Dependent Stiefel Manifolds in Deep Neural Networks, 2017. doi: 10.48550/arXiv.1709.06079.

Ichimura, T., Fujita, K., Koyama, K., Kikuchi, Y., Kusakabe, R., Minami, K., Inoue, H., Nishizawa, S., Tsuji, M., Nishiki, T., Hori, M., Maddegedara, L., and Ueda, N. Fast scalable implicit solver with convergence of equation-based modeling and data-driven learning: earthquake city simulation on low-order unstructured finite element. In *Proceedings of the Platform for Advanced Scientific Computing Conference*, pages 1–12. ACM, 2021. doi: 10.1145/3468267.3470616.

Jacquet, G. jacquetgottfried/SeismoALICE-v2: SeismoALICE on scaled data [software], 2025. doi: 10.5281/zenodo.15799533.

Kamae, K., Irikura, K., and Pitarka, A. A technique for simulating strong ground motion using hybrid Green's function. *Bulletin of the Seismological Society of America*, 88(2):357–367, 1998. doi: 10.1785/BSSA0880020357.

Kumar, K., Kumar, R., de Boissiere, T., Gestin, L., Teoh, W. Z., Sotelo, J., de Brebisson, A., Bengio, Y., and Courville, A. MelGAN: Generative Adversarial Networks for Conditional Waveform Synthesis, 2019. doi: 10.48550/arXiv.1910.06711.

Lee, M. and Seok, J. Regularization Methods for Generative Adversarial Networks: An Overview of Recent Studies. Unpublished arXiv:2005.09165, 2020. doi: 10.48550/arXiv.2005.09165.

Lehmann, F., Gatti, F., Bertin, M., and Clouteau, D. Machine learning opportunities to conduct high-fidelity earthquake simulations in multi-scale heterogeneous geology. *Frontiers in Earth Science*, 10, 2022. https://www.frontiersin.org/articles/10.3389/feart.2022.1029160.

Lehmann, F., Gatti, F., Bertin, M., and Clouteau, D. Fourier Neural Operator Surrogate Model to Predict 3D Seismic Waves Propagation, 2023. http://arxiv.org/abs/2304.10242.

Li, C., Liu, H., Chen, C., Pu, Y., Chen, L., Henao, R., and Carin, L. ALICE: Towards Understanding Adversarial Learning for Joint Distribution Matching. *arXiv:1709.01215 [cs, stat]*, 2017. http://arxiv.org/abs/1709.01215.

McCallen, D., Petersson, A., Rodgers, A., Pitarka, A., Miah, M., Petrone, F., Sjogreen, B., Abrahamson, N., and Tang, H. EQSIM—A multidisciplinary framework for fault-to-structure earthquake simulations on exascale computers part I: Computational models and workflow. *Earthquake Spectra*, 37(2), 2021a. doi: 10.1177/8755293020970982.

McCallen, D., Petrone, F., Miah, M., Pitarka, A., Rodgers, A., and Abrahamson, N. EQSIM—A multidisciplinary framework for fault-to-structure earthquake simulations on exascale computers, part II: Regional simulations of building response. *Earthquake Spectra*, 37(2), 2021b. doi: 10.1177/8755293020970980.

Miyato, T., Kataoka, T., Koyama, M., and Yoshida, Y. Spectral Normalization for Generative Adversarial Networks. *arXiv:1802.05957 [cs, stat]*, 2018. http://arxiv.org/abs/1802.05957.

Mousavi, S. M., Sheng, Y., Zhu, W., and Beroza, G. C. STanford EArthquake Dataset (STEAD): A Global Data Set of Seismic Signals for AI. *IEEE Access*, 7:179464–179476, 2019. doi: 10.1109/ACCESS.2019.2947848.

Mousavi, S. M., Ellsworth, W. L., Zhu, W., Chuang, L. Y., and Beroza, G. C. Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature Communications*, 11(1):3952, 2020. doi: 10.1038/s41467-020-17591-w.

Nakata, R., Nakata, N., Ren, P., Bi, Z., Lacour, M., Erichson, B., and Mahoney, M. W. Simulating seismic wavefields using generative artificial intelligence. *The Leading Edge*, 44(2):123–132, 2025. doi: 10.1190/tle44020123.1.

Paolucci, R., Mazzieri, I., Smerzini, C., and Stupazzini, M. Physics-Based Earthquake Ground Shaking Scenarios in Large Urban Areas. In Ansal, A., editor, *Perspectives on European Earthquake Engineering and Seismology*, volume 34, pages 331–359. Springer International Publishing, 2014. doi: 10.1007/978-3-319-07118-3_10. Series Title: Geotechnical, Geological and Earthquake Engineering.

Paolucci, R., Gatti, F., Infantino, M., Smerzini, C., Özcebe, A. G., and Stupazzini, M. Broadband Ground Motions from 3D Physics-Based Numerical Simulations Using Artificial Neural Networks-Broadband Ground Motions from 3D PBSs Using ANNs. *Bulletin of the Seismological Society of America*, 108(3):1272–1286, 2018. doi: 10.1785/0120170293.

Pitarka, A. and Mellors, R. Using Dense Array Waveform Correlations to Build a Velocity Model with Stochastic Variability. *Bulletin of the Seismological Society of America*, 111(4):2021–2041, 2021. doi: 10.1785/0120200206.

Rekoske, J. M., Gabriel, A.-A., and May, D. A. Instantaneous physics-based ground motion maps using reduced-order modeling. *Journal of Geophysical Research: Solid Earth*, 128(8): e2023JB026975, 2023-08. doi: 10.1029/2023JB026975.

Salimans, T. and Kingma, D. P. Weight Normalization: A Simple Reparameterization to Accelerate Training of Deep Neural Networks. Unpublished arXiv:1602.07868, 2016. doi: 10.48550/arXiv.1602.07868.

