# Regularized Radar Extrapolation: Combining Advection Physics with Attention-ConvLSTM for Robust Nowcasting

1st Sushen Yadav
*University of Exeter*
Exeter, UK
sy448@exeter.ac.uk

2nd Dr Stefan Siegert
*University of Exeter*
Exeter, UK
S.Siegert@exeter.ac.uk

3rd Dr Aishwaryaprajna
*University of Exeter*
Exeter, UK
Aishwaryaprajna@exeter.ac.uk

***Abstract*—Predicting radar reflectivity patterns for short-term precipitation nowcasting is a critical challenge in weather forecasting. This process can be conceptualized as cloud particle movement governed by the 2D advection equation, a core simple principle in fluid dynamics. To address this challenge, we propose and evaluate four variants of Convolutional Long Short-Term Memory (ConvLSTM) models: Standard, Physics-informed, Attention-based, and Physics-Informed Attention-Based. The Standard ConvLSTM model serves as our baseline. Our Physics-informed variants incorporate physical constraints into the neural network architecture to improve generalization and interpretability. The Attention-based variant leverages recent developments in attention mechanisms for spatiotemporal prediction such as vision transformers capturing complex, long-range dependencies more effectively. Our novel Physics-Informed Attention-Based model combines the strengths of physics-informed learning and attention mechanisms. This fusion approach integrates purely data-driven and physics driven approaches. The model includes a nonlinear convolutional layer and a preprocessing layer designed to incorporate physical insights while leveraging data-driven learning. We've also experimented with dynamically scaling the grid for physics regularization to reduce computational load and training times. Experimental results demonstrate significant improvements in SSIM and MAE of the physics-informed and attention-based variants. We conduct ablation studies on extrapolation capabilities and analyze feature importance using LIME for deeper insights into our developed deep learning models.**

***Index Terms*—Precipitation nowcasting, Radar Extrapolation, Convolutional LSTM, Physics-informed neural networks,**
**Spatiotemporal prediction, Advection equation. Dynamic grid, Extrapolation analysis, LIME interpretability**

## I. Introduction

Weather nowcasting applications are without a doubt challenging due to the chaotic nature of the cloud particles. This behaviour is often considered as an ill-posed problem. The prediction of the next frame in a sequence of radar images can be especially unpredictable thus the need for robust solutions to improve accuracy. One promising approach to tackle this complexity is by incorporating physics laws and mechanics into the models. This integration not only reduces the data load but also enhances the accuracy of predictions with less data, addressing the challenge of radar-based precipitation nowcasting effectively [**2**]. Traditional nowcasting methods rely heavily on large datasets to make predictions. However, the inclusion of physical constraints allows for a more efficient use of data. This directly leads to more accurate predictions with less data. For instance, integrating physical laws into machine learning models has been shown to significantly enhance their performance in predicting weather phenomena [**3**]. This approach not only reduces computational load but also improves the generalizability of the models, making them more reliable in varied scenarios [**9**]. Seeing nowcasting as a spatiotemporal sequence prediction problem offers new perspectives and methodologies to address these challenges. The use of Convolutional Long Short-Term Memory (ConvLSTM) networks has demonstrated considerable potential in handling the spatiotemporal aspects of weather data [**7**]. These networks, when combined with physics-informed layers, can effectively capture the dynamics of weather systems. Furthermore, approaches such as the Physics-guided Neural Networks (PGNN) leverage the underlying physical processes to guide the learning process, resulting in models that are both data-efficient and accurate [**12**]. Incorporating physics into machine learning models not only addresses the ill-posed nature of the problem but also leads to a drastic decrease in training time. This is crucial for nowcasting applications where timely predictions are essential. For example, using an advection-free convolutional neural network for convective rainfall nowcasting has shown significant improvements in both speed and accuracy [**1**]. Similarly, studies have demonstrated that physics-constrained deep learning models can effectively handle high-dimensional surrogate modeling and uncertainty quantification, further enhancing their applicability in real-world scenarios [**14**, **17**, **18**]. As the fusion of data-driven approaches with physical insights continues to evolve, the integration of physical laws into machine learning models is likely to become a standard practice, further enhancing the capabilities of nowcasting applications [**1**, **2**, **4**]. This spatiotemporal sequence prediction approach not only addresses immediate forecasting needs but also contributes to the

*Under the scientific supervision of Dr Stefan Siegert (Senior research fellow), University of Exeter & Internal supervision of Dr Aishwaryaprajna, University of Exeter*

broader goal of understanding and predicting complex weather patterns.

The work outlined in this research bridges the gap between the purely data-driven approaches such as [**1**, **19**] and the purely physics-driven [**21**] approaches. It aims to show the potential of Physics informed neural networks (PINN) [**35**] and tries to gain more exploratory insight into the pain points specific to the application of PINNs to radar extrapolation, such as : The vanishing gradient and generalization problem , The increased computational cost of the training and how to minimize it through experiments on extrapolation capability and dynamic grid.

## II. Related Work

The field of radar-based precipitation nowcasting has seen significant advancements in recent years. Shi et al. [**7**] developed a novel formulation of precipitation nowcasting as a spatiotemporal sequence forecasting problem and proposed the ConvLSTM network. The ConvLSTM extends fully-connected LSTMs by introducing convolutional structures in both the input-to-state and state-to-state transitions, allowing it to capture spatiotemporal correlations more effectively. Building upon this work, Ritvanen et al. [**1**] created the Lagrangian Convolutional Neural Network (L-CNN) for convective rainfall nowcasting. The L-CNN incorporates physics by separating the growth and decay of rainfall from motion using the advection equation. This approach results in better representation of rainfall temporal evolution, less underestimation of rainfall, and more accurate representation of heavy rainfall growth and decay compared to reference models. Wang et al. [**10**] presented an advection-free CNN that uses a dense Lucas-Kanade algorithm to estimate motion fields and a custom loss function with PDE constraints based on the advection-diffusion equation. This physics-based approach preserves spatial coherence and improves nowcasting skill. Kashinath et al. [**6**] surveyed systematic approaches to incorporating physics and domain knowledge into machine learning models for weather and climate applications. They identified common PIML approaches, including custom loss functions, custom architectures to enforce constraints, symmetries and invariances, and building upon physics-based modeling frameworks. Case studies demonstrate that PIML can lead to improved physical consistency, accuracy, generalization, and interpretability compared to pure data-driven models. Other relevant work includes using GANs with temporal coherence for stochastic emulation of precipitation processes [**18**], graph neural networks that capture long-range interactions in PDE-based precipitation models [**17**], and neural operators in Fourier space for efficient solutions of the Navier-Stokes equations governing atmospheric dynamics [**15**]. Luo et al. [**8**] proposed a novel LSTM model with interaction dual attention for radar echo extrapolation, presenting an alternative approach to the problem. Wu et al. [**11**] used a gated attention recurrent neural network for radar-based precipitation nowcasting, offering another deep learning approach. The PredRNN [**20**] stands out as a significant contribution in radar-based precipitation nowcasting. It introduces a novel recurrent unit, the Spatiotemporal LSTM (ST-LSTM), which models both spatial and temporal dependencies in precipitation patterns. The ST-LSTM incorporates memory cells for each spatial location and updates them using information from neighboring locations and previous time steps, capturing the complex spatiotemporal dynamics of rainfall events. This approach has demonstrated improved performance in precipitation nowcasting compared to traditional convolutional LSTM models. Recent advancements include the development of transformer-based models for radar echo extrapolation. Chen et al. [**21**] introduced TempEE, a temporal-spatial parallel transformer that goes beyond autoregression for radar echo extrapolation. Geng et al. [**22**] proposed MS-RadarFormer, a transformer-based multi-scale deep learning model for radar echo extrapolation, further pushing the boundaries of nowcasting capabilities. The integration of physics into machine learning models continues to evolve, with recent works focusing on physics-informed neural networks for high-resolution weather reconstruction [**24**] and physics-informed regularization of deep neural networks [**28**]. These approaches demonstrate the ongoing efforts to combine data-driven methods with physical insights for improved weather prediction.

