

FuXi-RTM: A Physics-Guided Prediction Framework with Radiative Transfer Modeling

Qiusheng Huang^{1,2,3†}, Xiaohui Zhong^{1,3†}, Xu Fan³, Hao Li^{1,2,3*}

¹Artificial Intelligence Innovation and Incubation Institute, Fudan University

²Shanghai Innovation Institute

³Shanghai Academy of Artificial Intelligence for Science

Abstract

Similar to conventional video generation, current deep learning-based weather prediction frameworks often lack explicit physical constraints, leading to unphysical outputs that limit their reliability for operational forecasting. Among various physical processes requiring proper representation, radiation plays a fundamental role as it drives Earth’s weather and climate systems. However, accurate simulation of radiative transfer processes remains challenging for traditional numerical weather prediction (NWP) models due to their inherent complexity and high computational costs. Here, we propose FuXi-RTM, a hybrid physics-guided deep learning framework designed to enhance weather forecast accuracy while enforcing physical consistency. FuXi-RTM integrates a primary forecasting model (FuXi) with a fixed deep learning-based radiative transfer model (DLRTM) surrogate that efficiently replaces conventional radiation parameterization schemes. This represents the first deep learning-based weather forecasting framework to explicitly incorporate physical process modeling. Evaluated over a comprehensive 5-year dataset, FuXi-RTM outperforms its unconstrained counterpart in 88.51% of 3320 variable and lead time combinations, with improvements in radiative flux predictions. By incorporating additional physical processes, FuXi-RTM paves the way for next-generation weather forecasting systems that are both accurate and physically consistent.

1. Introduction

Accurate modeling of multi-channel spatiotemporal sequences with complex physical constraints presents significant challenges beyond conventional video prediction tasks. While both tasks essentially predict the next state

of physical systems, RGB video prediction primarily optimizes for perceptual quality and visual coherence, whereas physical system forecasting - exemplified by weather prediction - demands consistency across multidimensional, physically interacting variables. Unlike RGB videos with three fixed channels, weather forecasting involves dozens of interrelated variables across multiple atmospheric layers, governed by physical laws rather than mere visual coherence—requiring models that maintain consistency across all prediction channels simultaneously. To address this fundamental challenge, the most intuitive approach is to incorporate strong physical process constraints that explicitly model the underlying dynamics governing these complex systems, ensuring predictions remain consistent with established physical laws. Among various physical processes, radiative transfer presents an ideal starting point for such integration, given its well-defined physics and fundamental importance. Indeed, radiation is the primary source of energy that drives the Earth’s weather and climate systems, modulating temperature gradients, atmospheric pressure patterns, wind circulations, and moisture distribution [45, 47]. Therefore, accurately simulating radiative processes is essential for weather and climate forecasting, which has significant implications for socioeconomic planning and daily human activities. Numerical weather prediction (NWP) models, which have been foundational tools in meteorology since their emergence in the 1950s [8, 38], simulate radiative transfer processes by resolving interactions between shortwave (SW) and longwave (LW) radiation and atmospheric constituents (e.g., clouds, water vapor, and aerosols), as well as the Earth’s surface [40]. The accuracy of these models depends on spatial resolution, parameterization schemes [44], and the quality of initial conditions. While advancements in NWP models have steadily enhanced forecast accuracy over recent decades [1], challenges persist due to the inherent complexity of cloud microphysics and radiative transfer processes [20]. These challenges are particularly pronounced under cloudy conditions, where uncertainties in cloud properties and verti-

*Corresponding author.

