PHYSERL-INV: A PHYSICS-ENCODED INVERSE MODELING APPROACH FOR ARCTIC SNOW DEPTH PREDICTION

Anonymous authors
Paper under double-blind review

ABSTRACT

The integration of domain knowledge with inverse modeling has emerged as a powerful approach for solving complex physical systems in machine learning. While prior research has explored bijective mapping and complex contrastive learning, the potential of surjective mapping in combination with supervised representation learning remains largely unexplored. To address this gap, we propose Physics-encoded Representation Learning for Inverse modeling (PhysERL-Inv), which encodes the hydrostatic balance equation and uses supervised contrastive learning to predict the time evolution of Arctic snow depth. Evaluated against multiple baseline models, PhysERL-Inv significantly improves prediction performance, reducing error by 20% and demonstrating superior physical consistency. Our approach demonstrates the potential of leveraging surjective mapping to solve complex, ill-posed problems, with wide applicability in data-sparse domains.

1 INTRODUCTION

Understanding and modeling Earth's climate system requires accurate data on the spatial extent and properties of Arctic sea ice. A crucial, yet often unmeasured, property is snow depth. While its influence on sea ice is well-established (Li et al., 2024), reanalysis datasets like ERA5 do not include direct measurements of snow depth over sea ice (Hersbach et al., 2020). To address this data gap, we exploit Physics-encoded Neural Networks (PeNNs) (Faroughi et al., 2022), a class of models widely applied in scientific analysis to implement inverse modeling. In this study, we use sea ice physics derived from the hydrostatic equation (Kwasniok, 2022) to inform our network and predict snow depth.

Inverse modeling is a powerful tool for inferring intrinsic physical parameters and uncovering hidden characteristics of complex physical phenomena. These models extract meaningful insights from systems in various fields, including lake temperature modeling (Tayal et al., 2022), tomography (Bubba et al., 2019), seismic waveform analysis (Sun et al., 2020), materials science (Liao & Li, 2020), and hydrology (Ghorbanidehno et al., 2020). However, a common limitation is that many of these methods rely on physical priors or posterior distributions for parameter estimation (Karumuri & Bilionis, 2024). For example, the accuracy of the predicted variable directly depends on the accuracy of the physical mapping from other variables. This dependency can be problematic because any errors in the initial mapping will propagate through the model during training, potentially affecting the overall performance. While an end-to-end learning approach can offer a streamlined solution, it can also limit the ability to embed additional physical constraints that are relevant to the system (Faroughi et al., 2022). This is because the end-to-end model is highly dependent on its initialization process and the pre-defined forward mapping.

In addition, inverse modeling using complex forward-inverse flow transformations (Tarantola, 2005; Ghosh et al., 2022; Tayal et al., 2022) is a computationally demanding process. These works rely on a bijective mapping between the input and latent spaces. A bijective mapping is a strict, one-to-one correspondence where each input has a unique output and every possible output is reached. While this type of mapping is suitable for systems with static parameters (e.g., unchanging material properties), its strict nature can be restrictive in the context of a dynamic physical system like Arctic sea ice. In such dynamic cases, the underlying parameters are not static but change over time. A bijec-

tive mapping may not be flexible enough to capture these evolving relationships, as it rigidly links a single input state to a single output state, limiting its ability to account for the continuous evolution of the system. Furthermore, the work mentioned above builds upon traditional self-supervised learning methods that effectively learn general data representations from unlabeled data by identifying statistical patterns. However, these learned features, which are not constrained by or encoded with physical principles, might not be directly relevant or physically meaningful for a specific domain. As a result, a generic self-supervised model often requires significant fine-tuning with labeled data to accurately estimate parameters, as it may capture general dynamic patterns without understanding the precise physical relationships that govern the system.

To overcome the limitations of both traditional inverse modeling and conventional self-supervised learning, we introduce a novel, simplified hybrid framework called Physics-Encoded Representation Learning for Inverse modeling (PhysERL-Inv). This framework utilizes a surjective inversion mapping for time-varying parameter estimation, enabling it to exploit all possible values the predicted variable could take. It operates on the principle that the temporal evolution of a system can be effectively modeled using linear dynamics present in its physical variables (Kwasniok, 2022). We, therefore, use a linear physical equation derived from the hydrostatic balance equation (Kwok & Cunningham, 2008) of Arctic sea ice to relate the cause-and-effect phenomena between sea ice and snow variables. By using this mapping, we can detect hidden characteristics of sea ice parameters. We then encode these physical constraints into our neural network to ensure physically consistent predictions of snow depth. Additionally, the model incorporates representation learning through a supervised contrastive learning approach via physics encoding. This compels the model to capture the physical relationships between Arctic snow and sea ice, which are later inverted to estimate the hidden parameters needed to improve the prediction process. The main contributions of this paper are summarized as follows: **(1)** We present PhysERL-Inv, a novel hybrid framework that leverages a surjective inversion mapping integrated into a supervised representation learning process. This approach enables a more direct and efficient optimization for predicting snow depth through inverse modeling. **(2)** Moving beyond traditional inverse modeling, our framework achieves effective physics encoding by embedding constraints from the hydrostatic balance equation of Arctic sea ice thickness directly into the neural network. This ensures physically consistent inferences, leading to more accurate snow depth predictions. **(3)** We demonstrate the effectiveness of our proposed model by comparing it with multiple baselines.