Smerzini, C. and Villani, M. Broadband Numerical Simulations in Complex Near-Field Geological Configurations: The Case of the 2009 Mw 6.3 L'Aquila Earthquake. *Bulletin of the Seismological Society of America*, 102(6):2436–2451, 2012. doi: 10.1785/0120120002.

Smith, L. N. Cyclical Learning Rates for Training Neural Networks, 2017. http://arxiv.org/abs/1506.01186.

Sochala, P., De Martin, F., and Le Maître, O. MODEL REDUCTION FOR LARGE-SCALE EARTHQUAKE SIMULATION IN AN UNCERTAIN 3D MEDIUM. *International Journal for Uncertainty Quantification*, 10(2):101–127, 2020. doi: 10.1615/Int.J.UncertaintyQuantification.2020031165.

Touhami, S., Gatti, F., Lopez-Caballero, F., Cottereau, R., De Abreu Corrêa, L., Aubry, L., and Clouteau, D. SEM3D: A 3D High-Fidelity Numerical Earthquake Simulator for Broadband (0–10 Hz) Seismic Response Prediction at a Regional Scale. *Geosciences*, 12(3):112, 2022. doi: 10.3390/geosciences12030112.

Wright, A. and Fayaz, J. Structured generative modelling of earthquake response spectra with hierarchical latent variables in hyperbolic geometry. *Scientific Reports*, 16(1):906, 2026. doi: 10.1038/s41598-025-29902-6.

Zhang, H., Goodfellow, I., Metaxas, D., and Odena, A. Self-Attention Generative Adversarial Networks, 2019. http://arxiv.org/abs/1805.08318.

A Annex

A.1 Optimization and Regularization

(i) **Model peformance**
Our architecture on time series data have been compared using Classical CNN and Attention CNN with conformer. The performance of the model is presented using Goodness of Fit Score in the Figure A.1.