†These authors contributed equally to this work.

cal structures propagate errors in simulated radiative fluxes [48]. Further enhancement of conventional physics-based models faces increasing computational barriers. In contrast, deep learning is revolutionizing weather prediction [5], with data-driven models demonstrating superior computational efficiency and forecast skill for conventional meteorological variables (e.g., temperature, wind, and pressure) compared to the high-resolution deterministic forecasts (HRES) by the European Center for Medium-Range Weather Forecasts (ECMWF) [15], the world’s leading operational prediction center [2, 9–11, 25, 26, 35, 37, 57]. However, current deep learning-based forecasting models remain fundamentally physics-agnostic and may produce unphysical outputs, such as negative humidity [42]. This situation raises concerns about the physical plausibility and long-term stability of these predictions [53], particularly for radiative processes, which are underexplored yet.

To date, the integration of rigid physical processes constraints into deep learning-based weather forecasting models has not been realized. In this study, we propose FuXi-RTM, a hybrid physics-guided deep learning architecture that is designed to integrate data-driven weather forecasts with physics-aware constraints. FuXi-RTM consists of two main components: 1) a primary proven forecasting model based on FuXi [10], and 2) a deep learning-based radiative transfer model (DLRTM). This hybrid design combining the flexibility of data-driven models with weather domain-specific physics, enhancing the physical plausibility and accuracy of the forecasting model’s outputs while maintaining computational efficiency. Experimental results show that FuXi-RTM outperforms its unconstrained counterpart on 88.51% of 3320 variable and lead time combinations (and on 100% for radiative fluxes). By extending this architecture to incorporate other critical physical processes, such as convection, planetary boundary layer (PBL), land surface interactions, and cloud microphysics, deep learning-based weather systems can achieve unprecedented accuracy and physical consistency. Our key contributions as follows:

- We propose FuXi-RTM, the first physics-guided spatiotemporal sequence prediction that integrates explicit physical process modeling with excellent weather forecasting capabilities.
- We propose a straightforward yet effective framework that extends weather prediction models with radiative transfer capabilities without additional training, while achieving orders of magnitude improvement in computational efficiency over traditional schemes.
- We present comprehensive experiments demonstrating that our approach not only improves forecast accuracy across meteorological variables and radiation fluxes but also significantly enhances physical consistency, as validated through physical conservation evaluations.

2. Related work

2.1. Video Generation and Prediction Models

Video Generation research has evolved rapidly, focusing on generating coherent spatiotemporal sequences for tasks ranging from human motion forecasting to natural scene evolution. Recent advances in diffusion-based models [4, 6, 19, 22, 34, 43, 51, 54, 60] have demonstrated impressive results in capturing visual coherence and perceptual quality. However, these approaches face fundamental limitations when applied to atmospheric science due to weather data’s unique characteristics: high-resolution multi-variable structure creates prohibitive memory requirements, while the physical interdependence among diverse atmospheric variables necessitates expensive retraining of representation models (like VAEs [24, 49]) whenever variable combinations change.

2.2. Weather Forecast

To address these challenges, weather forecasting develops specialized Spatiotemporal architectures, with recent advances in graph-based frameworks [25, 37] and transformer-based approaches [2, 9–11, 26, 35, 57] demonstrating significant improvements. FuXi [10], a state-of-the-art model in this domain, prioritizes forecast skill against ground truth observations rather than perceptual quality. Despite their effectiveness, these weather-specific models remain fundamentally physics-agnostic, relying on data correlations without incorporating the physical laws governing atmospheric dynamics—raising concerns about their reliability under extreme conditions or extended prediction horizons.

2.3. Physics-Guided Prediction

In the broader field of video generation, several approaches have implicitly captured physical dynamics. Video generation models like WorldDreamer [52], Sora [6] and OpenSora [60] learn to generate physically plausible content through large-scale training on multimodal data, while specialized models such as DrivingFusion [28] and PSLG [27] incorporate domain knowledge to enhance physical consistency in specific scenarios. However, studies [21] reveal that current models struggle to abstract universal physical principles from data alone.