2 Related Works

Machine learning–enhanced inverse modeling offers a powerful way to uncover hidden physical parameters that govern observable geospatial phenomena, many of which cannot be measured directly. This approach has been widely applied in fields such as hydrology (Ghosh et al., 2022), water flow studies (Mo et al., 2020), and lake temperature modeling (Tayal et al., 2022). In seismic waveform inversion, for example, researchers have begun integrating theoretical knowledge of seismic wave propagation into deep learning frameworks (Adler et al., 2021). Deep neural networks are also being employed to tackle electrical impedance tomography (EIT) problems, which involve inverting the highly nonlinear and high-dimensional Dirichlet-to-Neumann (DtN) map (Fan & Ying, 2020). In addition, innovative architectures such as SwitchNet have been developed to address forward and inverse scattering problems for wave equations, offering computational efficiency while capturing the global nature of scattering phenomena (Khoo & Ying, 2021).

The self-supervised paradigm has also proven effective in addressing inverse problems across a wide range of scientific disciplines, extending beyond traditional domains like natural language processing (Fang et al., 2020) and computer vision (Bardes et al., 2022) to include scientific modeling (Scotti et al., 2023) and noninvasive medical digital twins Kuang et al. (2025). By exploiting complex pretext tasks and multi-stage training, these methods can extract meaningful representations directly from underlying physical structures (Bardes et al., 2022), which in turn enables parameter estimation without the need for explicit labels (Liu et al., 2021; Jing & Tian, 2020). A limitation of these methods is their heavy reliance on generative and regularization losses, which necessitate extensive hyperparameter tuning. Moreover, the use of bijective mappings (e.g., (Ghosh et al., 2022)) assumes perfect invertibility, which is not plausible with real-world data.

Physics-encoded Neural Networks (PeNNs) represent another powerful approach for integrating scientific principles into machine learning (Faroughi et al., 2022; Willard et al., 2022; Karpatne et al., 2024). By incorporating physical laws in the form of differential and linear equations (Rao et al., 2021; Chen et al., 2018; Kovachki et al., 2021; Innes et al., 2019), these models not only show improved performance but also adhere more closely to physical laws, transforming traditional black-box algorithms into more interpretable models. Despite these advances, a research gap remains: there is no unified framework that merges the rigor of physics-encoded methods with the efficiency of direct optimization for complex, time-varying inverse problems.

3 METHODOLOGY

3.1 PROBLEM FORMULATION AND PRELIMINARIES

We begin by defining our observational data. For each sample i, we have an input time series $X_i = [x_1^i, x_2^i, ..., x_T^i]$, where x_t^i represents the measurement at time step t, and T denotes the total number of time steps. Correspondingly, we have a target variable $Y_i = [y_1^i, y_2^i, ..., y_T^i]$, where y_t^i is the observed output at time step t. As shown in Figure 1, our model uses an augmented input, denoted as X', which is crucial for the supervised contrastive learning component. The core details of this supervised contrastive learning objective are further discussed in Section 3.3.

We aim to model the relationship between the input $X_i = \rho_s$ and target $Y_i = h_s$, particularly in the context of estimating snow depth by integrating snow and sea ice properties. To achieve this, we adapt the hydrostatic balance equation 1, which, as stated by Kwok & Cunningham (2008), describes the equilibrium where the weight of ice and snow is balanced by the buoyant force of seawater. Using this relation, we generate proxy X_i labels to inject physical attributes into the training process.

$$h_s = \left(\frac{\rho_w}{\rho_s}\right) h_f - \left(\frac{\rho_w - \rho_i}{\rho_s}\right) h_i \tag{1}$$

Where h_s is the snow depth, h_i is the sea ice thickness, and h_f is the ice freeboard height. The parameters ρ_w, ρ_i, and ρ_s represent the densities of seawater, sea ice, and snow, respectively. This equation provides a fundamental relationship governing the interaction between sea ice, overlying snow, and the buoyant force exerted by seawater.

3.2 MODEL DESIGN

Our PhysECL-Inv framework (Figure 1) is a physics-encoded autoencoder that combines three key components: an LSTM, a multi-head attention mechanism, and a multilayer perceptron (MLP). The LSTM processes the temporal sequences, capturing long-term dependencies within the time series data. Given the size of our dataset and the need to capture long-term climate patterns, the LSTM provided a more efficient and stable solution compared to the more computationally intensive Transformer model. The multi-head attention mechanism then weighs the importance of different temporal features, which is particularly crucial for identifying non-local relationships in climate data. A key concept in climate science is that events far away can influence a local system, and this mechanism helps the model focus on those relevant, non-local connections. Finally, the MLP is used to infer the underlying physical parameters by encoding physical constraints, enabling the model to learn complex, physically consistent representations.

Encoder LSTM: The encoder receives the input time-series sequence $X = [x_1, x_2, \ldots, x_T]$, where x_t represents the input at time step t, and T is the length of the input sequence. The encoder is an LSTM network, it is particularly suited for tasks where long-range temporal dependencies between events exist. In each step, the LSTM uses the current input x_t and its internal memory to calculate the hidden state h_{t-1}. This hidden state acts as a summarized representation of the input sequence up to the current time. The hidden state h_t represents a compressed overview of the input sequence up to time step t. Each LSTM cell has a cell state c_t, and the following three gates: forget gate f_t, output gate o_t, and input gate i_t, which serve as a memory mechanism and allow the network to preserve information from the past.