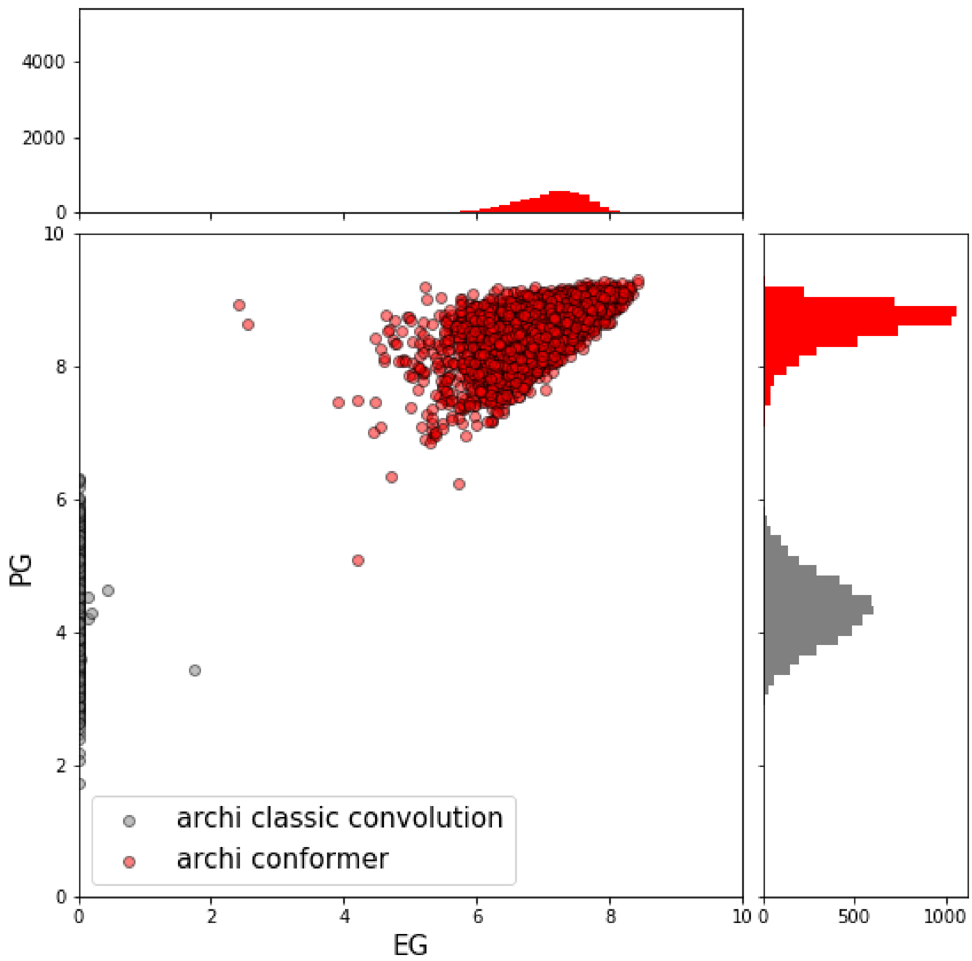


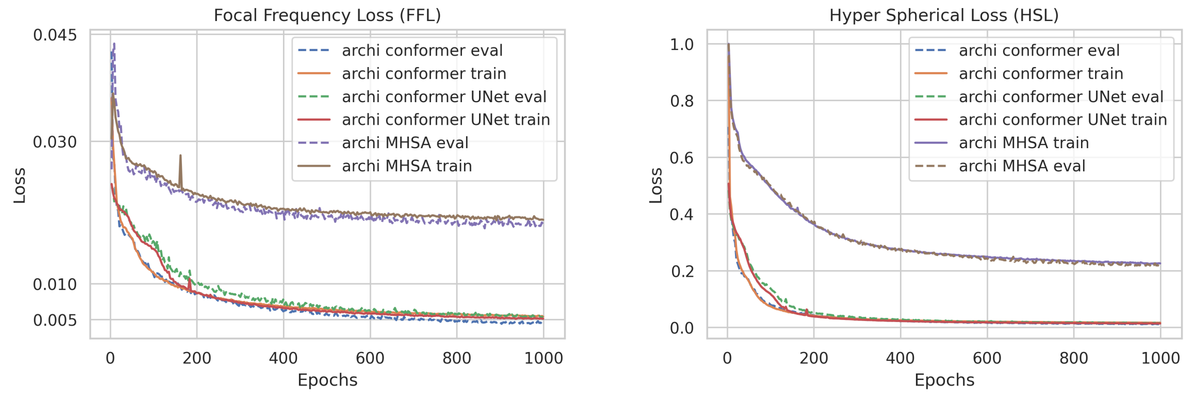
Figure A.1 The comparasion of the quality of the reconstruction of the seismogram time series ($G_y(F_y(\mathbf{y})) \sim \mathbf{y}$) of the STEAD is presented using both different architectures. The GOF proformed on the test set reveal the performance of our architectures compared to classic convolutions.

(ii) **Optimization loss**
The optimization is done through the loss of Focal Frequency and the hyper-spherical loss. After many tests, we kept the Conformer architecture with a bottleneck dimension 256. The result is not significantly different from the dimension 128, but the size 256 is tailored for the hybrid mapping, which we will train later. View Figure A.2. For the sake of performance, we have trained our loss function with different adversarial objectives, with HingeLoss (View Equation 22) or Least Square Mean Loss (View Equation 23) or Earth Motion Loss (Wasserstein). We found that the result is not significantly different between BCEWithlogit. The variant of adversarial loss with HingeLoss has an equation:

$$\begin{aligned}\mathcal{L}_{\mathcal{D}} &= \mathbb{E}_{y\sim P_{data}}\left[\max(0, 1 - D_{yz}(\mathbf{y}, F_y(\mathbf{y})))\right] + \mathbb{E}_{z\sim P_z}\left[\max(0, 1 + D_{yz}(G_y(\mathbf{z})), \mathbf{z}_y)\right] \\ \mathcal{L}_{\mathcal{G}} &= \mathbb{E}_{y\sim P_{data}}\left[D_{yz}(\mathbf{y}, F_y(\mathbf{y}))\right] + \mathbb{E}_{z\sim P_z}\left[-D_{yz}(G_y(\mathbf{z}), \mathbf{z}_y)\right]\end{aligned} \tag{22}$$

For Least Square Mean Loss :

$$\begin{aligned}\mathcal{L}_{\mathcal{D}} &= \mathbb{E}_{y\sim P_{data}}\left[(D_{yz}(\mathbf{y}, F_y(\mathbf{y})) - 1)^2\right] + \mathbb{E}_{\mathbf{z}_y\sim P_{\mathbf{z}_y}}\left[(D_{yz}(G_y(\mathbf{z}_y), \mathbf{z}_y))\right] \\ \mathcal{L}_{\mathcal{G}} &= \mathbb{E}_{y\sim P_{data}}\left[(D_{yz}(\mathbf{y}, F_y(\mathbf{y})) - 1)^2\right] + \mathbb{E}_{\mathbf{z}_y\sim P_{\mathbf{z}_y}}\left[(D_{yz}(G_y(\mathbf{z}_y), \mathbf{z}_y) - 1)^2\right]\end{aligned} \tag{23}$$



(a) FFL for the unitary test of broadband with different encoder/decoder architectures.

(b) HSL values range between 1.0 and 0. A lower HSL indicates better reconstruction.

Figure A.2 General investigation of architectures adapted for time series in Broadband signals. The comparison of different losses highlights the advantage of using the Conformer architecture for adversarial training. The HSL can be considered as a metric.

The Earth Motion (EM) Loss seems not properly adapted (WGAN and WGAN-GP) for our case.

The cycle consistency is made explicitly with FFL and HSL :

$$\mathcal{L}_{\text{Cycle}}(F_y, G_y, \mathbf{y}) = \text{FFL}(\mathbf{x}, G_y(F_y(\mathbf{y}))) + \text{HSL}(\mathbf{x}, G_y(F_y(\mathbf{y}))) \tag{24}$$

(iii) **Regularization technique**
We tested different weight regularization techniques designed to stabilize the training (Spectral NormMiyato et al. (2018), Weight Orthogonality Huang et al. (2017), Weight Norm Salimans and Kingma (2016)). *" Weight normalisation and penalisation methods improve the stability and performance of sample generation by GANs."* (Lee and Seok, 2020). The Spectral normalization is formulated in Equation 19.
Brock formulates the Orthogonal regularization as:

$$\begin{aligned} R_\beta(\mathbf{W}) : \\ = \beta \|\mathbf{W}^\top \mathbf{W} - \mathbf{I}\|_F^2 \end{aligned} \tag{25}$$

Formula in which $\|\cdot\|_F$ is the Forbinius norm, $\mathbf{I}$ the identity matrix. β is a hyper parameter, commonly choose with 10^{-4}.
The Weight Norm technique is a reparametrization method that decouples the magnitude of the weight from its directions. Assuming a neural network operation is expressed by :

$$y = \phi(\mathbf{w} \cdot \mathbf{x} + b)$$

ϕ is an activation functions , $\mathbf{w}$ is the k-dimensional matrix, $\mathbf{x}$ a k-dimensional vector of input. The weight regularization consists of the research of a parameter vector $\mathbf{v}$ and a scalar g, such a way that:

$$\mathbf{w} = \frac{g}{\|\mathbf{v}\|}\mathbf{v} \tag{26}$$

The norm of $\mathbf{w}$ is independent of $\mathbf{v}$ and equals g. Other techniques were listed by Lee *et al.*,Lee and Seok (2020). As highlighted by Kumar *et al.*, 2019 Kumar et al. (2019), even though the normalization technique should improve the training, some weight normalization may work better depending on the data set. In our case, the solution that works is the normalization of the weight of the layers with Spectral Normalization.