## III. Aims & Objectives

Radar-based extrapolation can be formulated as a spatiotemporal sequence forecasting problem, where the goal is to predict future radar reflectivity frames given a sequence of past observations. However, this task is inherently challenging due to the turbulent and non-linear nature of atmospheric processes.
We aim to develop specialised deep learning models and test the extrapolation capability, interpretability and the accuracy of the evolution of radar reflectivity patterns by only incorporating the underlying physics of advection.

## IV. Background

This section provides a brief overview of the key technical concepts and techniques underlying our radar nowcasting approach.

### A. Convolutional LSTM

A convolutional neural network is a sparsely connected multilayer network with parameter sharing that is designed to encode invariances and equivariances which are specific to an image. The LSTM, short form for Long Short-Term Memory is a specialized type of recurrent neural networks capable of learning long-term dependencies [**17**, **18**]. Convolutional Long Short-Term Memory (ConvLSTM) networks [**7**] extend the traditional LSTM architecture by replacing matrix multiplications with convolution operations.

### B. Physics-Informed Neural Networks

Physics-informed neural networks (PINN) combine traditional machine learning techniques with physical laws and domain knowledge to improve model accuracy and interpretability [**14**, **15**, **16**]. PINN can be broadly categorized into three types: soft, hard, and regularization-based approaches.

*1) Soft Physics-Informed Neural Networks:* Soft PINN incorporates physical knowledge as additional input features or as part of the loss function. They allow the model to learn from both data and physical constraints [**13**]. This approach is flexible and can handle noisy or incomplete physical knowledge.

*2) Hard Physics-Informed Neural Networks:* Hard PINN enforces physical constraints directly in the model architecture. The predictions always satisfy known physical laws [**11**] . This approach is viable only when the physical laws are well-understood and formulated.

*3) Regularization-Based Physics-Informed Neural Networks:* Regularization-based PINN uses physical knowledge to guide the learning process by adding physics-based terms to the loss function[**14**]. This approach can be seen as a middle ground between soft and hard PINN.

$$\mathcal{L}\text{total} = \mathcal{L}\text{data} + \lambda\mathcal{L}_{\text{physics}} \tag{1}$$

where $\mathcal{L}$data is the data-driven loss, $\mathcal{L}$physics is the physics-based regularization term, and $\lambda$ is a hyperparameter controlling the strength of the regularization [**36**].

### C. Attention Mechanism

Swin attention [**28**] is a hierarchical self-attention mechanism with shifting windows of attention designed for segmentation tasks. Unlike the standard Transformer which computes global self-attention across all tokens, Swin Transformer introduces a local window-based self-attention mechanism.

1) The input feature map is divided into non-overlapping windows of fixed size (e.g, 7x7).
2) Self-attention is computed within each window independently.
3) In alternate layers, the window partitioning is shifted to allow cross-window information exchange.

### D. Local Interpretable Model-agnostic Explanations

Local Interpretable Model-agnostic Explanations (LIME) have emerged as a powerful tool for elucidating the decision-making processes of complex machine learning models in the field of precipitation nowcasting [**2**].
LIME operates by creating local surrogate models that approximate the predictions of the original model in the vicinity of a specific instance [**4**, **37**].This helps us in understanding which features (e.g., specific radar echoes or atmospheric conditions) contribute [**5**] most significantly to a model's predictions.

## V. Design & Implementation

### A. Model Architecture

We propose and evaluate four variants of Convolutional Long Short-Term Memory (ConvLSTM) models: a) Standard ConvLSTM (baseline) [**7**] b) Physics-informed ConvLSTM (ConvLSTM-Physics) [**12**] c) Attention-based ConvLSTM (ConvLSTM-Attention) [**32**] d) Physics-informed Attention-based ConvLSTM The attention mechanism in models (c) and (d) utilizes a shifting window approach to capture long-range dependencies in the spatiotemporal data, similar to the approach used by Tang et al. [**27**] in their SwinLSTM model.

*1) Physics Regularization:* We incorporate physical constraints into models (b) and (d) using the 2D advection equation as a regularization term. This is achieved by computing the PDE residual and adding it to the loss function, following the approach of Raissi et al. [**13**]: The total loss for physics-informed models is a combination of data loss and physics loss, similar to the approach used by Pannell et al. [**29**] in their physics-informed regularization procedure

$$\frac{\partial R}{\partial t} + u\frac{\partial R}{\partial x} + v\frac{\partial R}{\partial y} = 0 \tag{2}$$

where,
R is the radar reflectivity, t is time, and u and v are the velocity components in the x and y directions, respectively. The PDE residual is computed and added to the loss function to penalize predictions that deviate from the physical law:

$$L_{total} = L_{data} + \lambda L_{physics}$$

where $L_{data}$ is the data loss (e.g., MSE), $L_{physics}$ is the physics loss, and $\lambda$ is a hyperparameter balancing the two terms.

*2) ConvLSTM:* The Convolutional LSTM (ConvLSTM) model, introduced by Shi et al. [**7**], extends the traditional LSTM by replacing matrix multiplications with convolution operations. This allows the model to capture spatial dependencies in addition to temporal ones, making it particularly suitable for spatiotemporal sequence prediction tasks like precipitation nowcasting. The core equations of ConvLSTM are as follows:

$$i_t = \sigma(W_{xi} \cdot X_t + W_{hi} \cdot H_{t-1} + W_{ci} \circ C_{t-1} + b_i) \tag{3}$$

$$f_t = \sigma(W_{xf} \cdot X_t + W_{hf} \cdot H_{t-1} + W_{cf} \circ C_{t-1} + b_f) \tag{4}$$

$$C_t = f_t \circ C_{t-1} + i_t \circ \tanh(W_{xc} \cdot X_t + W_{hc} \cdot H_{t-1} + b_c) \tag{5}$$

$$o_t = \sigma(W_{xo} \cdot X_t + W_{ho} \cdot H_{t-1} + W_{co} \circ C_t + b_o) \tag{6}$$

$$H_t = o_t \circ \tanh(C_t) \tag{7}$$

Here, $*$ denotes the convolution operator, $\circ$ represents the Hadamard product, $\sigma$ is the sigmoid function, and $W$ terms are convolutional filter weights. This formulation thereby allows the model to maintain spatial information throughout the processing of sequential data as radar echo extrapolation [**2**]