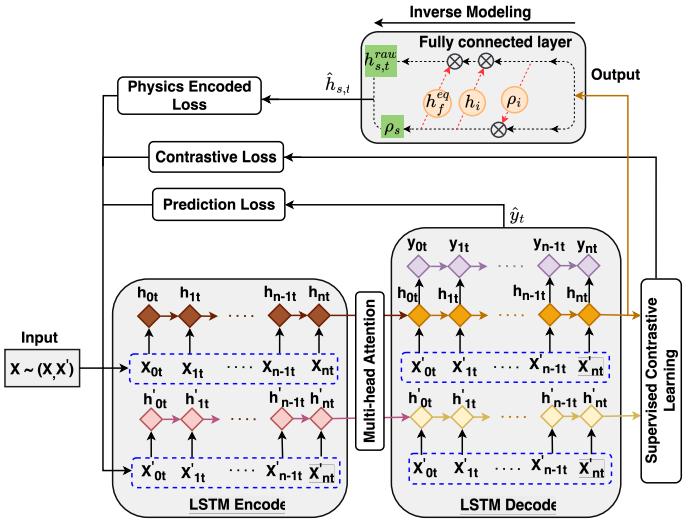

Figure 1: The Physics-Encoded Representation Learning for Inverse (PhysERL-Inv) framework presents an outline, showing the fusion of physics-based constraints, representation learning, and an invertibility approach based on hydrostatic balance equation.

The encoder processes the input sequence sequentially, updating its hidden and cell states at each time step. After processing the entire input sequence, the encoder produces a final hidden state h_T^{enc} and a final cell state c_T^{enc}, which capture the essential information needed for the decoder.

$$h_t^{\text{enc}}, c_t^{\text{enc}} = \text{LSTM}^{\text{enc}}(x_t, h_{t-1}^{\text{enc}}, c_{t-1}^{\text{enc}})$$

Multi-Head Self-Attention Mechanism: To capture intricate temporal dependencies and weigh the significance of different time steps, we employ a multi-head self-attention mechanism on the LSTM-encoded sequences. This module comprises multiple independent self-attention heads that learn to attend to different parts of the input, enabling the model to capture diverse temporal relationships. Each head computes attention as: The attention-enhanced input at time step t, denoted a_t, is computed as a weighted sum of the value vectors $V = \{v_1, \ldots, v_T\}$, where the weights are determined by the similarity between the query q_t and each key k_j, normalized via softmax:

$$a_t = \text{Attention}(q_t, K, V) = \sum_{j=1}^{T} \alpha_{tj} v_j, \quad \text{where } \alpha_{tj} = \frac{\exp\left(\frac{q_t^\top k_j}{\sqrt{d_k}}\right)}{\sum_{l=1}^{T} \exp\left(\frac{q_t^\top k_l}{\sqrt{d_k}}\right)}$$

The concatenated and linearly transformed outputs of these heads allow the model to focus simultaneously on various temporal contexts. This enriched representation is then passed to the decoder LSTM, which leverages it to perform the downstream tasks of prediction and parameter estimation.

Decoder LSTM: The decoder LSTM receives as input the sequence produced by the multi-head self-attention module, which enhances the encoder outputs with context-aware representations. It is initialized with the final hidden state h_T^{enc} and cell state c_T^{enc} from the encoder. The decoder processes this attention-enriched sequence to learn temporally-aware latent representation features.

At each time step t, the decoder updates its hidden state h_t^{dec} and cell state c_t^{dec} based on the previous states and the attention-informed input.

$$h_t^{\text{dec}}, c_t^{\text{dec}} = \text{LSTM}^{\text{dec}}(a_t, h_{t-1}^{\text{dec}}, c_{t-1}^{\text{dec}})$$

where a_t is the attention-enhanced input at time step t.

Inverse Modeling: To integrate physical principles into the learning process, our model leverages inverse modeling as a core mechanism to guide the mapping from input features to snow depth,

enabling physically consistent and flexible representations. In our implementation, only the final hidden state of the decoder, h_T^{dec}, is used as input to the inverse mapping module, which encapsulates temporally-aware features extracted from the entire input sequence and serves as the basis for surjective mapping. The model estimates a set of parameters through a small feedforward neural network f_{params}, which performs a nonlinear transformation of h_T^{dec}. Formally, the parameters are computed as: $[w, b, c] = f_{\text{params}}(h_T^{\text{dec}})$.

These parameters are interpreted as follows: w is snow-ice weight coefficient, b is coupling factor with predicted snow depth, and c is additive bias term. The parameter estimation module is implemented as a multilayer perceptron (MLP) composed of several fully connected layers with ReLU activations. This architecture enables the network to learn complex nonlinear mappings from the latent decoder representation to the physical parameter space.