(iv) **Mode collapse and Mode Dropping solutions**
We train either Early Stopping or CyclicLR to avoid mode collapse. Another common form of regularisation is early stopping, which stops the gradient descent after a few iterations when the error on a retained test set increases. Early stopping is a regularizer because it effectively restricts the parameter space Aggarwal (2018). The CycleLR (cyclical learning rate), Smith *et al.*, 2017 Smith (2017), is a learning scheduler that makes the learning rate oscillate between two values. This strategy was used for fast convergence and increased performance as a scheduler over the optimization parameters of the discriminator.

(v) **Training dynamics**
The ALICE training is performed using TTUR (Two Time-Scale Update Rule), proposed by Heusel *et al.*, 2018 Heusel et al. (2018). The optimization is done using Adam optimizer. The learning rate of the encoder/decoder is both $4 \cdot 10^{-4}$, and the learning rate of the discriminators is 10^{-4}. The data set contains 128,000 3D signals. Finally, the G.O.F serves as a metric.

The training process utilized 4 Graphics Processing Unit (GPUs) A100 leveraging CUDA 11.7. Our model underwent rigorous training across multiple servers, including the *Mesocentre* server at CentralSupelec, the *Metz* server, and finally, the *Maison de la Simulation*, Jean-Zay. The training phase was executed over 500 epochs using a batch size of 64. On a single node equipped with 4 NVIDIA A100 (80GB) GPUs, the total training time was approximately 72 hours (totaling 288 GPU-hours). Once trained, the model offers significant computational advantages for downstream applications: while a high-fidelity physics-based simulation of a 60 second event using SEM3D can require several thousand CPU-hours, our generative model performs the broadband synthesis in approximately 45 milliseconds per three-component station on a single consumer-grade GPU. This represents a multi-order-of-magnitude reduction in the time required for site-specific ground motion prediction.

A.2 Generator part

The different parts of the generators are presented in Table A.2 and Table A.3.

Coodness-of-Fit	
Verbal Value	**Numerical Value**
Excellent	8-10
Good	6-8
Fair	4-6
Poor	0-4

Table A.1 Verbal representation of the discrete G.O.F score

Layer	Opeation	Input Size	Fonction Activation	Stride	Output
ConvBlock I	Downsampling	[3, 4096]	LeakyReLU	2	[16, 2048]
ConvBlock II	Downsampling	[16, 2048]	LeakyReLU	2	[32, 1024]
ConvBlock III	Downsampling	[32, 1024]	LeakyReLU	2	[64, 512]
ConvBlock IV	Downsampling	[64, 512]	LeakyReLU	2	[128,256]
Conformer	Attention	[128,256]	–	–	[128,256]
ConvBlock V	Downsampling	[128,256]	LeakyReLU	2	[256,128]
ConvBlock VI	Downsampling	[256,128]	LeakyReLU	2	[512,64]
Conformer	Attention	[512,64]	–	–	[512,64]
Reshape	–	[512,64]	–	–	[32768]
Linear Layer	Projection	[32768]	–	–	[256]

Table A.2 Architecture of the Encoder

A.3 Discriminators part

More details about the discriminators part is presented in the tables A.4 and A.5.

A.4 Experiments

(i) **Splitting the latent space**
We have observed different effects of the latent space. One of the significant issues when concatenating the two different vectors of latent space is that the neural network only uses one latent space and focuses on the interpretability of only one vector. We have observed that this is the natural behaviour of the neural network. The reason behind this behaviour is that the same encoder does not constraint the $\mathbf{z}_{xx}$, while for the ground motion pass through $\mathbf{z}_{yy}$ is essential for its reconstruction and variability. In previous investigations, we have try reduce to zero the specific part of the physics-based simulations :

$$\|F_{xy}^{xx}(\mathbf{x})\|_2^2 = \|\mathbf{z}_{xx}\|_2^2 \tag{27}$$

Layer	Opeation	Input Size	Fonction Activation	Stride	Output
Linear	Projection	[256]	–	–	[32768]
Reshape	–	[32768]	–	–	[512,64]
ConvTransBlock I	Upsampling	[512,64]	LeakyReLU	2	[256, 128]
ConvTransBlock II	Upsampling	[128, 128]	LeakyReLU	2	[64, 256]
Conformer	Attention	[64,256]	LeakyReLU	2	[64, 256]
ConvTransBlock III	Downsampling	[64, 256]	LeakyReLU	2	[32, 1024]
ConvTransBlock IV	Attention	[32, 1024]	–	–	[16, 2048]
ConvTransBlock V	Downsampling	[16, 2048]	LeakyReLU	2	[16, 2048]
Conformer	Attention	[16,2048]	LeakyReLU	2	[16,2048]
ConvTransBlock VI	Downsampling	[16,2048]	Tanh	–	[3,4096]

Table A.3 Architecture of the Decoder for SeismoALICE

Layer	Opeation	Input Size	Fonction Activation	Stride	Output
ResBlockDense I	Downsampling	[512]	LeakyReLU	2	[512]
ResBlockDense II	Downsampling	[512]	LeakyReLU	2	[512]
ResBlockDense III	Downsampling	[512]	LeakyReLU	2	[512]
Linear Layer	Projection	[512]	–	–	[1]

Table A.4 Architecture of the Discriminator $D_{s_{yz}}$

Layer	Opeation	Input Size	Fonction Activation	Stride	Output
ResBlockDense I	Downsampling	[256]	LeakyReLU	2	[256]
ResBlockDense II	Downsampling	[256]	LeakyReLU	2	[256]
ResBlockDense III	Downsampling	[256]	LeakyReLU	2	[256]
Linear Layer	Projection	[256]	–	–	[256]

Table A.5 Architecture of the Discriminator $D_{s_{zy}}$

The nullification of the specific part, $\mathbf{z}_{xx}$, also affects the behaviour of the specific part of the broadband, $\mathbf{z}_{yy}$. Then, a drastic diminishing of the latent dimension to force the rearrangement of the latent space was not efficient either.