*3) Attention-ConvLSTM:* The Attention-based ConvLSTM model enhances the standard ConvLSTM by incorporating the self-attention mechanism. The attention mechanism is critical in enabling the model to focus on the most relevant spatial and temporal features. This approach has shown improved performance in various spatiotemporal prediction tasks, including radar echo extrapolation [**8**]. The attention mechanism can be integrated into the ConvLSTM framework as follows:

$$\alpha_t = \text{softmax}(W_a \cdot [H_{t-1}, X_t] + b_a) \tag{8}$$

$$\tilde{X}_t = \alpha_t \circ X_t \tag{9}$$

$$i_t = \sigma(W_{xi} \cdot \tilde{X}_t + W_{hi} \cdot H_{t-1} + W_{ci} \circ C_{t-1} + b_i) \tag{10}$$

$$f_t = \sigma(W_{xf} \cdot \tilde{X}_t + W_{hf} \cdot H_{t-1} + W_{cf} \circ C_{t-1} + b_f) \tag{11}$$

$$C_t = f_t \circ C_{t-1} + i_t \circ \tanh(W_{xc} \cdot \tilde{X}_t + W_{hc} \cdot H_{t-1} + b_c) \tag{12}$$

$$o_t = \sigma(W_{xo} \cdot \tilde{X}_t + W_{ho} \cdot H_{t-1} + W_{co} \circ C_t + b_o) \tag{13}$$

$$H_t = o_t \circ \tanh(C_t) \tag{14}$$

Here, $\alpha_t$ represents the attention weights, which are applied to the input $X_t$ before processing by the ConvLSTM cell. They enable the model to dynamically focus on different spatial regions or input channels, improving its ability to capture complex spatiotemporal dependencies. [**11**]

### B. Dynamic Grid Mechanism

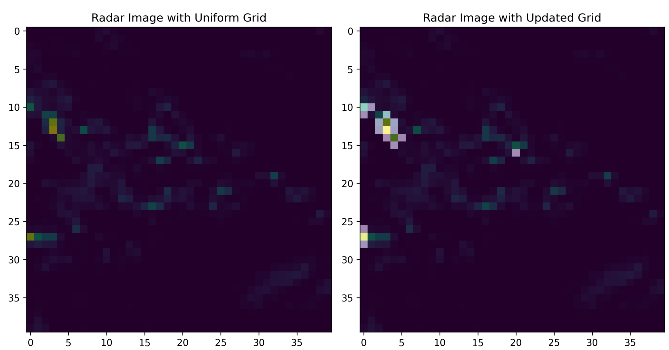


Fig. 1. Dynamic grid Visualisation

The dynamic grid mechanism adapts the computational grid based on the predicted particle movement. This is achieved by tracking the particle positions over time and adjusting the grid resolution accordingly. This approach helps to maintain numerical stability and accuracy, especially in scenarios with rapidly changing velocity fields inspired by the work of Leguen and Thome [**25**] on disentangling physical dynamics from unknown factors.

This can be observed in Fig. 1 where the updated grid is able to highlight the features that are more significant for the movement of cloud particles.

### C. Dataset

In our study, we generate two distinct types of synthetic datasets to simulate particle flow with gradient patterns:

1) Rectangle-based patterns (rect_movie.npy)
2) Radar-like reflectivity patterns (radar_movies.npy)

These datasets are designed to model the movement of particles from positions 1 to 11 while incorporating various velocity scenarios to ensure a diverse and representative dataset.

*1) Velocity Scenarios:* We implement three velocity scenarios, inspired by the approach of Xiong et al. [**28**] in their controlled physics-informed data generation:

1) Randomly generated velocities (40% of cases)
2) Constant velocities (40% of cases)
3) Exponentially increasing and decreasing velocities (20% of cases)

*2) Dataset Generation Process:* The synthetic data is generated using a 2D advection equation model. Let $u(x, y, t)$ represent the concentration of particles at position $(x, y)$ and time $t$. The advection equation is given by:

$$\frac{\partial u}{\partial t} + \mu_x \frac{\partial u}{\partial x} + \mu_y \frac{\partial u}{\partial y} = 0 \tag{15}$$

where $\mu_x$ and $\mu_y$ are the velocities in the $x$ and $y$ directions, respectively.

For the initial condition, we use a Gaussian wave function:

$$u_0(x, y) = \exp(-100((x - x_0)^2 + (y - y_0)^2)) \tag{16}$$

where $(x_0, y_0)$ is the center of the initial distribution.

The solution to the advection equation with this initial condition is:

$$u(x, y, t) = u_0(x - \mu_x t, y - \mu_y t) \tag{17}$$

*3) Rectangle-based Patterns:* For the rectangle-based patterns, we define $n_{blobs}$ (typically 3-5) rectangular regions within the domain. Each rectangle is characterized by its position $(x_{min}, y_{min}, x_{max}, y_{max})$ and individual velocity components $(v_x, v_y)$. The concentration within each rectangle is modeled using a gradient function:

$$u_{rect}(x, y) = \exp\left(\frac{x - x_{min}}{x_{max} - x_{min}} \cdot \frac{y - y_{min}}{y_{max} - y_{min}}\right) \tag{18}$$

The rectangles move according to their individual velocities, wrapping around the domain boundaries when necessary.

*4) Velocity Implementations:*

1) Randomly generated velocities:

$$\mu_x = \mathcal{U}(0, 1), \quad \mu_y = \mathcal{U}(0, 1) \tag{19}$$

where $\mathcal{U}(a, b)$ represents a uniform distribution between $a$ and $b$.

2) Constant velocities:

$$\mu_x = \mu_y = c \tag{20}$$

where $c$ is a constant randomly selected from $\mathcal{U}(0, 1)$ for each frame.

3) Exponentially increasing and decreasing velocities:

$$\mu_x(t) = \mu_y(t) = v_0 \cdot \exp(\sin(2\pi t)) \tag{21}$$

where $v_0$ is a base velocity randomly selected from $\mathcal{U}(0, 1)$ for each frame.

