

<https://doi.org/10.1038/s42005-024-01912-2>

On the design and evaluation of generative models in high energy density physics

Check for updates

Ankita Shukla¹✉, Yamen Mubarka², Rushil Anirudh², Eugene Kur², Derek Mariscal², Blagoje Djordjevic², Bogdan Kustowski², Kelly Swanson², Brian Spears², Peer-Timo Bremer², Tammy Ma², Pavan Turaga³ & Jayaraman J. Thiagarajan²

Understanding high energy density physics (HEDP) is critical for advancements in fusion energy and astrophysics. The computational demands of the computer models used for HEDP studies have led researchers to explore deep learning methods to enhance simulation efficiency. This paper introduces HEDP-Gen, a framework for training and evaluating generative models tailored for HEDP. Central to HEDP-Gen is Geom-WAE—a generalized Wasserstein auto-encoder accommodating both Euclidean and non-Euclidean latent spaces. HEDP-Gen establishes a rigorous evaluation standard, assessing not only reconstruction fidelity but also scientific validity, sample diversity, and latent space utility in geodesic interpolation and attribute traversal. A case study using hyperbolic geometry (Poincaré ball prior) demonstrates that non-Euclidean priors yield scientifically valid samples and stronger generalization in downstream tasks, advantages often missed by conventional reconstruction metrics.

Understanding high energy density physics (HEDP)¹ has profound implications in fusion energy research^{2–4}, astrophysics^{5–7} and the development of advanced materials^{8,9}. While powerful lasers¹⁰, particle accelerators¹¹, and other advanced tools have been traditionally used for experimental studies, computer models have become an integral part of most ongoing efforts in HEDP. In particular, synthetic diagnostic data from these computer models have been leveraged to evaluate different theories when the corresponding experiments are otherwise difficult or very expensive to create in a laboratory¹². High-performance computing capabilities have steadily increased over time, permitting larger parameter spaces to be probed, and/or higher fidelity simulations to be performed. However this increased capability still struggles to keep pace with increasingly complex experiments, and an increasing need to capture an ever-growing variety of multi-scale and multi-dimensional phenomena. Fully leveraging machine learning presents the possibility of probing larger design parameter spaces at higher simulation fidelity for a fixed amount of computing resource. This has inspired researchers to investigate the application of machine learning (ML) methods^{13,14}, specifically deep learning (DL)^{15,16}, to improve the simulation efficiency and thereby support analysis at scales not currently feasible. More specifically, deep learning-powered scientific models have emerged in applications including astrophysics, cosmology¹⁷, drug discovery¹⁸, material design¹⁹ and particle physics²⁰. At their core, these methods attempt to succinctly describe the dominant factors of variation in a dataset (commonly referred

to as *attributes*), while also capturing invariant properties and inductive biases that maybe essential in downstream tasks.

Computer simulations in HEDP are typically comprised of rich diagnostic data involving heterogeneous measurements and diverse modalities. A crucial step towards enabling scalable analysis of complex phenomena is to concisely approximate the data generation process. Current machine learning solutions generate data representations tailored to specific tasks, with limited versatility, making them difficult to adapt to different use cases. However, by approximating the true generative process with the inferred latent factors, we can produce general-purpose representations that can be adapted for any downstream task, e.g., surrogate modeling or synthetic data generation. Broadly referred to as generative modeling^{21–24}, this ML formulation involves automatic inference of key latent factors of the diagnostic data, and a succinct statistical description of the training data. Such a model can then be leveraged to identify key dependencies and even efficiently sample from the underlying *physics data manifold*. Typically trained in an unsupervised fashion, these generative models can provide latent representations that effectively encode the relevant invariances and capture key relationships between the covariates. In natural language and vision applications, such generative models have made a transformative impact through unprecedented reasoning capabilities and general-purpose representations that are useful across a myriad of downstream tasks. Among the large class of existing generative modeling solutions, *encode-decode*-style architectures are often

¹University of Nevada, Reno, NV, USA. ²Lawrence Livermore National Laboratory, Livermore, CA, USA. ³Arizona State University, Tempe, USA.

✉ e-mail: ankitas@unr.edu

preferred in science applications for their flexibility in incorporating latent priors (e.g., statistical priors) and in producing compact representations for even multimodal data. Furthermore, unlike other existing representation learning approaches^{25,26}, having an explicit decoder network helps us implement the surrogate as a mapping into the latent space from the input (*independent or control*) parameters. In particular, (multi-modal) variational autoencoders (VAE)²¹ and Wasserstein autoencoders (WAE)²⁷ have become popular in the domain of HEDP²⁸⁻³⁰. Despite their empirical success, we argue that existing efforts fall short in terms of validating the appropriateness of the chosen priors, reliably benchmarking the quality of the generative model, or even establishing the universality of the learned representations through suitable generalization studies. We believe conducting systematic efforts to address these critical gaps will not only lead to improved modeling solutions but also promote widespread adoption of these models across the community.

Motivated by the increased interest among HEDP practitioners to emulate the success of generative models from vision and NLP domains, we focus on the timely problem of designing and reliably evaluating HEDP generative models in this paper. In this regard, this work makes the following contributions:

First, we introduce HEDP-Gen, a comprehensive evaluation framework for HEDP generative models (see Fig. 1). Our framework builds upon multiple existing HEDP benchmarks from the literature (JAG analytics simulator³¹, PROBIES diagnostic data³², and a synthetic HYDRA dataset³⁰) and incorporates a wide array of metrics that enable a holistic assessment of these models. HEDP-Gen supports the following evaluations:

- Evaluates the fidelity and validity of the generated samples through three different criteria: *reconstruction fidelity*, *scientific validity* (based on known covariate relationships) and *sample diversity*.
- Assesses the effectiveness of the latent factors in the inferred representations by evaluating the *scientific validity* on samples synthesized via interpolation along a geodesic, and by quantifying the *ease of traversal* along a desired physical attribute.
- Measures the universality of the inferred representations through generalization experiments on the synthetic benchmark from³⁰, where the goal is to generalize across different sub-populations of data arising from changes to the *preheat* parameter.

Second, to demonstrate the utility of HEDP-Gen in rapid development of new solutions, we present a case study. Moving away from existing solutions²⁸ that extensively rely on statistical priors, we use HEDP-Gen to explore the utility of geometric priors in the design of HEDP generative models. Our approach shapes the latent space within an auto-encoding architecture using these priors, enabling meaningful navigation and manipulation metrics. Specifically, we employ hyperbolic geometry, such as the Poincaré ball prior³³⁻³⁵, implementing a geometry-guided variant (Geom-WAE) based on the WAE framework in HEDP-Gen. This Geom-WAE configuration can flexibly integrate various geometric priors, including the Euclidean Gaussian prior, applying geometric priors as a weak inductive bias suitable even if data only partially align with the chosen prior. As mentioned, our implementation and training methodology can be flexibly extended to other geometric priors as well (e.g., Euclidean, hyper-sphere, mixed curvature).

Method

In this section, we describe the proposed framework for design and evaluation of HEDP generative models. Note that we opt for an auto-encoding backbone, similar to state-of-the-art approaches²⁸⁻³⁰ and provide support for architecture design, training and rigorous evaluation of the generative models.

Data layer

To test our analysis we consider datasets spanning two disparate subfields of HEDP: inertial confinement fusion (ICF)^{2,4} and target normal sheath acceleration (TNSA)^{36,37}.

ICF: It utilizes powerful lasers to heat and compress a millimeter-scale capsule filled with deuterium-tritium (DT) fuel. The DT fuel is ignited by the intense pressures and temperatures in the central hotspot of the capsule, creating the conditions necessary for a self-sustained fusion reaction. The fusion reaction generates neutrons and X-rays, which can be imaged to help characterize the quality of the capsule implosion and subsequent hotspot formation. Scalar properties ranging from temperature to hotspot areal density are also measured by various available diagnostics. Given that real experiments are known to be expensive and hence highly sparse, ICF applications are predominantly driven using simulation data with synthetic