(ii) **Injection of the specific latent distribution**
The AdaIN equation, initially conceived for style transfer, has proven practical for accommodating one-to-many output scenarios. This strategic integration "aligns the mean and variance of the content features with those of the style features.". Particularly, the specific component of the latent space, intended for capturing high-frequency information, is propagated throughout all subsequent layers during the up-sampling process via Adaptive Instance Normalization (AdaIN).

$$\mathrm{AdaIN}(\mathbf{c}, \mathbf{s}) = \sigma(\mathbf{s}) \left(\frac{\mathbf{c} - \mu(\mathbf{c})}{\sigma(\mathbf{c})} \right) + \mu(\mathbf{s}) \tag{28}$$

The formula in which $\mathbf{c}$ is the content, s the style input $\sigma(\mathbf{s})$ and $\mu(\mathbf{s})$ are respectively its standard deviation and mean. We have tried to inject $\mathbf{z}_{yy}$ as a style when keeping $\mathbf{z}_{xy}$ and $\mathbf{z}_{yx}$ as common part. However, if this strategy is plausible for managing one-to-many mapping, this performance significantly affects the reconstructions. This tentative approach has not been adapted to our objectives.

(iii) **Variability of the representations**
To improve reconstruction quality and output variability, we investigated decoder configurations. We tested the classic Conformer decoder and variants with AdaIN and U-Net (Conformer+AdaIN, Conformer+AdaIN+UNet). These variations aimed to observe how specific aspects are passed to the generator (G_y) and interpreted.

(iv) **Ablation**
We have removed the explicit cycle in the hybrid loss (View Equation 29).

$$\mathcal{L}^{yx}_{\text{hybrid}} = \mathrm{FFL}(\mathbf{y}, G_y(F^{xx}_{xy}(\mathbf{x}), F^{yy}_{xy}(\mathbf{y}))) + \mathrm{HSL}(\mathbf{y}, G_y(F^{xx}_{xy}(\mathbf{x}), F^{yy}_{xy}(\mathbf{y}))) \tag{29}$$

This ablation aims to test the validity of this loss for training. However, our tests reveal that the performance of the hybrid mapping is not affected, expressing the robustness of our architecture.

A.5 Gaussian Distribution Part

We present in Figure A.4 and Figure A.3 the proof that our model respect the assumption that it could generate a realistic accelerogram by sampling Gaussian noise. In fact, the successful unconditional synthesis of broadband signals from $\mathcal{N}(0, I)$ serves as empirical verification that the adversarial objective has successfully regularized the latent space toward the Gaussian prior, satisfying the marginal consistency requirement of the ALICE framework.

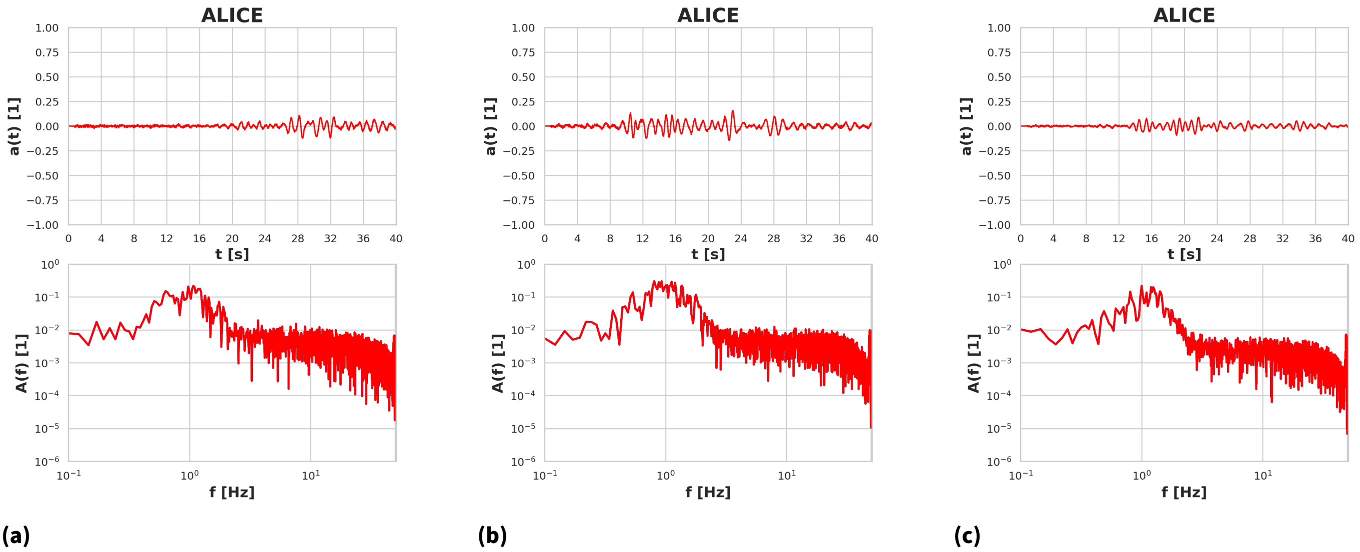


Figure A.3 We present the result for $G_x(\mathbf{z}_x)$. This is the proof of objective VI in Section § 5.

A.6 Comparison of different architecture for the Encoder and Decoder Network

We have tested diverse combination of neural network architecture before conclude to the must appropriate one four our training. View Figure A.5.