For weather forecasting, previous studies [30, 50, 55] have primarily focused on incorporating primitive equations that describe atmospheric motion using ODE solvers [3, 41]. These approaches, however, lack explicit constraints on physical processes like radiative transfer. In operational NWP models, physical processes are represented through parameterization schemes [44], which approximate complex atmospheric interactions based on simplified physical assumptions. Recent work has explored using deep learning

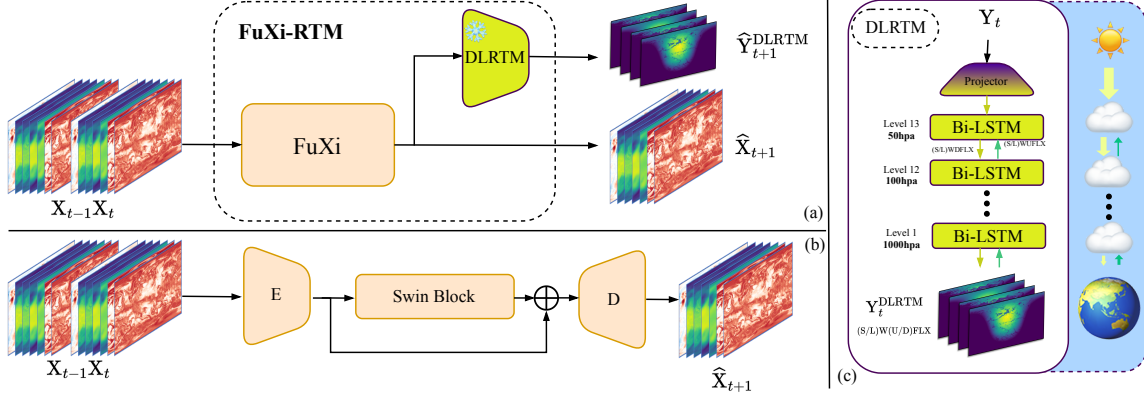


Figure 1. **The overall structure of our method.** (a) Schematic of FuXi-RTM. (b) Architecture of FuXi-base. (c) Schematic of the deep learning-based radiative transfer model (DLRTM) utilizing a bidirectional long short-term memory (Bi-LSTM) architecture.

to emulate these parameterization schemes [17, 39, 56, 58, 59, 61, 62], yet integrating such surrogate models with deep learning-based weather forecasting remains an open challenge.

2.4. Radiative Transfer Modeling

Atmospheric radiative transfer calculations traditionally rely on line-by-line radiative transfer models (LBLRTM) [12, 13], though their computational costs make them impractical for operational forecasting. Instead, radiation parameterization schemes [40] are employed in NWP models, using simplified physics to represent radiation-atmosphere interactions [38]. In the deep learning context, Yao et al. [56] found that bidirectional LSTM architectures achieve superior accuracy for radiation modeling, providing a foundation for our DLRTM implementation.

3. Method

3.1. Data

ERA5 [18], the fifth iteration of the ECMWF reanalysis dataset, represents the most comprehensive and accurate global reanalysis archive available and is widely used in developing deep learning-based weather forecasting framework. For this study, we employ the 6-hourly ERA5 dataset with a 0.25° resolution (equivalent to 721×1440 latitude-longitude grid points). Unlike conventional video data with pixels in height (H) and width (W) dimensions and sequential frames, weather data uses grid points representing specific geographic locations defined by latitude and longitude coordinates. Each grid point stores meteorological variables analogous to how pixels store color intensities. Similarly, forecast lead times at 6-hour intervals replace frame indices for temporal progression. The input to FuXi-RTM consists of four-dimensional cubes $\mathbf{X}_t, \mathbf{X}_{t-1} \in \mathbb{R}^{1 \times C \times H \times W}$, where

$C = 83$ represents the number of meteorological variables, and $H = 721, W = 1440$ correspond to the spatial dimensions of the global grid. Each such "frame" captures the complete global atmospheric state at time step t , analogous to frames in video generation tasks.