HEDP-Gen

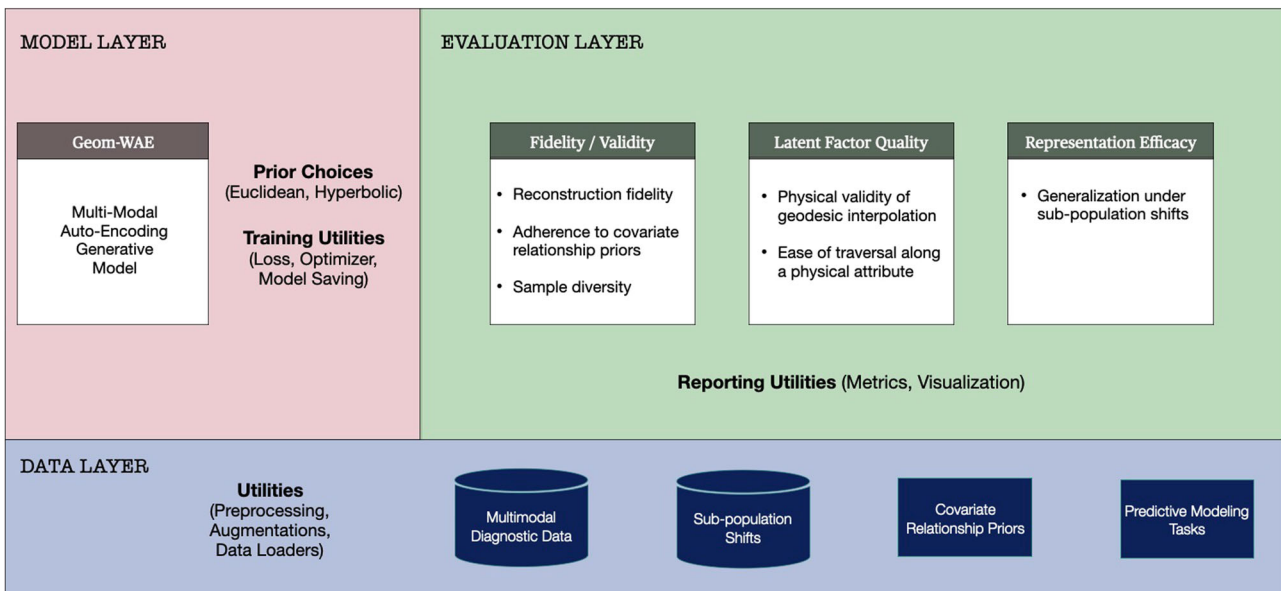


Fig. 1 | An overview of HEDP-Gen framework. Given the inherent challenges in designing and evaluating generative models for diagnostic data from high energy density physics, we propose a general framework. HEDP-Gen is comprised of *Model*, *Data*, and *Evaluation* layers that support flexible incorporation of different prior

choices, data loading and curation utilities and assessment mechanisms. The overall architecture of HEDP-Gen aims to enable rapid development and testing of HEDP generative models.

diagnostics ranging from spectrally resolved (i.e. hypercolor) images generated via X-ray and neutron cameras to a variety of scalars arising from a host of spectrometers and radiochemical diagnostics.

We use two ICF datasets. The first comes from the 1D semi-analytic JAG model³¹, which solves the capsule implosion dynamics and provides a multi-modal set of synthetic diagnostic data. The second uses HYDRA³⁸ to simulate a large ensemble of capsule implosions, capturing small perturbations around an experiment performed at the National Ignition Facility. We refer to these as the JAG and HYDRA-based synthetic benchmark, details for the same are given below:

- (a) JAG benchmark³¹: This comprises 100K samples from an 1D semi-analytic simulation code. The simulation result for each configuration in a 5-dimensional design space corresponds to diagnostic X-ray images, sized 64×64 each, from 4 different, pre-defined energy levels. In order to train our generative models, we view the 4 images at different energies as different channels of an image tensor with size $64 \times 64 \times 4$. In addition to the images, the simulation output also contains 15 summary diagnostics from the final state of the simulation, including yield, ion temperature, pressure, and other physical quantities. Our goal here is to obtain a concise low-dimensional vector representation for the multimodal simulation outputs;
- (b) HYDRA-based synthetic benchmark³⁰: This dataset contains 9 input parameters and a set of outputs comprising X-ray images of size 60×60 and a vector containing 10 scalar quantities describing the behavior of a target capsule under compression. Six of the input parameters, namely energy, power and the geometrical asymmetry parameterized in terms of two spherical-harmonic modes at two different times, represent how the laser energy compresses the target capsule. The three remaining inputs are related to the hydrodynamic scaling, additional energy deposited in the fuel known as the preheat, and the fraction of the dopant in the outer portion of the capsule. The

outputs include peak times for neutron and X-ray emission, temperature, velocity, X-ray emission intensity, neutron yield, which is a key metric for ICF performance. The dataset variants comprises three variants, each with fixed “preheat” and “asymmetry” parameters at settings (5, 0), (20, -0.05), and (40, -0.05), denoted as PH05, PH20, and PH40 respectively while other parameters are randomly sampled. By viewing the design space perturbation as distribution shifts, we expect these datasets to represent challenging scenarios for the generalization of predictive models. Note the task we employ for this study is that we attempt to recover the input parameters given the output images/scalars using a deep-learning model.

TNSA: It uses high-intensity lasers interacting with thin (micron-scale) foils to produce a beam of highly energetic protons. Of particular interest is the energy spectrum and collimation of the beam. These can be assessed with the help of specialized diagnostics, including the recently-developed PROBIES diagnostic³². Experiments involving TNSA can be performed at a high repetition rate, which suggests the possibility of using ML for self-driving experimental platforms. Simulations in this area can help benchmark and pretrain ML models prior to deployment.

Our TNSA dataset consists of synthetic PROBIES data from an ensemble of TNSA simulations. PROBIES is an image-based diagnostic, so the dataset consists entirely of images (with associated simulation inputs). The images have a speckled structure due to the nature of the PROBIES diagnostic: the diagnostic consists of small square regions, each subdivided into a 3×3 grid of filters with varying thicknesses (see Fig. 2 of ref. 32). Such filters allow the diagnostic to extract localized information about the proton energy based on the image brightness behind each filter. We henceforth refer to this dataset as the PROBIES benchmark.

PROBIES benchmark³²: This benchmark contains 70K samples, wherein each simulation run corresponds to a configuration in the 6D

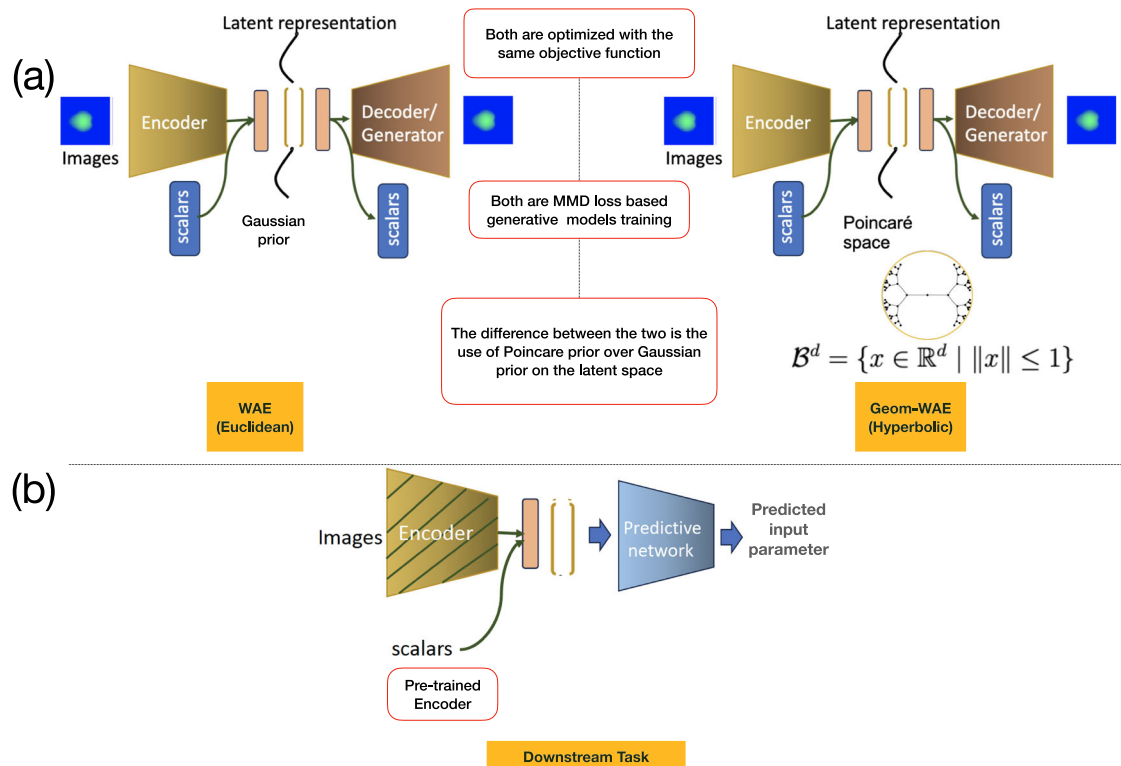


Fig. 2 | Network design in HEDP-Gen. We study the utility of hierarchical latent spaces in scientific generative models. To this end, we propose Geometry guided variant of Wasserstein auto-encoder (Geom-WAE) to enable flexible incorporation of the Poincaré prior in the latent space as shown in (a). The scientific benchmarks considered here are simulation data from HEDP codes and contain two modalities

(images and a set of scalars), which are encoded using separate heads before fusing into a common latent space. The trained encoder that maps input images to the geometric prior latent space can then be used for several downstream tasks as shown in (b).

parameter space (maximum proton energy, divergence parameter of the beam, total energy, laser intensity, laser pulse length and target thickness) and produces diagnostic images of size 300×300 each. In order to train our generative model, we view the diagnostic images as a 1-channel image tensor of size $300 \times 300 \times 1$.