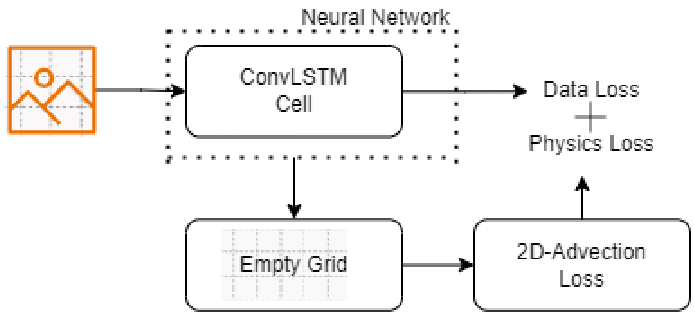


Fig. 2. PhyCell

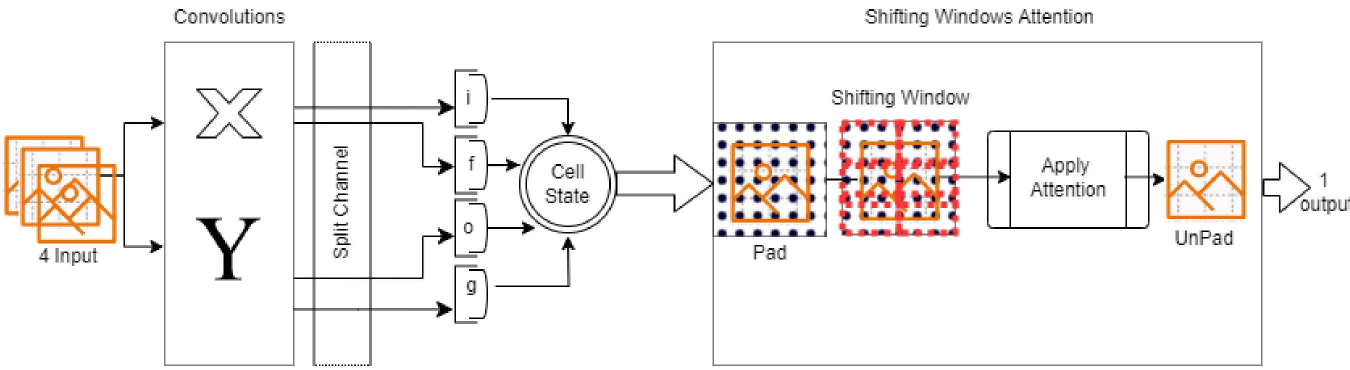


Fig. 3. Attention-ConvLSTM

*5) Data Structures:* The generated data is structured as a 4D numpy array with dimensions $(n_{frames}, n_x, n_y, n_t)$, where:

- $n_{frames}$: number of independent sequences (980 in our case)
- $n_x, n_y$: spatial dimensions (40x40 grid)
- $n_t$: number of time steps in each sequence (20 in our case)

This structure allows for efficient storage and retrieval of the spatiotemporal data and allows us to test various schemes such as training on 4 images and predicting the 5th; i.e, we can train up to 19 images and predict the 20th.

*6) Radar-like Reflectivity Patterns:* We utilise the real radar echo data. The evolution of these patterns follows a continuous field rather than discrete objects.

*7) Data Normalization:* To ensure consistency across frames and to mimic the characteristics of real radar reflectivity data, we normalize each frame to the range [0, 1]:

$$u_{norm}(x, y, t) = \frac{u(x, y, t) - \min(u)}{\max(u) - \min(u)} \tag{22}$$

This normalization step is crucial for maintaining consistent intensity scales across different frames and sequences.

This approach allows us to test and validate our predictive models under controlled conditions, mimicking both simplified (rectangle-based) and more complex (radar-like) real-world phenomena. The inclusion of different velocity scenarios ensures that our models are exposed to a wide range of dynamic behaviors.

### D. Training and Validation

The training and validation process for our radar nowcasting models encompasses a multifaceted approach. We employ three distinct training schemes to comprehensively assess the efficacy of our models where we train on 4 consecutive images and predict the 5th:

a) Standard training: This conventional approach utilizes the mean squared error (MSE) loss function, defined as:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \tag{23}$$

where $y_i$ represents the true value and $\hat{y}_i$ denotes the predicted value.

b) Physics-informed training: Following the pioneering work of Raissi et al. [**13**], we incorporate physical constraints into the loss function. The modified loss function is expressed as:

$$\mathcal{L}_{\text{Physics}} = \mathcal{L}_{\text{MSE}} + \lambda \mathcal{L}_{\text{PDE}} \tag{24}$$

where $\mathcal{L}_{\text{PDE}}$ represents the residual of the governing partial differential equation, and $\lambda$ is a hyperparameter controlling the influence of the physics-based regularization.

c) Physics-informed training with dynamic grid: Building upon the work of Le Guen et al. [**25**], we implement a dynamic grid approach. This method adapts the computational grid based on the evolving features of the radar reflectivity field.

For optimization, we utilize the Adam algorithm with an initial learning rate of 0.001.

Our evaluation metrics, chosen for their prevalence in radar nowcasting literature [**1**, **2**], include:

1. Mean Squared Error (MSE):

MSE squares the pixel-wise differences, giving more weight to larger errors. This is helpful when penalizing significant deviations in the prediction, as these could have a more significant impact on the overall image quality.
MSE provides a smooth error surface which is easier for gradient-based optimization algorithms during model training to converge to a good solution.

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{25}$$

2. Mean Absolute Error (MAE): MAE represents the average pixel-wise absolute difference between the predicted 5th image and the ground truth. It directly quantifies the average magnitude of errors in the prediction.
MAE treats all errors equally, regardless of their direction (positive or negative). This is beneficial in understanding the overall average deviation of the predicted image from the actual one.
MAE is less sensitive to outliers (extreme errors) compared to MSE, as it doesn't square the differences. This is crucial as the radar data is prone to occasional noise or anomalies.

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{26}$$

3. Structural Similarity Index (SSIM): SSIM measures the similarity between images based on their structural information (luminance, contrast, and structure). It aims to mimic how humans perceive image quality. SSIM is less sensitive to minor pixel-wise intensity differences that might not be noticeable to the human eye. Instead, it focuses on preserving the overall structural integrity of the predicted image. SSIM provides a different perspective on image quality compared to MAE and MSE. Using it alongside those metrics offers a more comprehensive evaluation of the model's performance.

$$\mathrm{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{27}$$

where $\mu_x$, $\mu_y$ are the average of $x$ and $y$, $\sigma_x^2$, $\sigma_y^2$ are the variance of $x$ and $y$, $\sigma_{xy}$ is the covariance of $x$ and $y$, and $c_1$, $c_2$ are variables to stabilize the division with weak denominator.

MAE, MSE, SSIM metrics used for analysis are the average over all test cases.

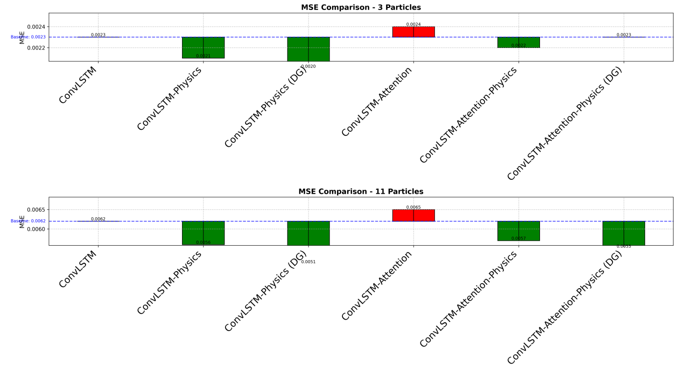


Fig. 4. Comparison of MSE across different models and particle numbers

## VI. Evaluation

### A. Extrapolation Analysis

The extrapolation analysis conducted in this study aims to assess the generalization capabilities of our models when confronted with scenarios of increasing complexity. Our methodology draws inspiration from the work of Saha et al. [**26**] on physics-incorporated convolutional recurrent neural networks for dynamical systems forecasting.

The core premise of our extrapolation analysis lies in training our models on datasets featuring single-particle movements and subsequently evaluating their performance on multi-particle scenarios. This approach allows us to gauge the models' ability to extrapolate learned patterns to more complex, realistic situations that may be encountered in actual radar nowcasting applications.