Each output parameter is computed independently using a final linear transformation followed by a non-linear activation. Specifically, given the decoder's final hidden state h_T^{dec}, the parameters are estimated as:

$$w = f(\mathbf{w}_w^\top h_T^{\text{dec}} + b_w), \quad = f(\mathbf{w}_b^\top h_T^{\text{dec}} + b_b), \quad c = f(\mathbf{w}_c^\top h_T^{\text{dec}} + b_c),$$

where f is a non-linear function (ReLU), and $\mathbf{w}_w, \mathbf{w}_b, \mathbf{w}_c$ and b_w, b_b, b_c are learnable parameters. The network is trained using gradient-based optimization to minimize the reconstruction error based on these estimated parameters. Through this process, it effectively learns to invert the forward process and uncover latent physical parameters driving snow depth.

Physics Encoding: Through inverse modeling, we use learned parameters to compute snow depth at every time step, which is an approach that leverages the physical parameters of the hydrostatic balance equation. This physical relationship is simplified into a linear form for our model, represented as:

$$\hat{s}_t = w \cdot \bar{\rho} + b \cdot \hat{y}_t + c, \tag{2}$$

where $\bar{\rho}$ represents the mean density, computed from the input sequence x. $\hat{y}_t$ denotes the predicted snow depth at time step t, as output by the decoder LSTM. w, b, c corresponds to the physical parameters estimated by surjective mapping.

The above formulation can be explained using the physical parameters of the hydrostatic balance equation.

$$\hat{h}_{s,t} = h_f^{\text{eq}} \cdot \bar{\rho}_{\text{snow}} + h_i \cdot h_{s,t}^{\text{raw}} + \rho_i, \tag{3}$$

Here, $\bar{\rho}_{\text{snow}}$ represents the mean snow density computed from the input sequence x, and $h_{s,t}^{\text{raw}}$ denotes the predicted snow depth at time step t. The parameters h_f^{eq}, h_i, and ρ_i correspond to physical parameters estimated by the inverse modeling approach, informed by the input features. This simplification assumes concurrent recovery of meaningful patterns analogous to the pretext target label generation using equation 1, thereby maximizing the embedding of physics into the neural network's training.

Physics Encoded Loss: The snow depth ($\hat{s}_t$) predicted via physics encoding is compared against the ground truth to calculate loss:

$$\mathcal{L}_{\text{PE-pred}} = \frac{1}{T} \sum_{t=1}^{T} (s_t - \hat{s}_t)^2, \tag{4}$$

where T is the number of time steps in the sequence. This loss encourages the model to make predictions that are as close as possible to the observed snow depth values.

Prediction: The snow depth prediction module leverages the full sequence of decoder hidden states $\{h_1^{\text{dec}}, h_2^{\text{dec}}, \ldots, h_T^{\text{dec}}\}$. At each time step t, a scalar snow depth prediction $\hat{y}_t$ is computed by applying a linear transformation to the corresponding decoder hidden state: $\hat{y}_t = \mathbf{W}_d h_t^{\text{dec}} + b_d$, where $\mathbf{W}_d$ and b_d are learnable parameters shared across time steps.

Prediction Loss: The predicted sequence $\{\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_T\}$ is supervised using an MSE loss against the ground truth snow depth values $\{y_1, y_2, \ldots, y_T\}$. This loss serves as the direct, data-driven path for learning to predict snow depth from the temporally aware decoder representations:

$$\mathcal{L}_{\text{pred}} = \frac{1}{T} \sum_{t=1}^{T} (y_t - \hat{y}_t)^2, \tag{5}$$

where T denotes the sequence length, y_t is the true snow depth at time step t, and $\hat{y}_t$ is the model's prediction.

3.3 Supervised Contrastive Learning

To encourage the model to learn stable and invariant representations, we incorporate a supervised contrastive learning objective. This approach trains the network to produce similar representations for inputs that are semantically equivalent. Our method is termed supervised because, unlike traditional methods that rely solely on data augmentation to define positive pairs, we use physics-based "labels" to generate them. As shown in Figure 1, given an original input sequence X and its augmented version X'(which is a physically-similar sample from our dataset), we first apply the same encoder and attention modules to both sequences. The augmented sequence X' is generated by applying a small perturbation by adding Gaussian noise to the original input. Both sequences are processed independently through the encoder LSTM, the multi-head self-attention mechanism, and the decoder LSTM. And then the final hidden states h_T and h'_T are used as embedding vectors in the contrastive loss. This objective encourages the model to learn a robust latent space where the representations of a time series and its noisy version are pulled closer together, while pushing them away from other unrelated samples.

Contrastive Loss: Let's say z_i and z_j are normalized embeddings of h_T and h'_T. A similarity matrix is computed, and the combined embeddings $[z_i, z_j]$ are used to create a combined similarity matrix. To focus on the similarity between non-identical pairs, the diagonal elements of this combined matrix are removed. Cross-entropy loss is then applied to this processed similarity matrix, using labels that indicate which pairs of embeddings correspond to augmentations of the same original input. The cross-entropy loss in this context is calculated as:

$$\mathcal{L}_{\text{Contrastive}} = -\frac{1}{N}\sum_{i=1}^{N}\log\left(\frac{\exp(s(z_i, z_i^+)/\tau)}{\sum_{j=1}^{N}\exp(s(z_i, z_j)/\tau)}\right), \tag{6}$$

where N is the batch size, and $s(z_i, z_j)$ represents the scaled similarity between embeddings z_i and z_j. z_i^+ denotes the positive pair of z_i (i.e., the embedding of the augmented version of the same input). τ is the temperature parameter (scale). This process encourages the model to produce similar embeddings for an input and its augmented version, while distinguishing them from other inputs.