HEDP-Gen implements data loading and processing utilities for different diagnostic benchmarks that represent a variety of scenarios commonly encountered in practice. The *JAG benchmark*³¹ contains multi-modal diagnostics and pre-specified covariate relationships. The *Hydra-based synthetic benchmark*³⁰ separates HYDRA capsule simulations into three distinct regions, each with its own unique physics, which we propose to leverage for evaluating the universality of the learned representation from one regime to another. Finally, the *PROBIES benchmark*³² provides a large-scale diagnostic benchmark for high repetition rate HED experiments. Note that our framework can readily accommodate additional datasets by repurposing the data loading and processing utilities. We now provide details of the three benchmarks included in this study.

Model layer

This core component of HEDP-Gen provides a generic implementation of a Wasserstein auto-encoding generative model, Geom-WAE, with support for extension with additional design choices and training utilities. We use WAE²⁷ as our base model as it is shown to be more flexible than the variational autoencoder (VAE)²¹ in terms of specifying the underlying prior. The WAE is a likelihood-based approach that minimizes the optimal transport (OT) cost for a given objective function $c: W_c(P_X, P_G)$ between the true *unknown* distribution P_X , and the distribution estimated by the latent variable model P_G . This is typically specified using a prior distribution $P(Z)$ in the latent space \mathcal{Z} . The training objective for WAE is similar to variational autoencoders²¹ in that the model optimizes for two distinct costs: (a) reconstruction cost W_G ; (b) a prior-enforcing regularizer that ensures the distribution of encoded samples matches the prior, P_Z . This can be implemented using an adversarial cost²² or distribution comparison metrics such as maximum mean discrepancy (MMD).

Without loss of generality, we assume that a typical diagnostic dataset comprises multiple modalities (e.g., X-ray images, collections of scalar quantities, spectra, etc.), and we implement encoder and decoder architectures to handle all these modalities jointly. Figure 2 (left) illustrates an example implementation of Geom-WAE with a single diagnostic image and a collection of scalar quantities, wherein the encoder ingests both modalities to produce a latent representation and the decoder correspondingly recovers both of them from a latent vector. Formally, we denote the encoder, parameterized by its weights θ , as $z = E_\theta(X_I, X_S)$ that learns to map the images (X_I) and scalars (X_S) into a low-dimensional representation $z \in \mathbb{R}^{z_{\text{dim}}}$, and similarly the decoder, with its weights ϕ , as $\hat{X}_I, \hat{X}_S = D_\phi(z)$ that maps these latent representations to the observation space of images and scalars. Here, \hat{X}_I and \hat{X}_S refers to the recovered scalar and image diagnostics. While the network architecture uses a combination of convolutional layers and fully connected layers, the optimization is carried out as follows:

$$\min_{\theta, \phi} \mathcal{L}(\theta, \phi) = \min_{\theta, \phi} \mathcal{L}_I + \lambda_s \mathcal{L}_S + \lambda_m \mathcal{L}_{\text{MMD}}. \quad (1)$$

Here, $\mathcal{L}_I = \|X_I - \hat{X}_I\|_2^2$ and $\mathcal{L}_S = \|X_S - \hat{X}_S\|_2^2$ use the ℓ_2 metric to evaluate the quality of reconstruction, and \mathcal{L}_{MMD} is implemented as the Maximum Mean Discrepancy (MMD) loss. In particular, we use the kernel version of MMD based on the Inverse Multiquadratics Kernel (IMK) given by $k(x, y) = \frac{C}{C + \|x - y\|_2^2}$, instead of the commonly used RBF kernel²⁷. Here λ_s controls the weighting given for recovering the scalars vs images in the reconstruction cost and λ_m expresses our belief in the prior, where a higher weight results in a stricter requirement for the chosen prior (e.g., Gaussian prior in the Euclidean space). In practice, all aspects of the model including the number of layers, weight initialization, optimizer, and normalization can be readily modified.

Note that, the implementation in Fig. 2 (left) assumes the latent space to be Euclidean (i.e., $\mathbb{R}^{z_{\text{dim}}}$). However, in this work, we hypothesize that the geometry of the latent space plays a central role in determining the quality of HEDP generative models. Consequently, we provide support for incorporating other geometric priors in the latent space, instead of the standard Euclidean assumption, which can hopefully lead to improved organization and manipulation of latent factors. Interestingly, the model layer is designed so that the same optimization in eqn. (1) can be used regardless of the choice of latent space geometry. We present a case study with hyperbolic geometry to illustrate the flexibility of Geom-WAE framework. An overview of the setup is shown in Fig. 2a.

Evaluation layer

While a variety of modeling choices can be adopted for building generative models with diagnostic data, the challenge of systematically evaluating scientific generative models persists. As we will discuss later, standard evaluation metrics such as reconstruction fidelity on held-out data tend to inflate the efficacy of these models and makes it challenging to choose the right model in practice. Hence, HEDP-Gen implements the suite of metrics and evaluation protocols for holistically evaluating HEDP generative models: (i) ability to approximate multimodal data distributions, while also preserving known covariate relationships; (ii) flexibility to recover key attributes of interest in the latent space; and most importantly, (iii) utility in designing downstream models that can effectively handle challenging distribution (or sub-population) shifts (e.g., changes in the underlying physics).

Evaluation 1: Fidelity and Validity of ICF Generative Models. Building accurate generative processes for multimodal simulation data can lead to significantly improved predictive models and experiment designs in ICF applications. Consequently, our first set of evaluations focus on understanding the quality of the samples produced, and measuring how well the synthesized distribution matches the true data distribution. We describe the metrics that we implement for this evaluation using the JAG and PROBIES benchmarks processed in our *Data Layer*.

We note that, in scientific applications, it is not straightforward to validate if the representations from a machine learned model encodes all attributes and correlations induced by the underlying physical process – or in other words – if the encoded latent space is close to the true “state space” of the physical system. Under specific assumptions, there exist theoretical guarantees that one can recover latent factors that accurately invert the true generative process³⁹. However, with the inherent limitations in data acquisition, sampling, and resolution, we can only approximate the recovery of the factors. However, even in the best case scenario where we have accurately recovered them, it is unclear how to validate. Widely adopted metrics from the vision literature, such as the FID score (Fréchet Inception Distance)⁴⁰, are based on pre-trained object classifiers and hence not applicable for evaluating scientific generative models. Finally, the trade-off between scientific validity and diversity in the samples synthesized by generative models is directly linked to the performance on the latent representations in downstream tasks such as surrogate modeling and experiment design. Consequently, we utilize the following metrics for a comprehensive evaluation of HEDP generative models:

(1.a) *Reconstruction fidelity.* For both JAG and PROBIES benchmarks, we measure the fidelity of reconstructions from the decoder through cross-validation, i.e., using a held-out test dataset (not used for training or hyperparameter tuning) with the trained encoder-decoder models. Note, we used the widely adopted mean squared error (MSE) and R-squared statistic (R^2) metrics, independently for images and the diagnostic scalars.

(1.b) *Scientific validity.* When available, known scientific constraints or correlations can be used to validate scientific generative models. For the case of 1D semi-analytic simulator JAG, we leverage the thermal equilibrium to compare the ion temperatures (one of the predicted scalar quantities) and estimates of electron temperature formed by ratios of X-ray image brightness values. Ideally, one expects a strong linear relationship between the predicted ion temperature, and the mean brightness statistic from the X-ray

images. By directly measuring how well the predictions conform to this expectation, we provide a metric for scientific validity.

(1.c) *Sample Diversity*. While sample diversity can be quantified using a number of different metrics, it is critical to ensure that the generative model effectively covers the underlying ICF data manifold. However, it is important to note that, achieving coverage at the cost of producing scientifically invalid samples is undesirable. Hence, we extend the scientific validity evaluation for JAG dataset to quantify the trade-off between validity and diversity of the synthesized samples:

(1.c.1) **Validity score**: Ratio of the number of samples that satisfy the scientific constraint to the total number of samples (higher is better). We consider a sample to be scientifically valid if the deviation of its ion temperature vs electron temperature correlation is less than 0.001 from the expected linear relationship;

(1.c.2) **Coverage**: Ratio of the difference between maximum and minimum values (of ion temperature and electron temperature) predicted from generated samples to the difference between maximum and minimum values observed in the training samples. An illustration for validity and coverage score is shown in the results.

Evaluation 2: Attribute Recovery in the Inferred Latent Spaces. An important characteristic of good generative models is that they have the ability to effectively encode key factors of the underlying physical process in the inferred latent space. This naturally enables their utility as a strong prior for common tasks such as design optimization. Even when sufficient training data is available, the choice of model architecture, training strategy or even optimization can play a crucial role in controlling how the latent factors are organized. Hence, we propose two evaluation strategies to better understand how well the inferred latent spaces adhere to the physical process.

(2.a) *Scientific validity of geodesic interpolation*. A desirable property of scientific generative models is that one should be able to meaningfully interpolate between samples in the latent space, without violating the known constraints or correlations. For this evaluation, we consider the JAG benchmark and investigate if the samples along the path of geodesic interpolation of any training data points also adhere to the known constraint. We sample points uniformly distributed between two training points in the latent space by linearly interpolating between them. handles points falling outside the Poincaréball by projecting them onto its boundary, ensuring the validity of the constraint.