The results of our extrapolation analysis are presented in Table II, which we can visualize using the following plot:

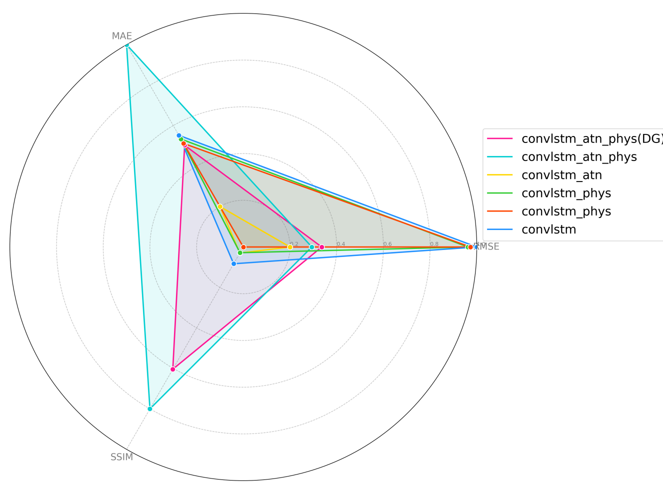


Fig. 5. Comparison metrics in extrapolation scenarios

As shown in the results, Every variant with physics regularization performs better than its counterpart in extrapolation scenarios. The results in Table II, Fig. 5 demonstrate that physics-informed models exhibit superior performance in extrapolating to more complex scenarios. This observation aligns with recent findings in the field of physics-informed neural networks for high-resolution weather reconstruction [**23**] and physics-informed regularization of deep neural networks [**24**]. However, No model variant is able to achieve its best performance in all the metrics. This is observed in the Fig. 5. Structural preservation and temporal consistency are at odds with each other as shown by the performance of models in MAE, MSE vs the SSIM. It can be concluded that the incorporation of dynamic grid does scale the structural similarity better than the variant without it, for e.g the convlstm_atn & the convlstm_atn_phys_dg.

## *B. Robust Radar Nowcasting*

The robust nowcasting experiments conducted in this study are inspired by recent advancements in radar echo extrapolation, such as the work of Wen et al. [**10**] on fusing environment grid point field information. Our approach aims to assess the models' performance on actual radar reflectivity patterns, to get a realistic evaluation of their nowcasting capabilities.

To this end, we utilize real radar movie data (radar_movies.npy) as our primary dataset. This dataset encompasses a diverse range of meteorological phenomena, which allows us to evaluate the models' ability to capture and predict complex and/or real-world weather patterns.

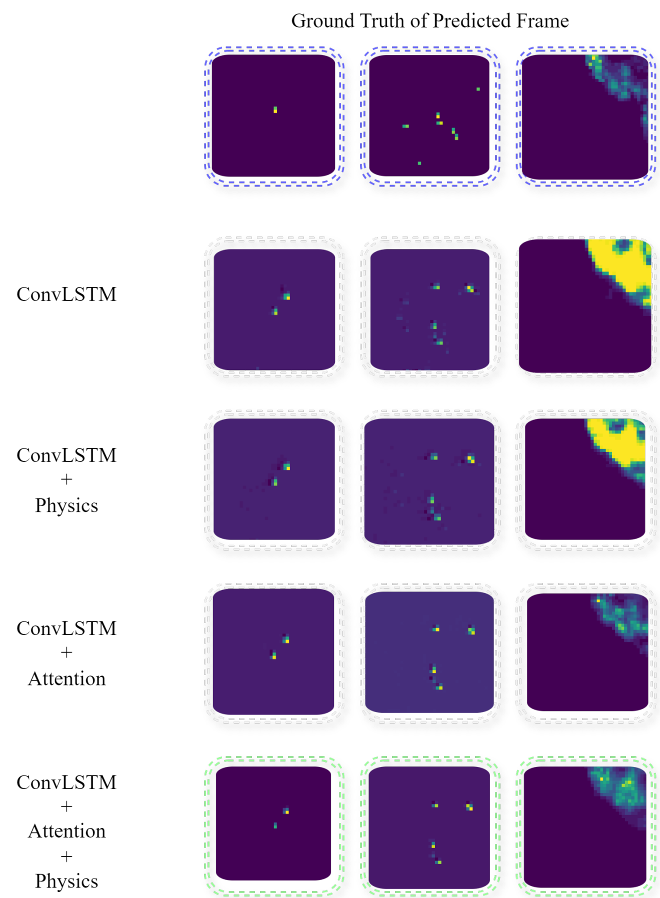


Fig. 6. Predicted frame comparison

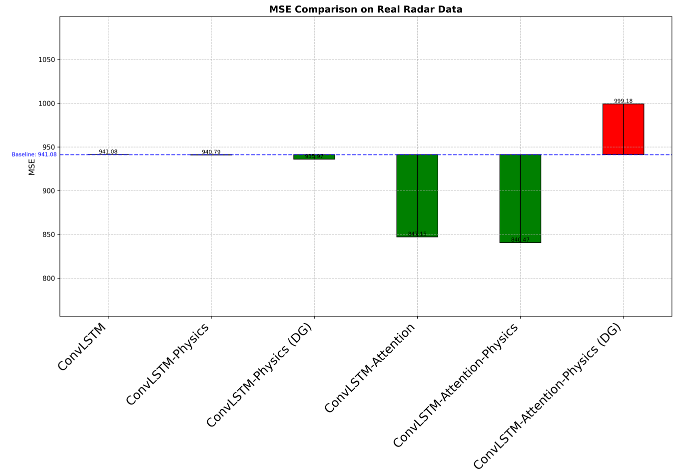


Fig. 7. Comparison of MSE on real radar data across different models and training schemes

The results demonstrate that physics-informed models, particularly those with attention mechanisms, show improved performance in terms of MSE and MAE. This aligns with recent findings in the field, such as the work of Lu et al. [**31**] on attention-driven tree-structured convolutional LSTM for high dimensional data understanding.

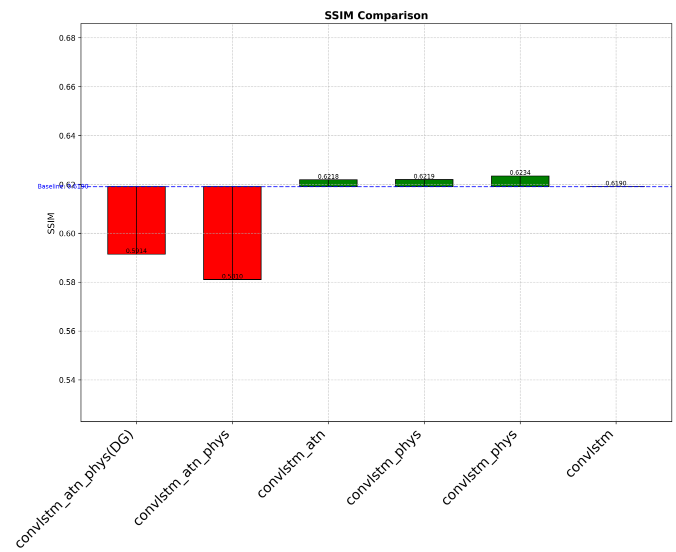


Fig. 8. Comparison of SSIM on real radar data across different models and training schemes

This comprehensive evaluation on real radar data provides crucial insights into the practical applicability of our models for operational weather nowcasting. The superior performance of physics-informed models with attention mechanisms underscores the potential of integrating domain knowledge and advanced neural network architectures in enhancing the accuracy and reliability of radar nowcasting systems.