A.7 Compare irrelevant aspects with different architectures

Our investigation has revealed, the different degradation issues on one-to-many mapping issues. Certain complex neural architectures exhibit a tendency to suppress low-amplitude signals in the early temporal range while maintaining fidelity to target amplitude values in the subsequent time domain. To mitigate this temporal bias we use Conformer architecture only as attention technic. View Figure A.6.

A.8 Quantitative Arrival time issues

Our analysis revealed degradation artifacts in the one-to-many mapping procedure, particularly affecting the temporal accuracy of seismic phase arrivals. The predicted arrival times for both P-waves and S-waves deviated from their expected values, indicating a systematic timing error in the mapping process. To address this issue, we introduced a specialized loss function that constrains both the original and hybrid signals during the automated time-picking procedure, thereby ensuring temporal consistency between the input and output waveforms. View Figure A.7.

Architecture	#Param. (M)	Broad.		PBS		Hybrid	
		EG	PG	EG	PG	EG	PG
Conformer	219,44	5.70 ± 1.07	8.21 ± 0.46	6.20 ± 1.35	9.13 ± 0.23	5.80 ± 1.07	8.32 ± 0.46
Conformer+AdaIN	217,40	6.41 ± 0.94	8.47 ± 0.43	5.52 ± 1.37	9.22 ± 0.23	5.12 ± 1.34	7.36 ± 0.87
Conformer+AdaIN+Unet	231,00	6.62 ± 0.80	8.72 ± 0.40	6.11 ± 1.39	9.21 ± 0.19	5.69 ± 0.97	7.87 ± 0.75

Table A.6 Table of Architecture of Generators for strategy Unified training ALICE with split latent space

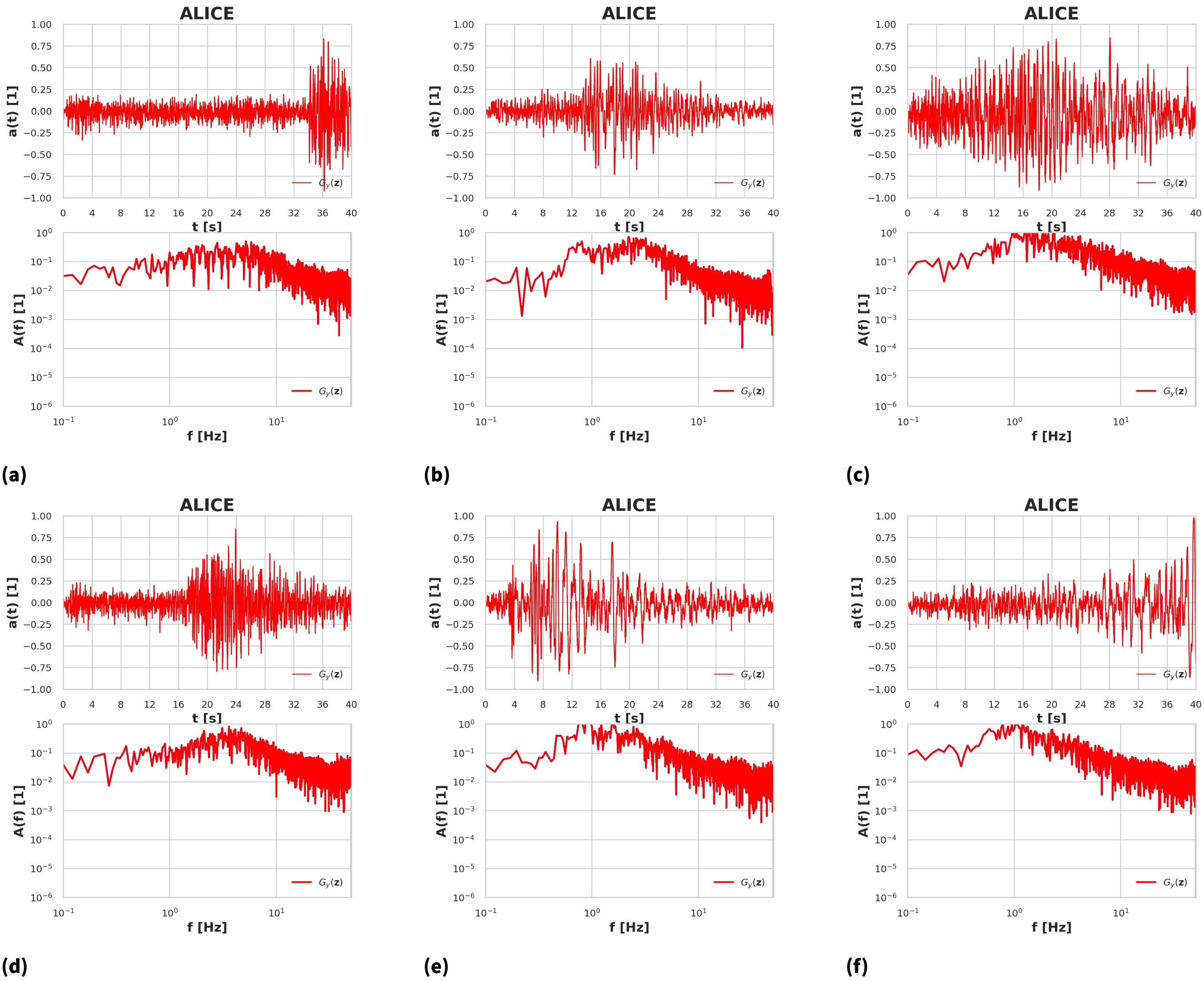


Figure A.4 We present the unconditional generation of broadband data from $\mathbf{z}_y \sim \mathcal{N}(0, \mathbf{I}) : G_y(\mathbf{z})$. The purpose of presenting different values is to represent the range of signals generated by our model visually. This exploration contributes to a comprehensive understanding of the model's performance and capacity to capture variability in broadband signal generation.