(2.b) *Ease of traversal along a desired attribute*. In addition to investigating samples along the geodesics, we also evaluate the ease of traversing along a desired target attribute in the latent space. To this end, we adopt a simple sequential optimization strategy to traverse the inferred manifold, with the goal of maximizing (or minimizing) any attribute. When the latent factors are organized effectively (i.e., sufficient disentanglement), we expect

to achieve the maximum value within the least number of steps. We choose the JAG benchmark for this evaluation and use image brightness, computed as the image integral, as the target attribute. For the sequential optimization, we start with an initial random sample of size 5 and employ a standard Bayesian optimization loop for 25 steps (with 1 sample chosen in each round) with the goal of maximizing the image integral. Note, we used 5000 candidates in each round with 15 restarts and adopted expected improvement as the acquisition function. More details about the optimization problem is present under Supplementary Note 4.

Evaluation 3: Generalizability of Predictive Models based on Learned Representations. The most common use of ICF generative models is to leverage them as pre-trained representation learners for downstream predictive modeling tasks. Hence, we systematically evaluate the generalizability of predictive models built upon the pre-trained representations, particularly under distribution shifts.

(3.a) **Generalization Fidelity**: We use the Hydra-based synthetic benchmark from³⁰ and measure the cross-generalization across different sub-populations of data arising from changes to the *preheat* parameter. As described earlier, in this evaluation, we consider the problem of predicting the input configuration (5 out of the 9 parameters) given the simulation outputs, i.e., inverse modeling. More specifically, we design a predictive model using data from PH5 and PH20 sub-populations, while the evaluation is performed on the unobserved PH40 subset. While the parameter ranges for all 5 variables are the same across different synthetic data subsets, the simulation outputs can vary, thus requiring the predictive model to handle unknown distribution shifts. Since the predictive model is designed using the pre-trained representations, the predictor f_{θ} is implemented a simple, fully connected network. The metric that we use to characterize the generalization performance is the standard $R - squared$ statistic on samples from the PH40 subset.

Results

Case study: using HEDP-Gen to study the utility of hyperbolic priors

In this section, we utilize the proposed HEDP-Gen framework for studying the performance of DGMs when using stronger inductive biases for the latent space. In particular, we leverage Poincaré (i.e., hyperbolic) geometry for shaping the latent spaces of HEDP generative models. Typically, most existing solutions assume the latent space as Euclidean and adopt a uniform or Gaussian distribution for sampling. In recent years, significant advances have been made towards relaxing the Euclidean assumption for learning models, e.g., graph neural networks⁴¹. Other examples include processing spherical images⁴² or networks that can exploit group-equivariance as an inherent property⁴³. While some of these ideas have been extended to generative models as well⁴⁴⁻⁴⁶, they are known to require specialized layers to

Table 1 | Details of the network architectures used to design generative models for the two datasets considered in our study

JAG dataset			
Encoder:	Conv(4, 4, 4, 64) + BatchNorm2d + ReLU	Decoder:	FC(z_{dim} , 1024) + ReLU
	Conv(4, 4, 64, 32) + BatchNorm2d + ReLU		FC(1024, number of scalar) → scalars
	Conv(4, 4, 32, 16) + BatchNorm2d + ReLU		DConv(4, 16, 32) + BatchNorm2d + ReLU
	FC(1024 + number of scalars, z_{dim})		DConv(4, 4, 32, 16) + BatchNorm2d + ReLU
			DConv(4, 4, 16, 4) → Images
PROBIES dataset			
Encoder:	Conv(4, 4, 1, 256) + BatchNorm2d + ReLU	Decoder:	FC(z_{dim} , 1296) + ReLU
	Conv(4, 4, 256, 128) + BatchNorm2d + ReLU		FC(1024, number of scalar) → scalars
	Conv(4, 4, 128, 64) + BatchNorm2d + ReLU		DConv(4, 4, 16, 32) + BatchNorm2d + ReLU
	Conv(4, 4, 64, 32) + BatchNorm2d + ReLU		DConv(4, 4, 32, 64) + BatchNorm2d + ReLU
	Conv(4, 4, 32, 16) + BatchNorm2d + ReLU		DConv(4, 4, 64, 128) + BatchNorm2d + ReLU
	FC(1296 + number of scalars, z_{dim})		DConv(4, 4, 128, 256) + BatchNorm2d + ReLU
			DConv(4, 4, 256, 1) → Images

enforce the desired geometry, followed by sampling from an appropriately defined prior during inference. We aim to provide a simple training algorithm that builds upon the standard WAE backbone (i.e., inherits the architecture, optimization and training strategy) and incorporates the geometric prior only as a weak inductive bias.

Formulation. Our implementation of Geom-WAE with a Poincaré prior is illustrated in Fig. 2a (right). Let us denote the encoder, parameterized by its weights, θ as $z = E_{\theta}(X_I, X_S)$ that learns to map the images (X_I) and scalars (X_S) into a low-dimensional Poincaréball $z \in \mathbb{B}^n$, and similarly the decoder with its weights, ϕ as $\hat{X}_I, \hat{X}_S = D_{\phi}(z)$ that maps these latent representations to the observation space of images and scalars. We emphasize that Geom-WAE differs from the standard WAE model only by the choice of the prior in the latent space. Hence, we use the general objective in eqn. (1) for training a Geom-WAE, regardless of the

geometric prior. The Poincaréball is a model for d -dimensional hyperbolic geometry, wherein all points are embedded in an d -dimensional open ball of radius $\frac{1}{\sqrt{c}}$, $c > 0$. The ball is expressed as $\mathcal{B}^d = \{x \in \mathbb{R}^d \mid \|x\| \leq 1\}$. In other words, the manifold is defined by the region inside a unit hypersphere, such that the distances grow exponentially as they approach the center of the shell. More formally, the distance between two points in the Poincaréball is given by³⁴:

$$d(\mathbf{u}, \mathbf{v}) = \operatorname{arccosh} \left(1 + 2 \frac{\|\mathbf{u} - \mathbf{v}\|^2}{(1 - \|\mathbf{u}\|^2)(1 - \|\mathbf{v}\|^2)} \right) \quad (2)$$

Training and inference. During training, we optimize for the objective outlined in eqn. (1), which contains the MMD term that requires us to sample from the Poincaréball. To do this, we first sample from a uniform distribution followed by projecting these samples onto the Poincaréball