TABLE I
ROBUST NOWCASTING RESULTS ON REAL RADAR DATA

| Model | MSE | MAE |
|---|---|---|
| ConvLSTM (Standard) | 941.0779 | 13039.1438 |
| ConvLSTM (Physics-Informed) | 940.7903 | 12701.3924 |
| ConvLSTM (Physics-Informed + Dynamic Grid) | 935.9675 | 13032.8909 |
| ConvLSTM-Attention (Standard) | 847.1528 | 12867.7427 |
| ConvLSTM-Attention (Physics-Informed) | **840.4709** | **12337.6249** |
| ConvLSTM-Attention (Physics-Informed + Dynamic Grid) | 999.1830 | 17771.2780 |

TABLE II
EXTRAPOLATION ANALYSIS RESULTS

| Model | 3 Particles | | 11 Particles | |
|---|---|---|---|---|
| | MSE | MAE | MSE | MAE |
| ConvLSTM | 0.0023 | 8.4720 | 0.0062 | 21.8797 |
| ConvLSTM-Physics | 0.0021 | 7.9750 | 0.0056 | 21.1488 |
| ConvLSTM-Physics (Dynamic Grid) | **0.0020** | 7.1409 | **0.0051** | 19.2754 |
| ConvLSTM-Attention | 0.0024 | 6.9694 | 0.0065 | 19.1884 |
| ConvLSTM-Attention-Physics | 0.0022 | 6.9971 | 0.0057 | 16.7913 |
| ConvLSTM-Attention-Physics (Dynamic Grid) | 0.0023 | **6.6932** | 0.0055 | **16.0350** |

### *C. LIME Interpreatability Results*

The LIME results shown by Fig. 12, 11, 10, 9 are the top 5 most important labels and the heatmap of the features that contributed to the prediction of these features. Our approach of LIME explanations is a simpler implementation of the LRP, a technique that attributes the model's output back to its input features [**34**]. It is observed that the physics regularization changes the importance of different features as per the scaling of the physics loss with respect to the data loss. Furthermore, the self shifting window attention mechanism effectively increases the role a region plays in the prediction of the next frame. This is directly responsible for more accurate predictions where the clouds move away from the frame of reference of the radar listening device observing them.

## VII. CONCLUSION

We tackle the complex problem of radar-based precipitation nowcasting [**33**] by integrating physical laws with advanced deep learning neural networks. Nowcasting as a spatiotemporal sequence prediction problem addresses the chaotic and ill-posed nature of radar extrapolation. To this end, we develop variants of Convolutional Long Short-Term Memory (ConvLSTM) models, building upon the work of Shi et al. [**7**] and incorporating physics-informed approaches inspired by Kashinath et al. [**6**]. Our approach addresses the need for better and reliable nowcasting and boosts Mean Absolute Error (MAE) closer to ground truth improving model reliability. The use of ConvLSTM networks combined with attention mechanisms and physics-informed layers effectively handles the complexities of weather prediction [**8**, **11**]. Our experiments with real radar reflectivity data show that physics-informed models maintain physical consistency and offer improved accuracy and computational efficiency, crucial for operational nowcasting. This aligns with the findings of Flora et al. [**3**] and Cuomo et al. [**9**] on the benefits of integrating physical laws into machine learning models for weather phenomena prediction. The fusion of data-driven approaches with physical insights provides a robust solution to nowcasting challenges. As this field evolves, integrating physical laws into machine learning is expected to become standard practice, further enhancing nowcasting capabilities [**1**, **2**, **4**]. This research underscores the potential of physics-informed machine learn-

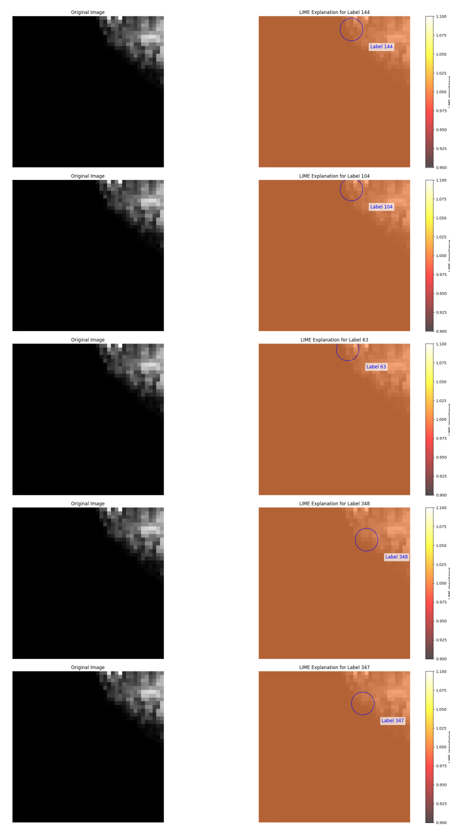

Fig. 9. Top 5 LIME explanations for Attention-ConvLSTM with Physics and dynamic grid

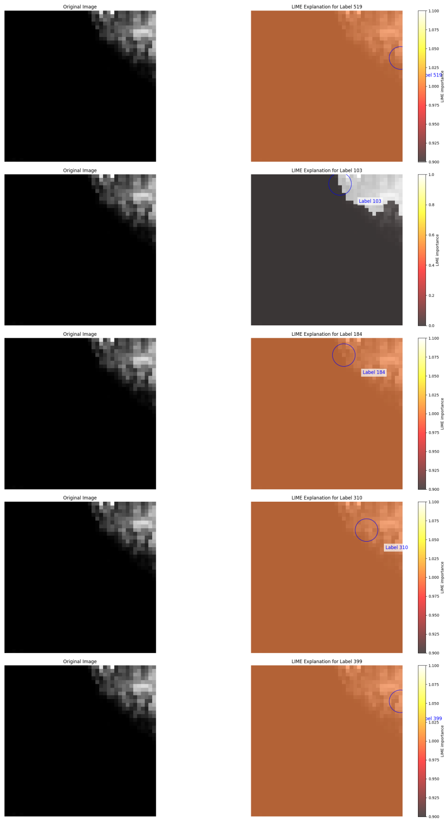

Fig. 10. Top 5 LIME explanations for Attention-ConvLSTM with Physics

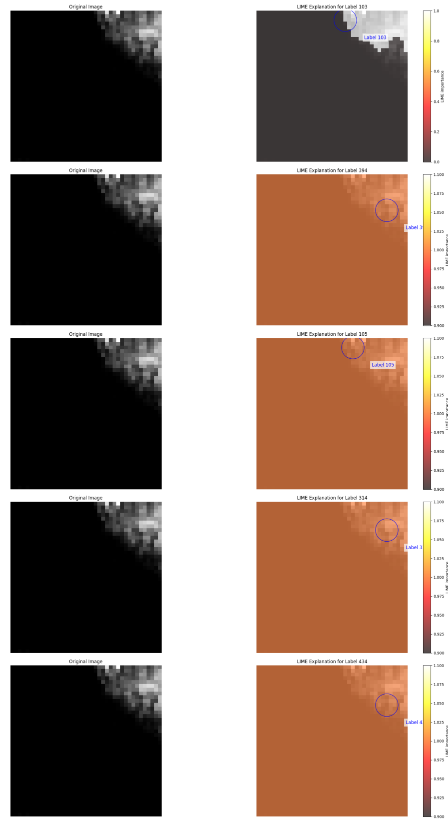

Fig. 11. Top 5 LIME explanations for Attention-ConvLSTM

ing to improve accuracy, efficiency, and reliability in weather forecasting, contributing to the broader goal of understanding and predicting short term complex weather patterns.