Total Loss The total loss is calculated by adding the combined model loss to the weighted sums of the absolute values L1 and squared values L2 of the model's weights. The hyperparameters λ_1 and λ_2 determine the influence of each regularization term on the overall loss. We used both L1 and L2 regularization to combine their unique benefits. The L2 component was essential for mitigating overfitting and improving the model's generalization capabilities on unseen data. The L1 component, while not explicitly for sparsity in our model, helped to further control the complexity of the network by encouraging smaller, non-zero weights. This combined approach resulted in a more stable and robust model, which is critical for complex and noisy climate datasets. The combined model loss consists of the snow depth loss, the estimated snow depth loss, and the contrastive loss, as defined below:

$$\text{Total Loss} = \mathcal{L}_{\text{PE-pred}} + \mathcal{L}_{\text{pred}} + \mathcal{L}_{\text{Contrastive}} + \lambda_1\sum_{i=1}^{n}|w_i| + \lambda_2\sum_{i=1}^{n}w_i^2 \tag{7}$$

4 Experimental Results

4.1 Data and Implementation

The dataset comprises ERA5 reanalysis data (Hersbach et al., 2020) from the European Centre for Medium-Range Weather Forecasts (ECMWF). ERA5 ingests a wide variety of observational data, a significant portion of which comes from remote sensing instruments (Hersbach et al., 2020). We acquired spatiotemporal data from January 1, 1995, to 2011 (10,958 time steps), covering the central Arctic Ocean region, which roughly corresponds to the latitude range of approximately $70°Nto85°$N. The data has a spatial resolution of 0.25° × 0.25° (approximately 25 km). It includes key parameters related to snow depth and sea ice thickness: snow albedo, snow density, and

sea ice concentration. We therefore process these variables in the context of sea ice thickness and freeboard height, which are critical for understanding the dynamics within this specific Arctic region. To extract spatially aggregated information for our time series prediction, spatial averaging was applied across the grid points within the orange highlighted area to produce daily time series. Z-score normalization was then performed using the training set.

The model was trained and evaluated using a chronologically split dataset spanning from January 1, 1995, to March 11, 2011, with the first 5,461 days used for training and the subsequent 589 days reserved for testing. Training was conducted over 500 epochs using the Adam optimizer with a learning rate of 0.0005. Data were processed in shuffled batches of 16. The model's learnable parameters were optimized by minimizing a composite loss function that included the MSE for both actual and estimated snow depth predictions, as well as a contrastive loss. The model input X includes the snow density field, while the target Y corresponds to the normalized snow depth proxy derived from Equation 1. Input and target sequences are constructed with a fixed length of 10 days, and supervision is applied at the final time step of each sequence to guide learning. The model architecture consists of a two-layer encoder LSTM with 64 hidden units and a dropout rate of 0.4, followed by a four-headed self-attention mechanism that captures temporal dependencies. An identically configured decoder LSTM then processes the attended representations, which are passed to a fully connected layer for snow depth prediction and a three-layer feed-forward network with ReLU activations for physical parameter estimation. The raw outputs for the physical parameters are subsequently transformed to ensure they remain within meaningful and stable ranges.

4.2 Baselines

We compare prediction accuracy across four baseline models to evaluate the effectiveness of our proposed approach. A vanilla LSTM is included to isolate the contribution of the autoencoder architecture and the invertibility mechanism relative to a standard sequential model. The BiLSTM baseline BiLSTM (Ghosh et al., 2022) extends this comparison by capturing bidirectional temporal dependencies, serving as a reference for the benefits of incorporating invertibility into more complex recurrent structures. Neural ODE (Chen et al., 2018) provides a continuous-time framework well-suited for modeling dynamic physical parameters, making it a natural baseline for assessing performance in parameter estimation. Finally, ResNet-50 (He et al., 2016), viewed as an approximate ODE solver with inherent reversibility properties, offers a strong foundation for comparison due to its established role in linking deep residual learning to differential equation frameworks. We compare the prediction accuracy among four different models.

4.3 Results and Discussion

Table 1 provides a comprehensive comparison of the proposed PhysERL-Inv framework against several deep learning baselines, including LSTM, BiLSTM, NeuralODE, and ResNet50. While these models demonstrate a varying capacity to capture underlying data patterns, their architectural constraints present distinct limitations. For instance, recurrent models excel at capturing temporal dependencies but often struggle with long-term stability, while NeuralODE's continuous-time formulation is limited by the difficulty of learning highly nonlinear dynamics directly from data. Similarly, ResNet50, although effective at extracting hierarchical features, lacks a mechanism to embed domain-specific physical constraints.

Table 1: Model performance comparison across baselines and PhysERL-Inv. The percentage improvements are calculated with respect to the PhysERL-Inv model, which serves as the benchmark.