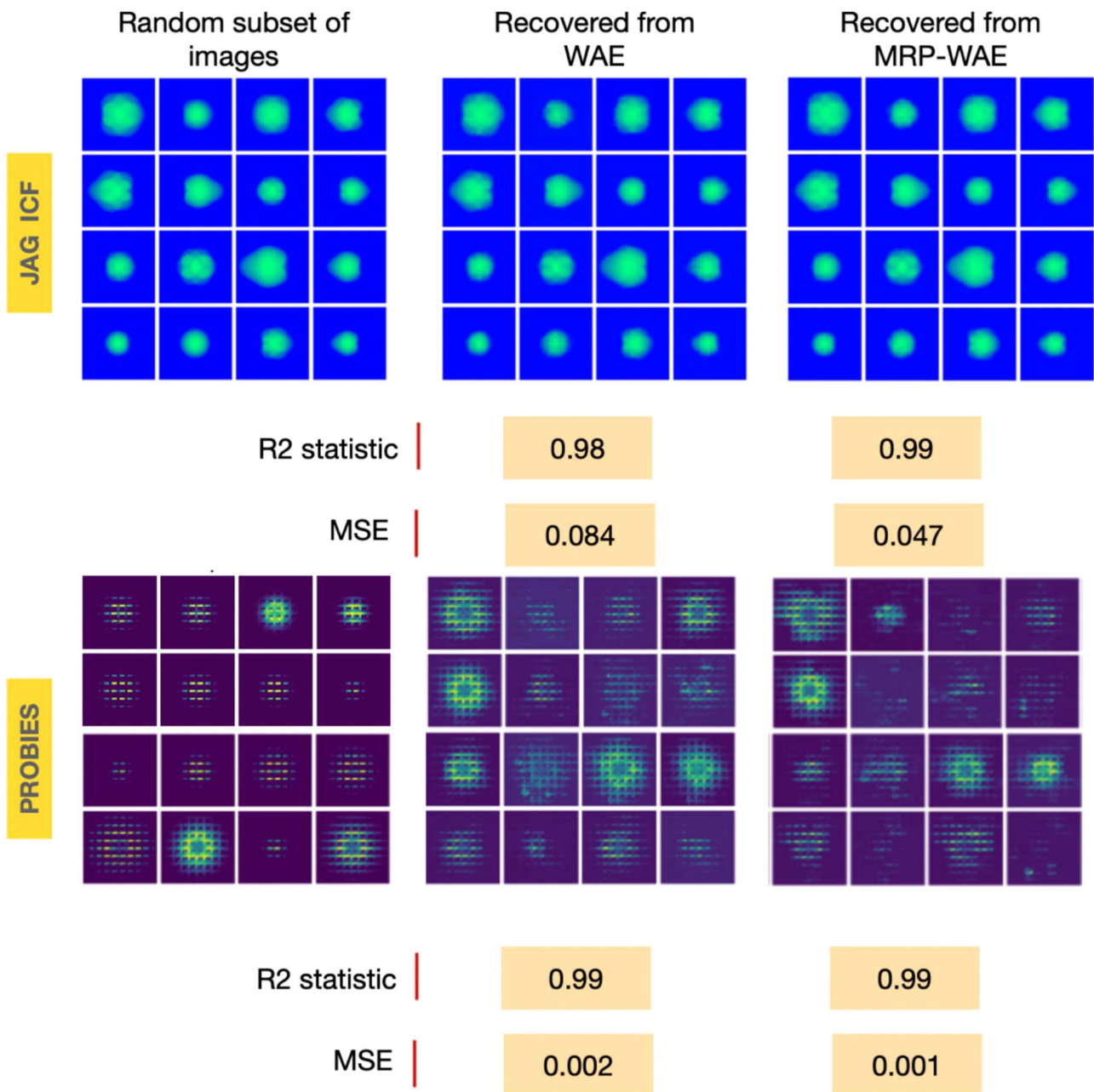


Fig. 3 | Reconstruction fidelity. Effectively embedding and recovering unseen test samples is the most commonly used evaluation protocol for generative models in sciences. Here, we show sample (held-out) images from JAG and HRRL PROBIES

datasets, and their corresponding reconstructions obtained using a Geom-WAE (Euclidean) and Geom-WAE (Poincaré). We also report the R^2 and MSE scores from HEDP-Gen.

using a stereoscopic projection. Sampling from hypersquare is done uniformly from a region defined by the bound parameter. A point p on the Poincaréball is obtained by transforming the point x as $p = \frac{x}{1 + \sqrt{1 + |x|^2}}$. The bound parameter effectively determines the “severity” of the prior, in that a larger bound value emphasizes a stronger hierarchy, with fewer points closer to the origin (the root of the tree). An illustration of the same is shown in Supplementary Fig. 2. Note, since the sampling step is outside the computational graph, any strategy can be used in theory to sample from a hyperbolic space. Our approach naturally generalizes to cases where we wish to *stitch* together custom priors such as mixed curvature spaces⁴⁶, and our technique allows these sampling steps to be non-differentiable without affecting the training or inference steps.

Representations for downstream tasks. To produce compact representations of multi-modal data for downstream tasks, we pass the samples through the pre-trained encoder as shown in Fig. 2, and to synthesize new data, one can sample from the Poincaréprior and pass it through the decoder. One of the objectives of this work is to determine if the Poincaréprior helps in mitigating some of the failures that are generally associated with generative models, namely producing samples that violate the scientific manifold or representations failing to generalize under sub-population shifts. To this end, we use the *evaluation layer* from HEDP-Gen to systematically benchmark the different prior choices.

Network design. For our implementation, we used the Geomstats library⁴⁷ to perform sampling from a Poincaréprior. The dimensionality of the latent space z_{dim} needs to be specified by the user. While we

experimented with a variety of choices for z_{dim} , we observed that a 4-dimensional latent space was already sufficient. The bound parameter of the Poincaréball is also a critical hyperparameter and our hyperparameter search revealed that setting the parameter *bound* = 10 can effectively capture the training data hierarchy on the Poincaréball. The hyperparameters λ_s and λ_m were set to $1e3$ and 10 , respectively. The network architectures used for building generative models with the two benchmark datasets in Table 1. For a fair comparison, we adopted the same training protocol for both cases—each model was trained for 400 epochs using the Adam optimizer with $\beta_1 = 0.5$ and $\beta_2 = 0.999$ and learning rate $3e^{-3}$. Note that, all our models were implemented using PyTorch, and the experiments were carried out using 22 GB, Quadro RTX 6000 GPUs. We also set the random seeds for both Torch and NumPy libraries, to establish a reproducible starting point for the training process, ensuring that both networks start from the same initial state, mitigating any potential biases introduced by differing initializations. The details on the network training are provided in Table 1.

Empirical results

Result 1: Fidelity and Validity of ICF Generative Models. To begin with, we compare the reconstruction fidelity (evaluation 1.a) metrics, namely MSE and R^2 scores for held-out diagnostic data between the two models, and report the results in Fig. 3. Our JAG and PROBIES benchmarks results indicate that Geom-WAE with the Poincaréprior provides consistent improvements over using the standard Euclidean space. Note that the reported metrics were aggregated across all image channels in each case. In addition to the summary metrics, we also include examples of the reconstructions from the two models. Despite the apparent improvements in the

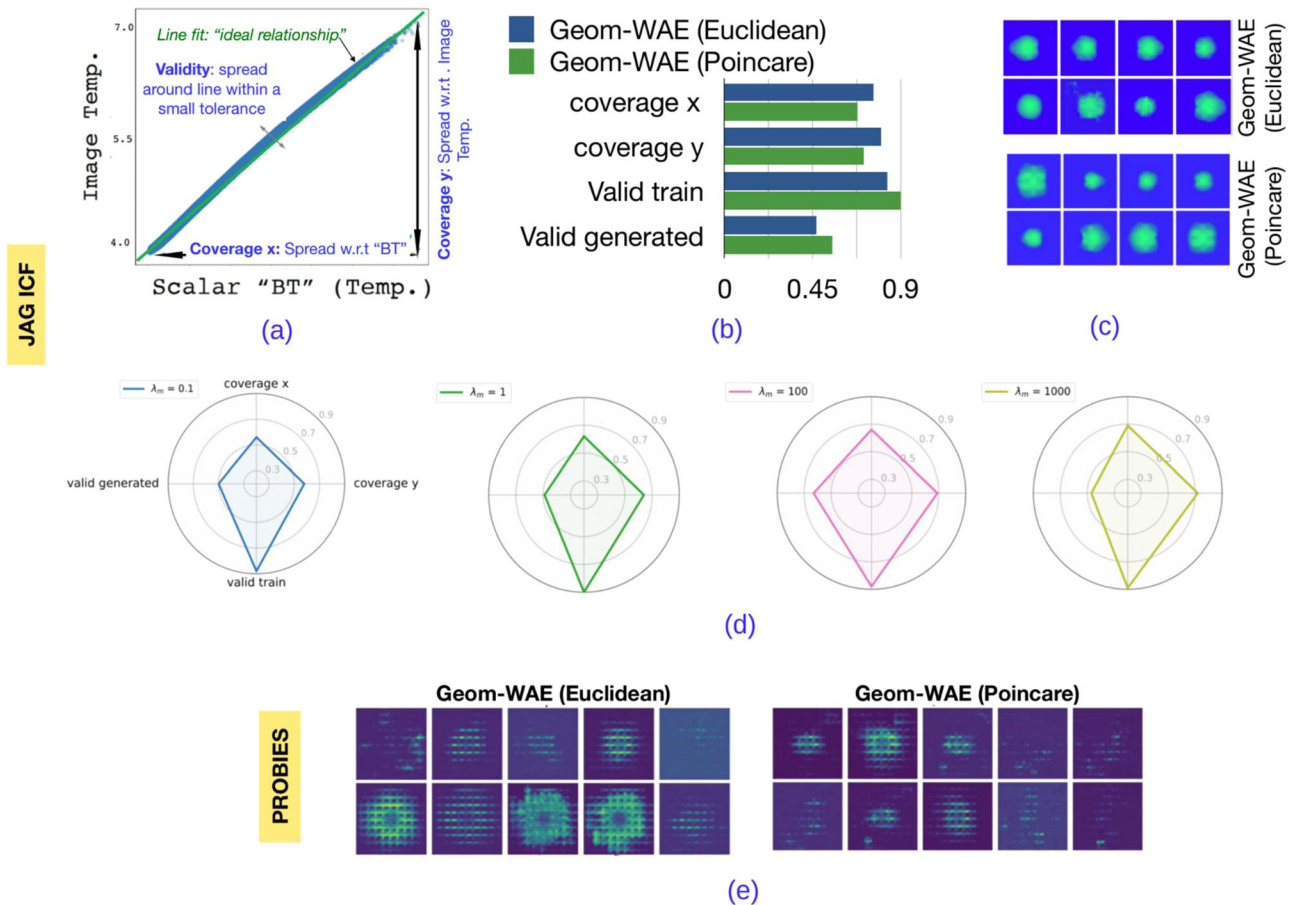


Fig. 4 | Scientific validity and sample diversity. **a** describes the coverage and validity metrics used by HEDP-Gen; **(b)** Comparison of different geometric priors reveals that the Poincaré prior leads to superior generation; **(c)** illustrates a set of realizations generated from the two models; **(d)** shows the effect of the hyper-parameter λ_m in

eqn. (1), and we find that using higher values of λ_m while improving coverage can compromise the scientific validity. Finally, **(e)** illustrates realizations from the two models trained on the PROBIES dataset. Due to the lack of known scientific priors, domain experts typically visually inspect the reconstruction quality.

reconstruction quality, as we will discuss in the rest of this section, such fidelity metrics do not provide a holistic characterization of generative models.