*Declaration of Originality*. I am aware of and understand the University of Exeter's policy on plagiarism and I certify that this assignment is my own work, except where indicated by referencing, and that I have followed the good academic practices

*Declaration of Ethical Concerns*. This work does not raise any ethical issues. No human or animal subjects are involved neither has personal data of human subjects been processed. Also no security or safety critical activities have been carried out

## References


[1] J. Ritvanen, B. Harnist, M. Aldana, T. Mäkinen, and S. Pulkkinen. Advection-Free Convolutional Neural Network for Convective Rainfall Nowcasting. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16:1654–1667, 2023.

[2] G. V. Ayzel, M. Heistermann, A. A. Sorokin, O. R. Nikitin, and O. Lukyanova. All convolutional neural networks for radar-based precipitation nowcasting. *Procedia Computer Science*, 2019.

[3] S. Kim, S. Hong, M. Joh, and S.-K. Song. DeepRain: ConvLSTM Network for Precipitation Prediction using Multichannel Radar Data. *CoRR*, abs/1711.02316, 2017.

[4] M. L. Flora, C. K. Potvin, A. McGovern, and S. Handler. A Machine Learning Explainability Tutorial for Atmospheric Sciences. *Artificial Intelligence for the Earth Systems*, 3(1):2024, Jan. 2024.

[5] S. Lundberg and S.-I. Lee. A Unified Approach to Interpreting Model Predictions. *arXiv*, eprint: 1705.07874, 2017.

[6] C.-Y. Lai, P. Hassanzadeh, A. Sheshadri, M. Sonnewald, R. Ferrari, and V. Balaji. Machine learning for climate physics and simulations. In *Proceedings of the Conference*, 2024.

[7] K. Kashinath et al. Physics-informed machine learning: case studies for weather and climate modelling. *Phil. Trans. R. Soc. A*, 379:20200093, 2021.

[8] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-k. Wong, and W.-c. Woo. Convolutional LSTM Network: a machine learning approach for precipitation nowcasting. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1*, pages 802–810. MIT Press, Cambridge, MA, USA, 2015.

[9] C. Luo, X. Li, Y. Ye, Y. Wen, and X. Zhang. A Novel LSTM Model with Interaction Dual Attention for Radar Echo Extrapolation. *Remote Sens.*, 13:164, 2021.

[10] S. Cuomo, V. S. Di Cola, F. Giampaolo et al. Scientific Machine Learning Through Physics–Informed Neural Networks: Where we are and What's Next. *J Sci Comput*, 92:88, 2022.

[11] L. Wang, Y. Dong, C. Zhang, and Z. Heng. Extreme and severe convective weather disasters: A dual-polarization radar nowcasting method based on physical constraints and a deep neural network model. *Atmospheric Research*, 289:106750, 2023.

[12] G. Wu, W. Chen, and H. Jung. Gated Attention Recurrent Neural Network: A Deep Learning Approach for Radar-Based Precipitation Nowcasting. *Water*, 14:2570, 2022.

[13] A. Karpatne, W. Watkins, J. S. Read, and V. Kumar. Physics-guided Neural Networks (PGNN): An Application in Lake Temperature Modeling. *CoRR*, abs/1710.11431, 2017.

[14] M. Raissi, P. Perdikaris, and G.E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686-707, 2019.

[15] Y. Zhu, N. Zabaras, P.-S. Koutsourelakis, and P. Perdikaris. Physics-constrained deep learning for high-dimensional surrogate modeling and uncertainty quantification without labeled data. *Journal of Computational Physics*, 394:56-81, 2019.

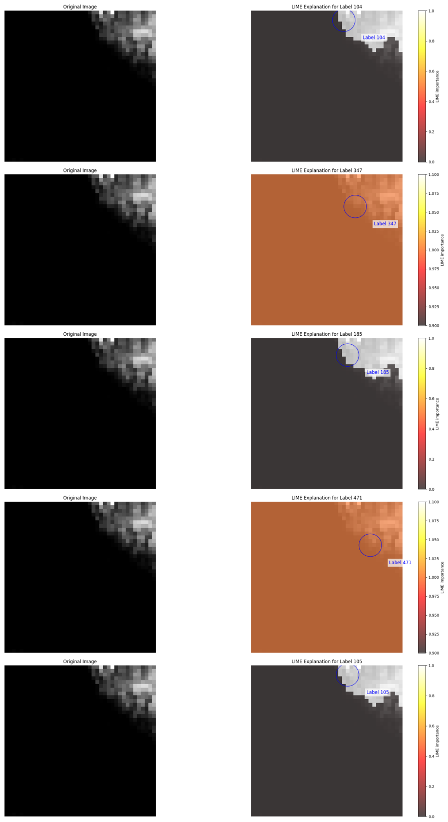


Fig. 12. Top 5 LIME explanations for ConvLSTM

[16] C. Jiang, K. Kashinath, P. Prabhat, and P. Marcus. Enforcing Physical Constraints in CNNs through Differentiable PDE Layer. In *ICLR 2020 Workshop on Integration of Deep Neural Models and Differential Equations*, 2019.

[17] A. Rupe, Mr Prabhat, J. P. Crutchfield, N. Kumar, V. Epifanov, K. Kashinath, O. Pavlyk, F. Schlimbach, M. Patwary, S. Maidanov, and V. Lee. DisCo: Physics-Based Unsupervised Discovery of Coherent Structures in Spatiotemporal Systems. In *2019 IEEE/ACM Workshop on Machine Learning in High Performance Computing Environments (MLHPC)*, 2019.

[18] K. Xu, L. Wen, G. Li, L. Bo, and Q. Huang. Spatiotemporal CNN for Video Object Segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[19] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-Image Translation with Conditional Adversarial Networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5967-5976, Honolulu, HI, USA, 2017.

[20] L. Chen, X. Zhong, F. Zhang, Y. Cheng, Y. Xu, Y. Qi, and H. Li. FuXi: a cascade machine learning forecasting system for 15-day global weather forecast. *npj Climate and Atmospheric Science*, 6, Nov. 2023.

[21] Yunbo Wang, Haixu Wu, Jianjin Zhang, Zhifeng Gao, Jianmin Wang, Philip S. Yu, and Mingsheng Long. PredRNN: A Recurrent Neural Network for Spatiotemporal Predictive Learning. *CoRR*, abs/2103.09504, 2021.

[22] S. Chen, T. Shu, H. Zhao, G. Zhong, and X. Chen, "TempEE: Temporal-Spatial Parallel Transformer for Radar Echo Extrapolation Beyond Autoregression," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1-14, 2023.