A.9 Quantitative Arrival Time Analysis (Ablation Study)

As shown in the Table A.7, the baseline model suffered from significant timing drift, primarily due to the zero-padding bias (samples 4096 to 6000) Which attracts the generated energy toward the window boundaries. The introduction of a binary validity mask and the $\mathcal{L}_{EQT}$ regularizer effectively "pins" the P and S arrivals to their physical locations. The Median Absolute Error (MEdAE) is reduced by approximately 95%, reaching sub-second precision (0.19 – 0.22 s), which is well within the acceptable margin for seismic structural analysis.

Configuration	MedAE P-wave (s)	MedAE S-wave (s)
Baseline (Raw / Unmasked)	6.54 s	3.96 s
SeismoALICE-v2 (With Mask + $\mathcal{L}_{EQT}$)	0.19 s	0.22 s

Table A.7 Median Absolute Error (MedAE) for P-wave and S-wave detection

A.10 Distributional Coverage and One-to-Many Mapping

To assess the probabilistic soundness of the SeismoALICE-v2 generator and detect potential mode-dropping, we evaluated the distributional coverage of the one-to-many mapping. For a fixed physics-based input, we generated an ensemble of stochastic realizations and compared their scalar intensity measures—Peak Ground Acceleration (PGA), Arias Intensity (I_a), and significant duration (D_{5-95})—against matched real records. Figure A.8 illustrates these distributions:

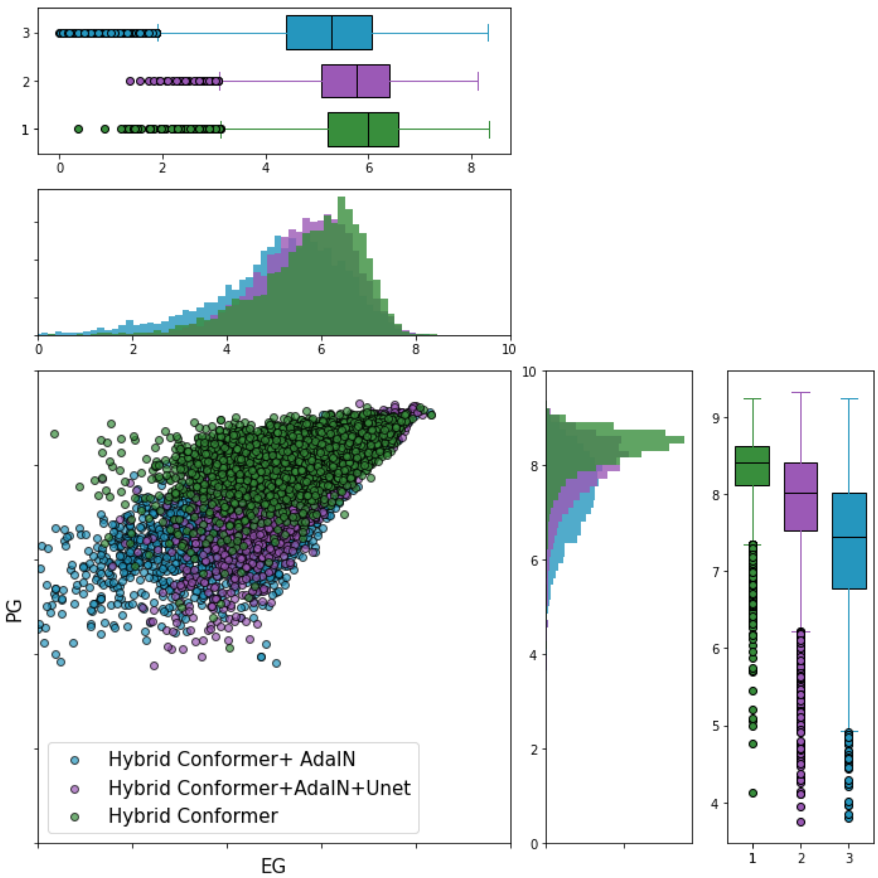
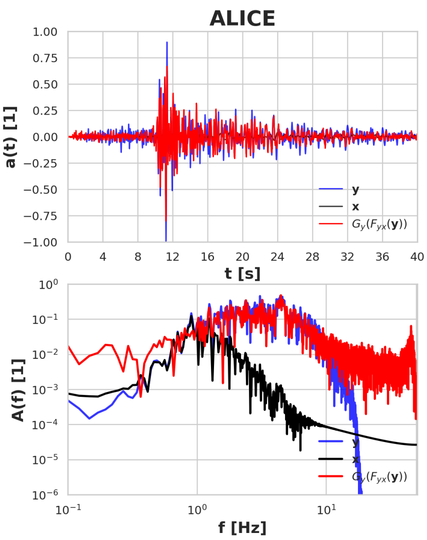
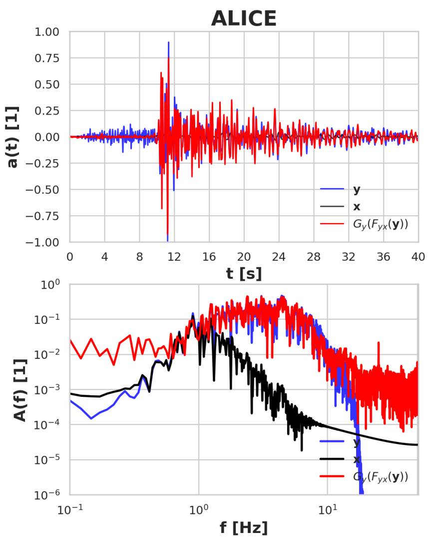