Once the model is trained, points can be sampled from Poincaréball and passed through the decoder to generate images and scalars to evaluate the scientific validity of these generative models. To assess the scientific validity of those samples, HEDP-Gen can verify known covariate relationships when available. In our evaluation, the diagnostic data from the JAG simulator contains a known scientific constraint. However, note that

the generator is not explicitly conditioned to respect the scientific prior during training, and hence, the synthesized samples are not always guaranteed to result in a valid simulation output. Our results show that, despite producing similar reconstruction errors, different Geom-WAE models vary significantly in terms of the scientific validity metric. In Fig. 4a we explain our metric validity score and coverage with illustration, more details for same are discussed under Evaluation 1. We show our results comparing with the standard WAE model in Fig. 4b. When the generated samples have the same range spread as the training data, the

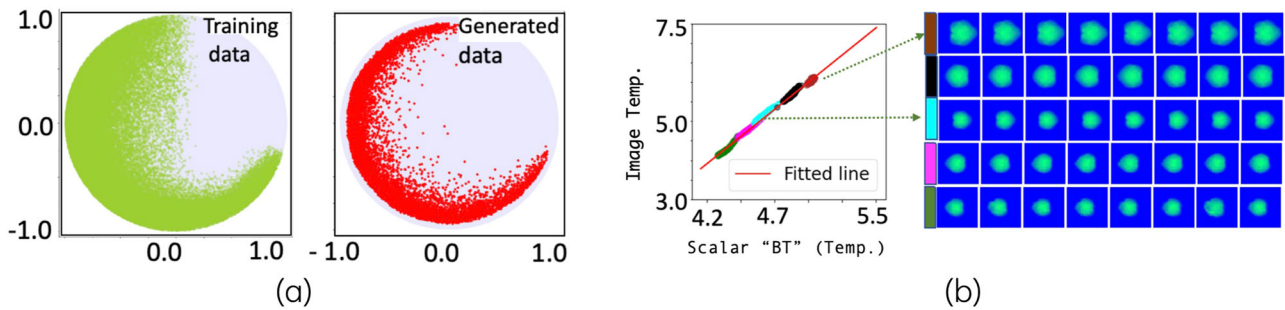


Fig. 5 | Embedding space visualization. **a** shows 2D embeddings of both training and generated data from the JAG benchmark using HoroPCA; **b** shows samples generated by sampling in the latent space in the neighborhood of five different

training samples. The plot illustrates that the resulting generated samples in the neighborhood of the training data do not compromise the scientific validity.

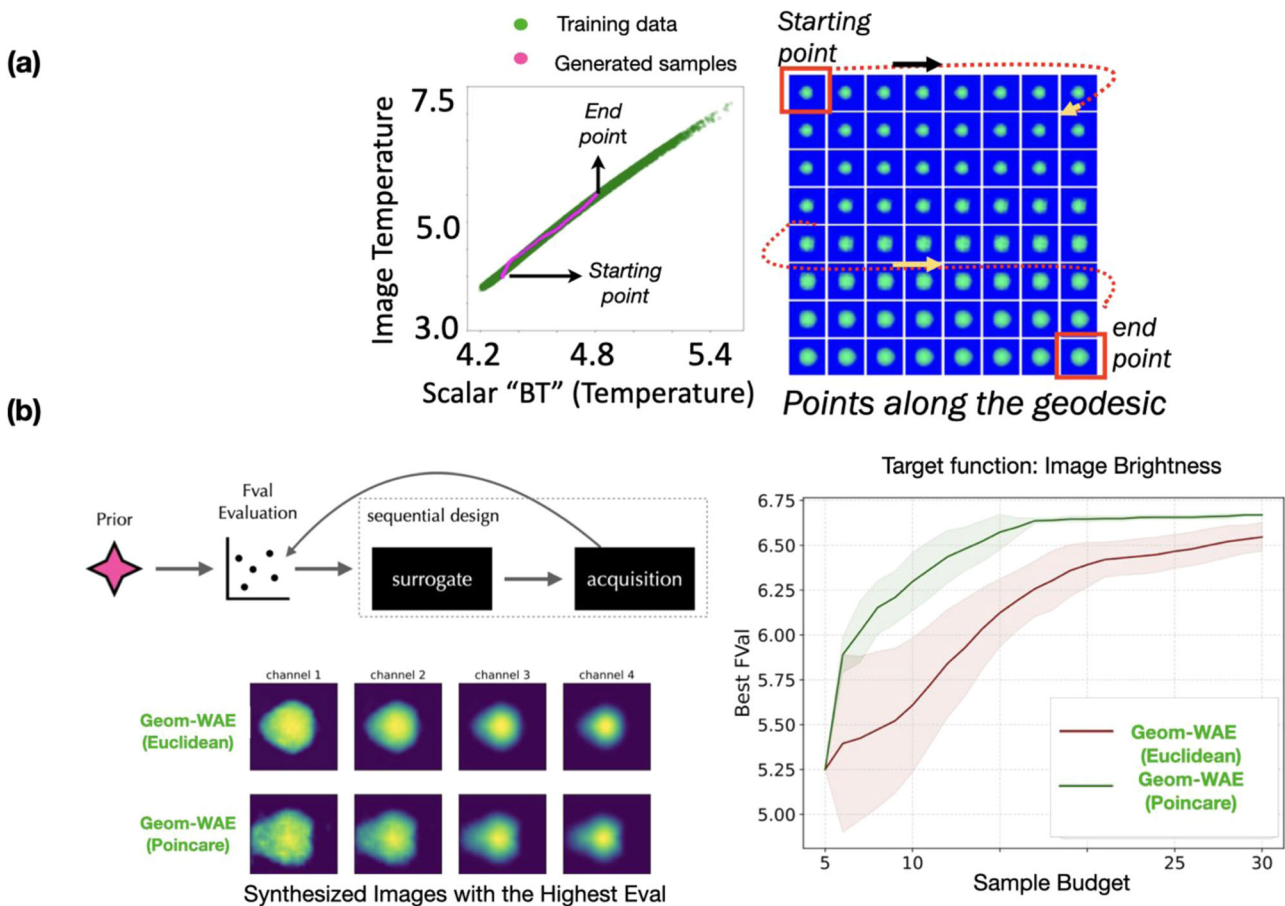


Fig. 6 | Attribute recovery in the inferred latent spaces. **a** Shows geodesic interpolation between two points in the latent space with hyperbolic prior. We show the points generated along this geodesic are scientifically valid (top left), and the corresponding images show high visual fidelity (top right). **b** The schema on top shows the setup of sequential optimization to evaluate the effectiveness of this hyperbolic latent space in optimizing for an attribute of interest. Specifically, we want to find regions in the latent space that produce images with the maximum brightness, as it

captures the amount of energy in the current experiment. This not only measures the effectiveness of structure in the latent space, but also the ability to sample from the Poincaré prior. The sample images in the left panel, show that Geom-WAE (Poincaré) consistently outperforms its Euclidean counterpart in achieving higher peaks for the target function and often does so more rapidly, as illustrated in the right panel. The shaded region represents variations observed across five trials.

value for the coverage metric will be at the maximum value of 1. The coverage metric is mathematically defined as $\frac{v^s_{\max} - v^s_{\min}}{v^t_{\max} - v^t_{\min}}$, here v^s is the value from generated samples and v^t is for training samples. A model that performs well across the four quantities, namely valid train (referring to reconstructed training samples that satisfy the scientific constraint), valid generated (referring to generated samples that satisfy the scientific constraint), coverage x (spread in scalar values) and coverage y (spread in temperature value) would allow generating samples without compromising the reconstruction fidelity. We find that Geom-WAE with a Poincaré is consistently superior in the validity metric across both train and generated samples, thus indicating that it better captures the inherent latent structure of the diagnostic data. Figure 4d illustrates the validity (evaluation 1.b) and coverage (evaluation 1.c) metrics of Geom-WAE (Poincaré) for different values of λ . While increasing the λ_m imposes a stricter prior, we empirically see that $\lambda_m = 10$ gives overall better performance as compared to others. When such scientific priors are not readily available, HEDP-Gen supports a qualitative evaluation by visualizing the generated samples as shown in Fig. 4e. Since domain scientists are required to perform this evaluation, additional sample selection strategies can be used to show only a few, yet diverse, samples for visualization⁴⁸⁻⁵¹.