Model	MSE	RMSE	MSE Improv. (%)	RMSE Improv. (%)	MSE Rank	RMSE Rank
LSTM	0.4679	0.6840	**23.73**	**12.67**	3	3
NeuralODE	0.5066	0.7117	**29.59**	**16.09**	2	2
ResNet50	0.4308	0.6563	**17.17**	**8.99**	4	4
BiLSTM	0.5263	0.7255	**32.20**	**17.68**	1	1
PhysERL-Inv	**0.3568**	**0.5973**	–	–	–	–

In contrast, PhysERL-Inv consistently outperforms all baselines across the evaluation metrics, ranking first in both MSE and RMSE. This is clearly demonstrated by significant percentage improve-

ments, including a 32.20% reduction in MSE and a 17.68% reduction in RMSE compared to the best-performing baseline (BiLSTM). The results underscore the critical value of embedding physical knowledge directly into deep learning architectures to create a more accurate and reliable framework.

We conducted an ablation study (Table 2) to evaluate the role of supervised contrastive learning (SCL) within the PhysERL-Inv framework, specifically testing its robustness in data-limited conditions.The results clearly show that the PhysERL-Inv model performs consistently better when SCL is included. We observed systematic reductions in both mean squared error (MSE) and root mean squared error (RMSE) across all training scenarios. This study highlights two key findings. First, SCL provides meaningful benefits by helping the model learn more accurate and generalizable representations. Second, and most importantly, these improvements are consistent even when the amount of training data is significantly reduced. This suggests that SCL is particularly valuable in settings where observational data are scarce, as it helps the model maintain stability and limit the degradation in accuracy that typically occurs in data-limited situations. In short, this study confirms that SCL is a crucial component for strengthening the PhysERL-Inv framework, making it a more robust and reliable tool for scientific prediction.

Table 2: Ablation study comparing PhysERL-Inv performance with and without supervised contrastive learning (SCL)

Training Data	Without SCL		With SCL	
	MSE	RMSE	MSE	RMSE
Sample 1 (80%)	0.6601	0.8125	**0.5926**	**0.7698**
Sample 2 (60%)	0.6588	0.8117	**0.6037**	**0.7770**
Sample 3 (50%)	0.8675	0.9314	**0.7940**	**0.8911**

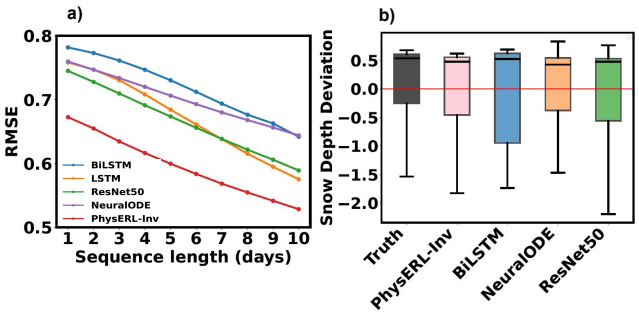

Figure 2: Comparison of model performance: (a) RMSE over time across baseline models and PhysERL-Inv; (b) box-and-whisker plot of snow depth deviations for the ground truth, PhysERL-Inv, and baseline models.

The figure (2a) presents a comparison of the Root Mean Square Error (RMSE) for five distinct models, illustrating how their performance changes with varying sequence lengths. The analysis demonstrates that as the sequence length increases, the RMSE for all models generally decreases, which indicates that a greater amount of historical data leads to improved prediction quality. The PhysERL-Inv model consistently shows the lowest RMSE across all sequence lengths, suggesting it is the most stable model among those compared. In contrast, the BiLSTM model generally has the highest RMSE, and the LSTM, ResNet50, and NeuralODE models perform between these two extremes.

The box-and-whisker plots in Figure 2b illustrate the distributions of snow depth anomalies for the ground truth, baseline models, and the proposed PhysERL-Inv. A horizontal dashed red line at zero represents perfect alignment with the average snow depth condition. Among the baselines, BiLSTM shows reasonably consistent performance, with a median close to zero and a spread comparable to the ground truth, though it produces fewer negative outliers. This suggests that BiLSTM captures the overall shape of the distribution while underrepresenting extreme values. ResNet50 also has a median near zero, but its predictions show a narrower spread than the ground truth, indicating the model may underestimate the full variability of the anomalies. In contrast, NeuralODE exhibits greater variability with a median slightly above zero, suggesting an upward bias, and contains more significant negative outliers, pointing to reduced stability and increased deviation from the true distribution. PhysERL-Inv demonstrates the closest agreement with the ground truth. Its predictions have medians very near zero and a spread that matches the true anomaly distribution. The relatively low number of outliers indicates stable predictions and an accurate representation of the overall distribution. Overall, PhysERL-Inv provides the most reliable and consistent estimates, successfully capturing both the central tendency and variability of snow depth anomalies.

Figure 3 illustrates the PhysERL model's prediction of seasonal changes in snow depth from late 2009 through early 2011, demonstrating a strong correlation of r=0.81 with the true values. The plot's most prominent feature is the distinct seasonal pattern. During the winter season, both the true and predicted snow depths show a consistent downward trend, indicating a period of accumulation. This is followed by a sharp upward trend as snow melts during the transition to spring. During the summer months, both truth and PhysERL remain near zero, showing minimal deviation. The model's reliability is further highlighted by the shaded band, which represents one standard deviation from the predicted values and contains the majority of the true data points. The ability of the PhysERL model to accurately capture these seasonal dynamics, from accumulation to melt, is crucial for improving our understanding of how snow cover impacts sea ice thickness. The results also reinforce a broader point about the future of computational climate science. Its path lies not in replacing physical models with machine learning, but in integrating the strengths of both approaches. By bridging physics and AI, PhysERL-Inv exemplifies how hybrid frameworks can move beyond short-term predictive gains to deliver models that enhance process-level understanding, provide robustness under extrapolation, and ultimately expand the toolkit available for tackling pressing questions in Earth and environmental sciences.