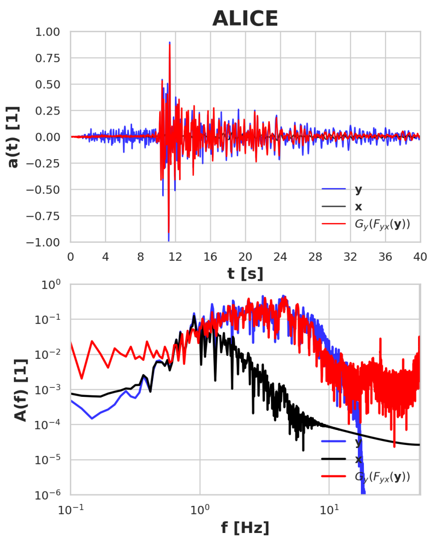
Figure A.5 List of G.O.F for different models trained for ALICE split latent space. We have tested different architectures. Encoder and Decoder made with Conformer, Conformer+ AdaIN and Conformer + AdaIN + Unet



(a) Quality reconstruction of a signal with the Conformer architecture



(b) Quality reconstruction of a signal with the Conformer+Unet architecture



(c) Quality reconstruction of a signal with the Conformer+AdaIN+UNet architecture

Figure A.6 Comparative analysis of temporal reconstruction artifacts across model variants, $G_y(F_{xy}(\mathbf{y}))$. Readers should observe that while the standalone Conformer A.6a preserves the signal's stochastic texture and late-stage coda, the more complex architectures exhibit a significant failure to reconstruct the accelerogram during its **early stages**. In A.6b and A.6c, the generators systematically underestimate the **low-amplitude components** of the signal (pre-arrival and P-wave onset) while only preserving high-amplitude peaks. This suppression of ambient noise and initial transients suggests that the addition of UNet skip-connections and AdaIN layers introduces an undesirable inductive bias that filters out crucial low-energy information. This results in a "dead" pre-arrival window, ultimately compromising phase-picking accuracy and the overall physical realism of the synthesized ground motion.

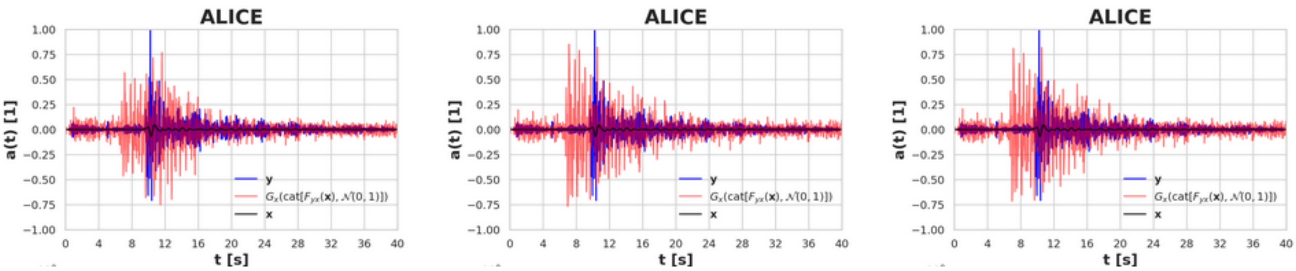


Figure A.7 Illustration of non-alignment of phase for a seismogram. We solve this issue with a loss performed through the EQTransformers (Equation 16).

- **Arias Intensity (I_a):** The histograms show an almost perfect overlap ($D = 0.116, W_1 = 0.033$). This agreement confirms the model's ability to respect the cumulative energy balance, avoiding the "energy-leakage" artifacts common in standard GANs.
- **Significant Duration (D_{5-95}):** The generated ensemble accurately captures the temporal evolution of ground motion without being too impulsive or artificially prolonged. This ensures that the non-stationary envelope of the seismic signal is preserved.
- **Peak Ground Acceleration (PGA):** While the KS distance is high ($D = 0.976$), this is a known numerical artifact of the dataset preprocessing. Real records are strictly normalized to a peak value of 1.0, creating a Dirac delta distribution with zero variance. The generative model, by injecting stochastic latent noise, naturally introduces physically plausible variability around this mean. The low Wasserstein distance ($W_1 = 0.185$) confirms that the model successfully respects amplitude constraints while providing the diversity required for stochastic simulations.

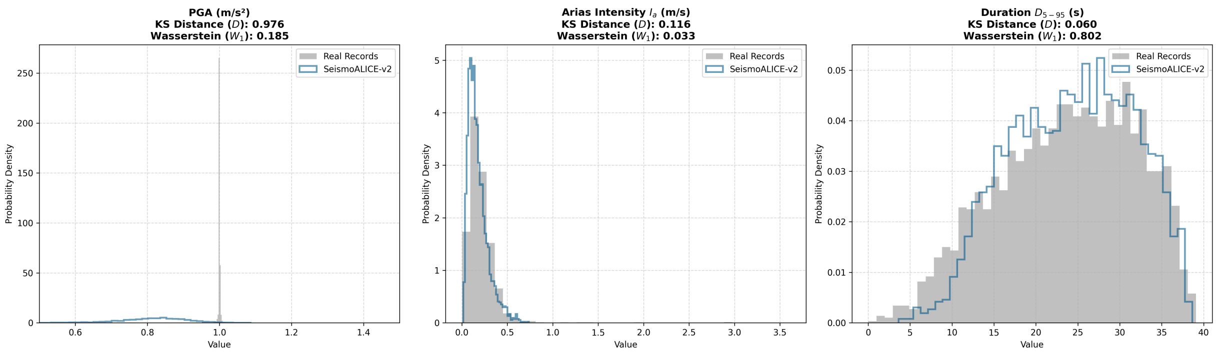


Figure A.8 Distributional coverage analysis for One-to-Many generation. Histograms compare real records (blue) and SeismoALICE-v2 generated ensembles (orange) for PGA, I_a, and D_{5-95}. Metrics include Kolmogorov-Smirnov (D) and first Wasserstein distance (W_1).