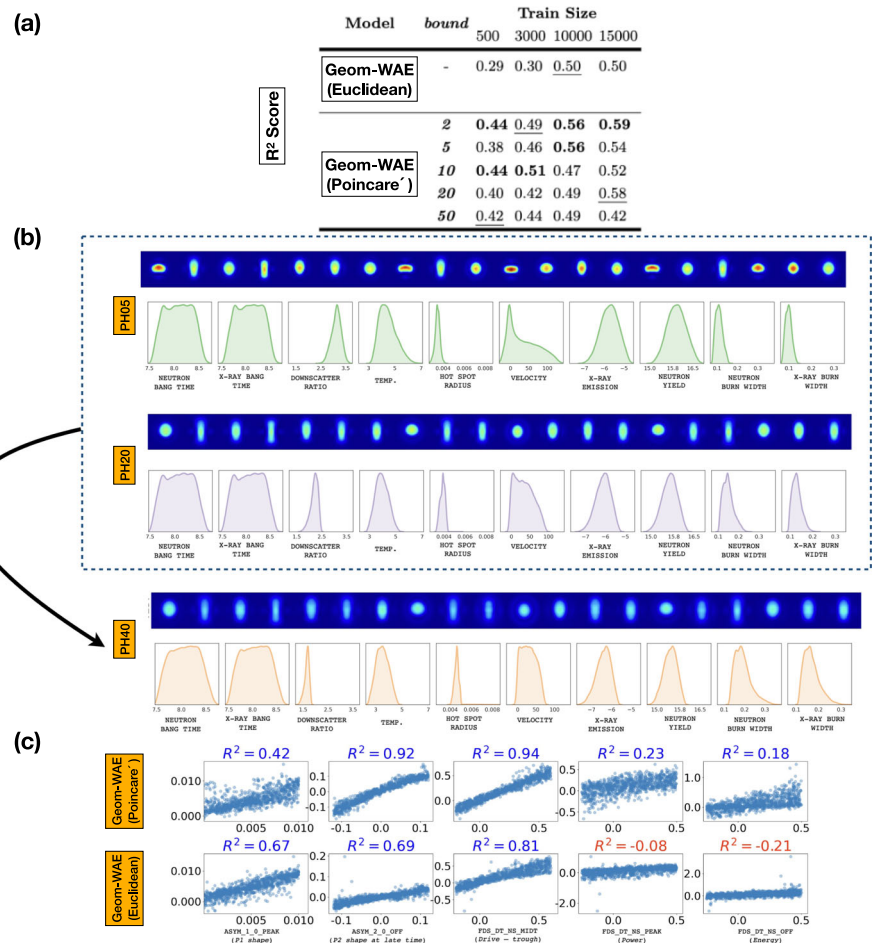
Result 2: Attribute Recovery in the Inferred Latent Spaces. Before we present our results on effectively traversing the inferred latent spaces for recovering physical attributes, we first study the latent organization in a Geom-WAE (Poincaré) model. To this end, we consider the JAG 1D diagnostic data, transform the training samples to their corresponding latent representations, and subsequently embed them into a 2D visualization using HoroPCA⁵². From the illustration in Fig. 5, as expected, we first notice that the majority of the training samples are embedded on the Poincaréball. Interestingly, this latent organization is also reflected with the generated samples that are sampled using a pre-defined prior on the Poincaréball. This

evidences Geom-WAE’s ability to impose hyperbolic geometry during the WAE training. Furthermore, in order to characterize the distribution of training data on the Poincaréball, we sample the neighborhood of each training example (using a standard normal around that example and projecting them onto the Poincaréball). Surprisingly, the neighborhood samples are found to be scientifically valid (i.e., recovers the covariate relationship), as illustrated in Fig. 5b. The visualization of marginal distribution of the training data and generated samples is present in Supplementary Fig. 1.

Next, we use HEDP-Gen to evaluate the scientific validity of samples generated via geodesic interpolation (evaluation 2.a) between latent embeddings of two training samples and report the results in Fig. 6a. Note that this evaluation only used the JAG 1D simulator data and the latent space from our Geom-WAE (Poincaré) model. We observed that the generated samples along the geodesic satisfy the scientific constraint, indicating that the path taken in the latent space is reflective of the underlying data manifold. The figure shows few generated image samples from this path, with the top left showing the starting point and the bottom right the end-point. We noticed that the inferred latent space accurately captures the semantic information in the dataset— e.g., indicated by the direction of the tail in the image.

We also present our results on ease of traversal in the latent space to maximize the image brightness (evaluation 2.b). We find that, when compared to the conventional Euclidean latent spaces, Geom-WAE (Poincaré) consistently identifies candidates with higher temperatures and, more importantly, achieves significantly faster convergence. This behavior can be attributed to both the efficacy of the Poincaréball to better approximate the true data manifold as well as the benefit of Poincaré prior in effectively navigating the latent space. We repeated the experiment for 5 independent trials and reported the convergence curves in Fig. 6(b).

Fig. 7 | Generalizability of predictive models based on learned representations. HEDP-Gen utilizes a supervised, predictive modelling task to assess the utility of the learned representations in downstream applications. **a** lists the R^2 scores for the the surrogate models trained with representations from Geom-WAE (Euclidean) and Geom-WAE (Poincaré) models at varying training sizes; (bold and underline denotes best and second best results) **(b)** shows examples of the distribution shift in the form of scalars and images in PH05 and PH20 respectively. The distribution shift is visually apparent, and our task is to learn representations that can predict accurately when trained on data with preheat parameters PH05 and PH20, while being tested on PH40. In **(c)** we show a sample result in predicting the design parameters using 10,000 labeled training examples.



Result 3: Generalizability of Predictive Models based on Learned Representations. Lastly, we compare the generalization fidelity (evaluation 3.a) metric and R^2 scores for the two models in Fig. 7. Our results in predictive modeling task on PH40 dataset indicate that the hyperbolic realization of Geom-WAE contains domain invariant features as compared to standard model, resulting in better generalization fidelity. The Poincaré realization of Geom-WAE achieves a better score in all test cases in predicting the target domain's design parameter as shown in Fig. 7a. For the Poincaré prior realization of Geom-WAE, we vary along two factors: *bound* that controls type of Poincaré prior enforced in the latent space, and the number of labeled examples available to train the predictor network f_θ . An illustration of predictive network is shown in Fig. 2b. Note that, in all cases we use a latent space of size $d = 8$, and train WAE on 15000 unlabeled images and scalars. Once trained, we use the encoder as a feature extractor and train f_θ using N_{sup} labeled examples.

Discussion

This paper introduced HEDP-Gen, a framework for training and evaluating generative models tailored for HEDP application. To evaluate its effectiveness, we present a study to explore the effect of imposing hyperbolic prior in latent space of generative model. We impose Poincaré prior on the latent space to induce Poincaré geometry as opposed to the standard model that assumes latent space to be Euclidean. The underlying geometry of the Poincaré ball allows us to improve the generation of samples that satisfy the scientific prior, while also maintaining reconstruction fidelity. Our results show that Geom-WAE with Poincaré prior is holistically superior to Geom-WAE with Gaussian prior in Euclidean space, which is essentially the naive WAE model. Our Geom-WAE model with Poincaré prior is able to generate a significantly higher ratio of scientifically valid samples with reduced reconstruction error when compared to its Euclidean counterpart. Furthermore, using latent spaces with non-Euclidean geometry leads to improved generalization of surrogate models, even in low data regimes. We believe this relaxed prior in the latent space was beneficial as the underlying geometry is not known to have a specific geometric structure. Therefore, to address this going forward we want to explore mixed-geometric structures that can go beyond hyperbolic geometry and could benefit the generation of scientific data as well as downstream tasks.

Data availability

The dataset JAG ICF is available at <https://github.com/LLNL/macc>. PROBIES, HYDRA, and generated datasets are available upon reasonable request from the corresponding author.

Code availability

The code is available at <https://github.com/goodailab/journalcode>.

Received: 18 January 2024; Accepted: 11 December 2024;