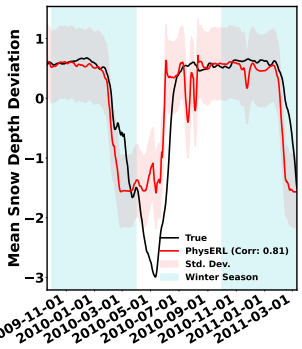

Figure 3: Time series of predicted and true mean snow depth seasonal pattern

Figure 4 compares the Probability Density Function (PDF) of the ground truth and the PhysERL-Inv model's predictions. This comparison provides a more comprehensive evaluation than point-estimate metrics like MSE by assessing the model's ability to learn the underlying statistical distribution of the data. In climate science, where a perfect one-to-one match in geospatial grids is often not expected, the alignment of the probability distributions becomes a more crucial evaluation criterion. A strong correspondence between the predicted and true PDFs indicates that the model is not merely a regression function for individual data points but is capable of generalizing the data's generative process. This is crucial for capturing the system's overall statistical behavior, including the frequency and likelihood of different outcome magnitudes.

The PhysERL-Inv model successfully captures the unimodal nature of the measured deviations in the ground truth distribution (Fig. 4). By incorporating physics encoding, it captures real-world variabilities, reducing prediction error and producing more physically consistent and robust predictions.

5 Discussion and Future Work

We introduce PhysERL-Inv, a novel framework that integrates physics encoding through inverse modeling with representation learning via supervised contrastive learning. Our framework uniquely implements a surjective inverse mapping, which allows for multiple plausible physical explanations for observed data, making it particularly well-suited for modeling the dynamic, time-varying behaviors of seasonal systems. When applied to complex polar geospatial data, PhysERL-Inv effectively captured the seasonal evolution of snow depth. We demonstrate our model's robustness by consistently achieving a 20% performance gain over multiple baselines, effectively capturing both seasonal trends and finer-scale variability. Through our unique, simplified inverse modeling approach using Arctic sea ice data, we demonstrate its potential for predicting snow depth when observational data is unavailable. This is achieved by using hidden parameters of Arctic variables and incorporating physics relevant to snow depth to forecast future conditions. Ultimately, our work showcases the practical value of physics-encoded machine learning in environmental systems and highlights its potential for fostering interdisciplinary collaboration within geoscience.

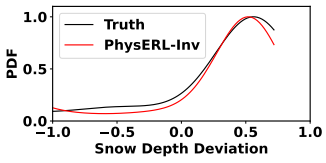

Figure 4: PDF of mean snow depth anomalies for true values and PhysERL-Inv predictions

AUTHOR CONTRIBUTIONS

This work uses a large language model to refine and clarify the language, improve sentence structure, and polish the writing.

ACKNOWLEDGMENTS

This work uses a large language model to refine and clarify the language, improve sentence structure, and polish the writing.

REFERENCES

A. Adler, M. Araya-Polo, and T. Poggio. Deep learning for seismic inverse problems: Toward the acceleration of geophysical analysis workflows. *IEEE Signal Processing Magazine*, 38:89–119, 2021.

A. Bardes, J. Ponce, and Y. LeCun. Vicregl: Self-supervised learning of local visual features. *Advances In Neural Information Processing Systems*, 35:8799–8810, 2022.

T. Bubba, G. Kutyniok, M. Lassas, M. März, W. Samek, S. Siltanen, and V. Srinivasan. Learning the invisible: A hybrid deep learning-shearlet framework for limited angle computed tomography. *Inverse Problems*, 35:064002, 2019.

R. Chen, Y. Rubanova, J. Bettencourt, and D. Duvenaud. Neural ordinary differential equations. *Advances In Neural Information Processing Systems*, 31, 2018.

Y. Fan and L. Ying. Solving electrical impedance tomography with deep learning. *Journal Of Computational Physics*, 404:109119, 2020.

H. Fang, S. Wang, M. Zhou, J. Ding, and P. Xie. Cert: Contrastive self-supervised learning for language understanding. *ArXiv Preprint ArXiv:2005.12766*, 2020.

S. Faroughi, N. Pawar, C. Fernandes, M. Raissi, S. Das, N. Kalantari, and S. Mahjour. Physics-guided, physics-informed, and physics-encoded neural networks in scientific computing. *ArXiv Preprint ArXiv:2211.07377*, 2022.

H. Ghorbanidehno, A. Kokkinaki, J. Lee, and E. Darve. Recent developments in fast and scalable inverse modeling and data assimilation methods in hydrology. *Journal Of Hydrology*, 591:125266, 2020.

R. Ghosh, A. Renganathan, K. Tayal, X. Li, A. Khandelwal, X. Jia, C. Duffy, J. Nieber, and V. Kumar. Robust inverse framework using knowledge-guided self-supervised learning: An application to hydrology. In *Proceedings Of The 28th ACM SIGKDD Conference On Knowledge Discovery And Data Mining*, pp. 465–474, 2022.