[23] H. Geng, F. Wu, X. Zhuang, L. Geng, B. Xie, and Z. Shi, "The MS-RadarFormer: A Transformer-Based Multi-Scale Deep Learning Model for Radar Echo Extrapolation," in *Remote Sensing*, vol. 16, no. 2, p. 274, 2024.

[24] Á. Moreno Soto, A. Cervantes, and M. Soler, "Physics-informed neural networks for high-resolution weather reconstruction from sparse weather stations," in *Open Research Europe*, vol. 4, p. 99, 2024.

[25] M. A. Nabian and H. Meidani, "Physics-Informed Regularization of Deep Neural Networks," Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA, 2024.

[26] V. Le Guen and N. Thome, "Disentangling Physical Dynamics from Unknown Factors for Unsupervised Video Prediction," 1 EDF R and D, Chatou, France; 2 CEDRIC, Conservatoire National des Arts et Métiers, Paris, France, 2024.

[27] P. Saha, S. Dash, and S. Mukhopadhyay. Physics-incorporated convolutional recurrent neural networks for source identification and forecasting of dynamical systems. *Neural Networks*, 144:359-371, 2021.

[28] S. Tang, C. Li, P. Zhang, and R. Tang. SwinLSTM: Improving Spatiotemporal Prediction Accuracy using Swin Transformer and LSTM. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 13470-13479, October 2023.

[29] J. Xiong, O. Fink, J. Zhou, and Y. Ma. Controlled physics-informed data generation for deep learning-based remaining useful life prediction under unseen operation conditions. *Mechanical Systems and Signal Processing*, 197:110359, August 2023.

[30] J.J. Pannell, S.E. Rigby, and G. Panoutsos. Physics-informed regularisation procedure in neural networks: An application in blast protection engineering. *International Journal of Protective Structures*, 13(3):555-578, 2022.

[31] Y. Wen, J. Zhang, D. Wang, C. Wang, and P. Wang. Research on Radar Echo Extrapolation Method by Fusing Environment Grid Point Field Information. *Atmosphere*, 14(6):980, 2023.

[32] Y. Lu, B. Kong, F. Gao, K. Cao, S. Lyu, S. Zhang, S. Hu, Y. Yin, and X. Wang. Attention-driven tree-structured convolutional LSTM for high dimensional data understanding. *Frontiers in Physics*, 11, 2023.

[33] H. Tian and J. Chen. Deep Learning with Spatial Attention-Based CONV-LSTM for SOC Estimation of Lithium-Ion Batteries. *Processes*, 10(11):2185, 2022.

[34] Z. Sokol, D. Kitzmiller, P. Pesice, and J. Mejsnar. Comparison of precipitation nowcasting by extrapolation and statistical-advection methods. *Atmospheric Research*, 123:17–30, April 2013.

[35] Rajasekaran, GK Sriram, A Malini, and Vandana Sharma. Hybrid explainable srnn-lstm architecture for irradiance, temperature and wind speed forecasting. 2023.

[36] D. Davini, B. Samineni, B. Thomas, A. H. Tran, C. Zhu, K. Ha, G. Dasika, and L. White. Using physics-informed regularization to improve extrapolation capabilities of neural networks. In *NeurIPS Workshop on Machine Learning and the Physical Sciences*, 2021. Neural Information Processing Systems Foundation.

[37] D. Garreau and U. von Luxburg. Explaining the Explainer: A First Theoretical Analysis of LIME. In *AISTATS 2020 - 23rd International Conference on Artificial Intelligence and Statistics*, Aug 2020, Palermo /Online, Italy. ffhal-03233013v2f.

# VIII. Appendix

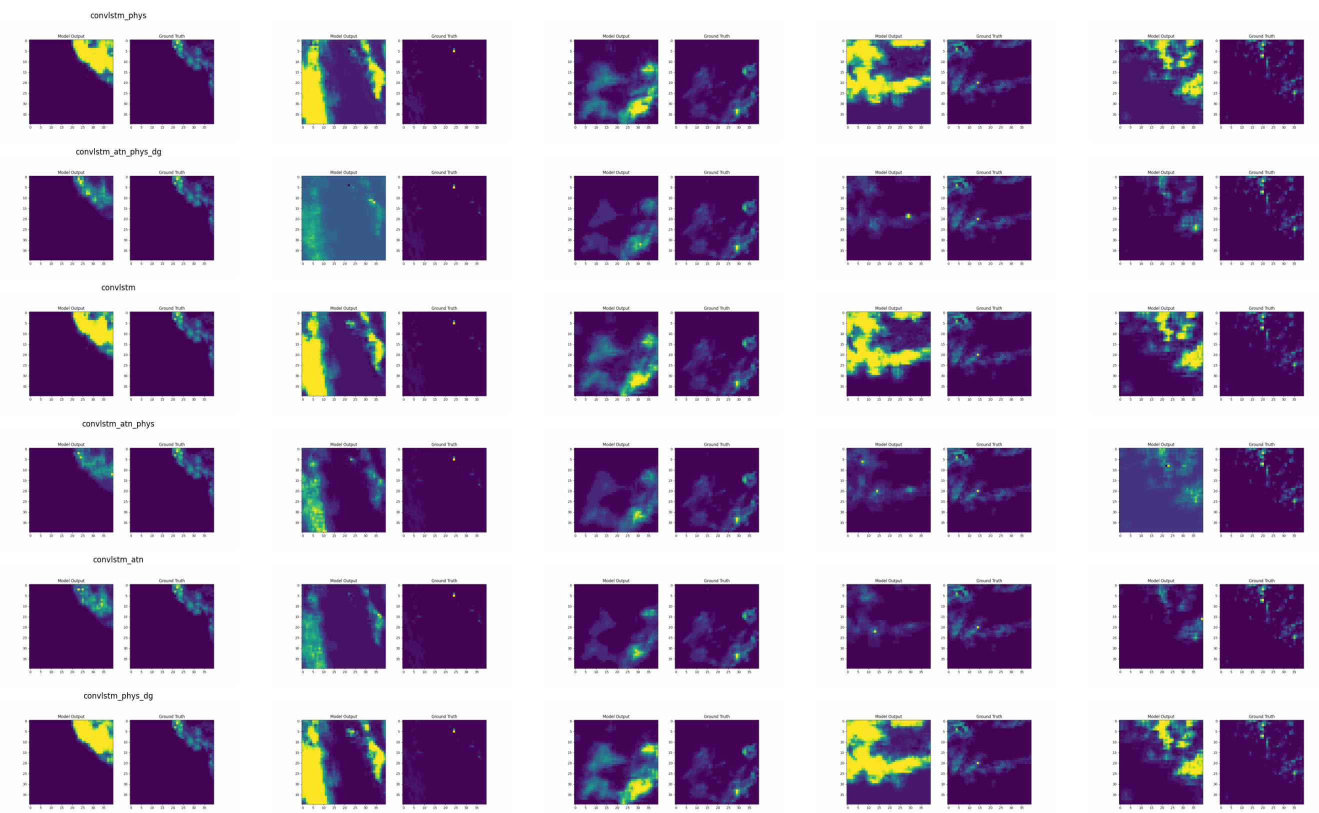


Fig. 13. Radar Movie Results