Published online: 08 January 2025

References

- Drake, R. P. and Drake, R. P. Introduction to high-energy-density physics. in *High-Energy-Density Physics: Foundation of Inertial Fusion and Experimental Astrophysics*, 1–20 (Springer International Publishing, 2018).
- Hurricane, O. et al. Physics principles of inertial confinement fusion and us program overview. *Rev. Mod. Phys.* **95**, 025005 (2023).
- Craxton, R. et al. Direct-drive inertial confinement fusion: a review. *Phys. Plasmas* **22**, 110501 (2015).
- Lindl, J. D. et al. The physics basis for ignition using indirect-drive targets on the national ignition facility. *Phys. Plasmas* **11**, 339 (2004).
- Ishak, B. High-energy-density physics: foundation of inertial fusion and experimental astrophysics. *Contemp. Phys.* **59**, 308 (2018).
- Ryutov, D. et al. Similarity criteria for the laboratory simulation of supernova hydrodynamics. *Astrophys. J.* **518**, 821 (1999).
- Swift, D. C. et al. Mass–radius relationships for exoplanets. *Astrophys. J.* **744**, 59 (2011).
- Ma, Y. et al. Transparent dense sodium. *Nature* **458**, 182 (2009).
- Drozdov, A., Erements, M., Troyan, I., Ksenofontov, V. & Shylin, S. I. Conventional superconductivity at 203 kelvin at high pressures in the sulfur hydride system. *Nature* **525**, 73 (2015).
- Hurricane, O. & Herrmann, M. High-energy-density physics at the national ignition facility. *Annu. Rev. Nucl. Part. Sci.* **67**, 213 (2017).
- Sharkov, B. Y., Hoffmann, D. H., Golubev, A. A. & Zhao, Y. High energy density physics with intense ion beams. *Matter Radiat. Extrem.* **1**, 28 (2016).
- Hatfield, P. W. et al. The data-driven future of high-energy-density physics. *Nature* **593**, 351 (2021).
- Hatfield, P. et al. Using sparse Gaussian processes for predicting robust inertial confinement fusion implosion yields. *IEEE Trans. Plasma Sci.* **48**, 14 (2019).
- Albertsson, K. et al. Machine learning in high energy physics community white paper, in *Journal of Physics: Conference Series*, Vol. 1085, 022008 (IOP Publishing, 2018).
- Kasim, M. F. et al. Building high accuracy emulators for scientific simulations with deep neural architecture search. *Mach. Learn.: Sci. Technol.* **3**, 015013 (2021).
- Yang, C. et al. *Preparing Dense Net for Automated HYDRA Mesh Management via Reinforcement Learning*. (Lawrence Livermore National Laboratory (LLNL), 2019).
- Mustafa, M. et al. Cosmogon: creating high-fidelity weak lensing convergence maps using generative adversarial networks. *Comput. Astrophys. Cosmol.* **6**, 1 (2019).
- Bian, Y. & Xie, X.-Q. Generative chemistry: drug discovery with deep learning generative models. *J. Mol. Model.* **27**, 1 (2021).
- Zhou, T., Song, Z. & Sundmacher, K. Big data creates new opportunities for materials research: a review on methods and applications of machine learning for materials design. *Engineering* **5**, 1017 (2019).
- Paganini, M., de Oliveira, L. & Nachman, B. Calogan: simulating 3d high energy particle showers in multilayer electromagnetic calorimeters with generative adversarial networks. *Phys. Rev. D.* **97**, 014021 (2018).
- Kingma, D. P. & Welling, M. Auto-encoding variational bayes. In *2nd International Conference on Learning Representations, ICLR* (2014).
- Goodfellow, I. et al. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **27** 2672–2680 (2014).
- Croitoru, F.-A., Hondru, V., Ionescu, R. T. & Shah, M. Diffusion models in vision: a survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **45**, 10850–10869 (2023).
- Papamakarios, G., Nalisnick, E., Rezende, D. J., Mohamed, S. & Lakshminarayanan, B. Normalizing flows for probabilistic modeling and inference. *J. Mach. Learn. Res.* **22**, 2617 (2021).
- Chen, T., Kornblith, S., Norouzi, M. & Hinton, G. A simple framework for contrastive learning of visual representations. *Proceedings of the International Conference on Machine Learning 2020*, 1597–1607 (PMLR, 2020).
- Caron, M. et al. Emerging properties in self-supervised vision transformers. In *Proc. IEEE/CVF International Conference on Computer Vision*, 9650–9660 (IEEE, 2021).
- Tolstikhin, I., Bousquet, O., Gelly, S. & Schölkopf, B. Wasserstein auto-encoders. In *6th International Conference on Learning Representations (ICLR)* (2018).
- Anirudh, R., Thiagarajan, J. J., Bremer, P.-T. & Spears, B. K. Improved surrogates in inertial confinement fusion with manifold and cycle consistencies. *Proc. Natl. Acad. Sci. USA* **117**, 9741 (2020).
- Kustowski, B. et al. Transfer learning as a tool for reducing simulation bias: application to inertial confinement fusion. *IEEE Trans. Plasma Sci.* **48**, 46 (2019).

30. Kustowski, B. et al. Suppressing simulation bias in multi-modal data using transfer learning. *Mach. Learn. Sci. Technol.* **3**, 015035 (2022).
31. Gaffney, J. A. et al. The JAG inertial confinement fusion simulation dataset for multi-modal scientific deep learning., In Lawrence Livermore National Laboratory (LLNL) Open Data Initiative. <https://doi.org/10.6075/JORVOM27> (2020).
32. Mariscal, D. et al. A flexible proton beam imaging energy spectrometer (probies) for high repetition rate or single-shot high energy density (hed) experiments. *Rev. Sci. Instrum.* **94**, 023507 (2023).
33. Ratcliffe, J. G., Axler, S. & Ribet, K. *Foundations of hyperbolic manifolds*, Vol. 149 (Springer, 1994).
34. Nickel, M. & Kiela, D. Poincaré embeddings for learning hierarchical representations. *Adv. neural Inf. Process. Syst.* **30**, 6338 (2017).
35. Sarkar, R. Low distortion delaunay embedding of trees in hyperbolic plane. In *Proc. International Symposium on Graph Drawing* 355–366 (Springer, 2011).
36. Snavely, R. et al. Intense high-energy proton beams from petawatt-laser irradiation of solids. *Phys. Rev. Lett.* **85**, 2945 (2000).
37. Wilks, S. et al. Energetic proton generation in ultra-intense laser–solid interactions. *Phys. Plasmas* **8**, 542 (2001).
38. Marinak, M. M. et al. Three-dimensional hydra simulations of national ignition facility targets. *Phys. Plasmas* **8**, 2275 (2001).
39. Kirchhof, M., Kasneci, E. & Oh, S. J. Probabilistic contrastive learning recovers the correct aleatoric uncertainty of ambiguous inputs. In *Proc. International Conference on Machine Learning* 17085–17104 (PMLR, 2023).
40. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B. & Hochreiter, S. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Adv. Neural Inform. Process. Syst.* **30**, 6629–6640 (2017).
41. Defferrard, M., Bresson, X. & Vandergheynst, P. Convolutional neural networks on graphs with fast localized spectral filtering. *Adv. Neural Inf. Process. Syst.* **29**, 3844 (2016).
42. Cohen, T. S., Geiger, M., Köhler, J. & Welling, M. Spherical CNNs. In *6th International Conference on Learning Representations* (2018).
43. Cohen, T. & Welling, M. Group equivariant convolutional networks. In *Proc. 33rd International conference on machine learning*, Vol. 48, 2990–2999 (PMLR, 2016).
44. Xu, J. & Durrett, G. Spherical latent spaces for stable variational autoencoders. In *Proc. Conference on Empirical Methods in Natural Language Processing*, 4503–4513 (Association for Computational Linguistics, 2018).
45. Mathieu, E., Le Lan, C., Maddison, C. J., Tomioka, R. & Teh, Y. W. Continuous hierarchical representations with poincaré variational autoencoders. *Adv. Neural Inf. Process. Syst.* **32**, 12565–12576 (2019).
46. Skopek, O., Ganea, O.-E. & Bécigneul, G. Mixed-curvature variational autoencoders. In *Proc. International Conference on Learning Representations* (2020).
47. Miolane, N. et al. Geomstats: a python package for riemannian geometry in machine learning. *J. Mach. Learn. Res.* **21**, 1 (2020).
48. Razavi, A., Van den Oord, A. & Vinyals, O. Generating diverse high-fidelity images with vq-vae-2. *Adv. Neural Inf. Process. Syst.* **32**, 14866–14876 (2019).
49. Ma, Y. J., Inala, J. P., Jayaraman, D. & Bastani, O. Likelihood-based diverse sampling for trajectory forecasting, in *Proc. IEEE/CVF International Conference on Computer Vision*, 13279–13288 (IEEE, 2021).
50. Kviman, O., Melin, H., Koptagel, H., Elvira, V. & Lagergren, J. Multiple importance sampling elbo and deep ensembles of variational approximations. In *Proc. International Conference on Artificial Intelligence and Statistics*, 10687–10702. (PMLR, 2022).
51. Gao, Z. et al. Mitigating the filter bubble while maintaining relevance: targeted diversification with vae-based recommender systems. In *Proc. 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2524–2531 (Association for Computing Machinery, 2022).
52. Chami, I. et al. Low-dimensional hyperbolic knowledge graph embeddings. In *Proc. 58th Annual Meeting of the Association for Computational Linguistics*, 6901–6914 (Association for Computational Linguistics, 2020).

Acknowledgements

This work was performed under the auspices of the US Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. A.S. and P.T. were supported by B644146 from LLNL to ASU. Document release number LLNL-JRNL-843999-DRAFT.

Author contributions

A.S. conceived the presented idea and, along with R.A., J.J.T., and E.K., developed the overall formulation. Y.M. extended the approach to a use case with the help of R.A. and J.J.T., R.A., E.K., and J.J.T. contributed substantially to technical discussions throughout this work. While B.S. and P.-T.B. and P.T. and T.M. helped in setting the high-level objectives, D.M., B.D., B.K., and K.S. provided insights and expert knowledge to work with different datasets used in this work. All authors contributed extensively to the work presented in this paper and provided expert knowledge. All authors edited the manuscript. While A.S. led the paper-writing efforts, J.J.T., R.A., and E.K. contributed significantly to several sections.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42005-024-01912-2>.

Correspondence and requests for materials should be addressed to Ankita Shukla.

Peer review information *Communications Physics* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025