K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings Of The IEEE Conference On Computer Vision And Pattern Recognition*, pp. 770–778, 2016.

Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, Adrian Simmons, Cornel Soci, Saleh Abdalla, Xavier Abellan, Gianpaolo Balsamo, Peter Bechtold, Gionata Biavati, Jean Bidlot, Massimo Bonavita, Giovanna De Chiara, Per Dahlgren, Dick Dee, Michail Diamantakis, Rossana Dragani, Johannes Flemming, Richard Forbes, Manuel Fuentes, Alan Geer, Leo Haimberger, Sean Healy, Robin J. Hogan, Elías Hólm, Marta Janisková, Sarah Keeley, Patrick Laloyaux, Philippe Lopez, Cristina Lupu, Gabor Radnoti, Patricia de Rosnay, Iryna Rozum, Freja Vamborg, Sebastien Villaume, and Jean-Noël Thépaut. The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049, 2020. doi: https://doi.org/10.1002/qj.3803. URL `https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3803`.

M. Innes, A. Edelman, K. Fischer, C. Rackauckas, E. Saba, V. Shah, and W. Tebbutt. A differentiable programming system to bridge machine learning and scientific computing. *ArXiv Preprint ArXiv:1907.07587*, 2019.

L. Jing and Y. Tian. Self-supervised visual feature learning with deep neural networks: A survey. *IEEE Transactions On Pattern Analysis And Machine Intelligence*, 43:4037–4058, 2020.

A. Karpatne, X. Jia, and V. Kumar. Knowledge-guided machine learning: Current trends and future prospects. *ArXiv Preprint ArXiv:2403.15989*, 2024.

S. Karumuri and I. Bilionis. Learning to solve bayesian inverse problems: An amortized variational inference approach using gaussian and flow guides. *Journal Of Computational Physics*, 511: 111903, 2024.

Y. Khoo and L. Ying. Switchnet: a neural network model for forward and inverse scattering problems. *SIAM Journal On Scientific Computing*, 43:A1105–A1132, 2021.

N. Kovachki, Z. Li, B. Liu, K. Azizzadenesheli, K. Bhattacharya, A. Stuart, and A. Anandkumar. Neural operator: Learning maps between function spaces. *ArXiv Preprint ArXiv:2108.08481*, 2021.

K. Kuang, F. Dean, J. B Jedlicki, D. Ouyang, A. Philippakis, D. Sontag, and A. Alaa. Med-real2sim: Non-invasive medical digital twins using physics-informed self-supervised learning. *Advances In Neural Information Processing Systems*, 37:5757–5788, 2025.

F. Kwasniok. Linear inverse modeling of large-scale atmospheric flow using optimal mode decomposition. *Journal Of The Atmospheric Sciences*, 79:2181–2204, 2022.

R. Kwok and G. Cunningham. Icesat over arctic sea ice: Estimation of snow depth and ice thickness. *Journal Of Geophysical Research: Oceans*, 113, 2008.

Haili Li, Chang-Qing Ke, Xiaoyi Shen, Qinghui Zhu, Yu Cai, and Lanhua Luo. The varied role of atmospheric rivers in arctic snow depth variations. *Geophysical Research Letters*, 51(14):e2024GL110163, 2024. doi: https://doi.org/10.1029/2024GL110163. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2024GL110163. e2024GL110163 2024GL110163.

T. Liao and G. Li. Metaheuristic-based inverse design of materials–a survey. *Journal Of Materiomics*, 6:414–430, 2020.

X. Liu, F. Zhang, Z. Hou, L. Mian, Z. Wang, J. Zhang, and J. Tang. Self-supervised learning: Generative or contrastive. *IEEE Transactions On Knowledge And Data Engineering*, 35:857–876, 2021.

S. Mo, N. Zabaras, X. Shi, and J. Wu. Deep-learning-based inverse modeling approaches: A subsurface flow example. *Water Resources Research*, 56:e2019WR026731, 2020.

C. Rao, H. Sun, and Y. Liu. Hard encoding of physics for learning spatiotemporal dynamics. *ArXiv Preprint ArXiv:2105.00557*, 2021.

P. Scotti, A. Banerjee, J. Goode, S. Shabalin, A. Nguyen, A. Dempster, N. Verlinde, E. Yundler, D. Weisberg, K. Norman, and Others. Reconstructing the mind's eye: fmri-to-image with contrastive learning and diffusion priors. *Advances In Neural Information Processing Systems*, 36: 24705–24728, 2023.

J. Sun, Z. Niu, K. Innanen, J. Li, and D. Trad. A theory-guided deep-learning formulation and optimization of seismic waveform inversion. *Geophysics*, 85:R87–R99, 2020.

Albert Tarantola. *Inverse problem theory and methods for model parameter estimation*. SIAM, 2005.

K. Tayal, X. Jia, R. Ghosh, J. Willard, J. Read, and V. Kumar. Invertibility aware integration of static and time-series data: An application to lake temperature modeling. In *Proceedings of the SIAM International Conference on Data Mining (SDM)*, 2022. Available: https://epubs.siam.org/doi/10.1137/1.9781611977172.79.

J. Willard, X. Jia, S. Xu, M. Steinbach, and V. Kumar. Integrating scientific knowledge with machine learning for engineering and environmental systems. *ACM Computing Surveys*, 55(1):1–37, 2022.