Specifically, The FuXi-RTM model forecasts 83 variables, including 5 upper-air atmospheric variables across 13 pressure levels (50, 100, 150, 200, 250, 300, 400, 500, 600, 700, 850, 925, and 1000 hPa), and 18 surface variables. The upper-air atmospheric variables are geopotential (Z), temperature (T), fraction of cloud cover (CC), specific cloud liquid water content (CLWC), and specific humidity (Q). The surface variables are 2-meter temperature (T2M), 2-meter dewpoint temperature (D2M), 10-meter u wind component (U10M), 10-meter v wind component (V10M), 100-meter u wind component (U100M), 100-meter v wind component (V100M), mean sea-level pressure (MSL), surface pressure (SP), low cloud cover (LCC), medium cloud cover (MCC), high cloud cover (HCC), total cloud cover (TCC), surface albedo (FAL), surface net solar radiation (SSR), surface solar radiation downwards (SSRD), total sky direct solar radiation at surface (FDIR), top net thermal radiation (TTR), and total precipitation (TP). A complete list of these variables and their abbreviations is provided in Tab. 1 of the supplementary material. Notably, the u and v components of wind across pressure levels are excluded from the model, as they are not necessary for examining radiative constraints.

3.2. FuXi model

Before introducing our proposed FuXi-RTM, we briefly describe the baseline FuXi model architecture. The FuXi model takes as input two consecutive global atmospheric state data cubes $\mathbf{X}_t, \mathbf{X}_{t-1}$ and aims to predict the atmospheric state at the next time step $\hat{\mathbf{X}}_{t+1}$. As an autore-

gressive model, FuXi extends forecast lead times by recursively feeding its outputs back as inputs. This recursive process continues until reaching the desired forecast horizon (typically 10 days in our experiments). As shown in Fig.1, the architecture follows an encoder-processor-decoder paradigm. First, 3D convolutional layers encode the input data into feature representations. These features are then processed through a backbone network consisting of Swin Transformer V2 blocks[31], which effectively capture long-range dependencies and multi-scale features crucial for global weather systems. Finally, 3D transposed convolutions decode the processed features to generate the predicted atmospheric state. Skip connections between the encoder and decoder preserve detailed information throughout the network. In our implementation, the baseline model FuXi-base employs 30 Swin Transformer V2 blocks, reducing the parameter count from approximately 1.5 billion in the original FuXi to 1.1 billion while maintaining strong predictive capabilities.

3.3. Deep learning based RRTMG model

Model designs. Inspired by Yao et al [56], we develop a DLRTM based on the Bi-LSTM model. Our DLRTM consists of three repeated Bi-LSTM layers, each containing a forward and a backward LSTM layer with feature dimensions of 96 and 128, respectively. As illustrated in Fig.1, the DLRTM processes data $\mathbf{Y}_t \in \mathbb{R}^{1 \times 71 \times H \times W}$ selected from $\mathbf{X}_t$, across 13 pressure levels in an atmospheric column, ranging from 50 hPa at the top of the atmosphere (TOA) to 1000 hPa near the surface. Note that, DLRTM operates independently on each grid point (i, j) , processing one vertical atmospheric column at a time. This column-wise approach allows for efficient parallel computation across the global grid. At each level, the model processes 11 input variables (listed in Tab.2 in the supplementary material), including 5 upper-air variables that vary with pressure levels and 6 single-level variables (e.g. solar zenith angles and land-sea mask) that remain constant across levels. The DLRTM generates $\mathbf{Y}_t^{\text{DLRTM}} \in \mathbb{R}^{1 \times (4 \times 13) \times H \times W}$, including four output variables at each layer: shortwave upward fluxes (SWUFLX), shortwave downward fluxes (SWDFLX), longwave upward fluxes (LWUFLX), and longwave downward fluxes (LWDFLX).

A critical challenge in atmospheric modeling arises from Earth's varying topography. For example, consider a mountainous region like Tibet, where surface pressure might be approximately 600 hPa. In such locations, the standard pressure levels of 700, 850, 925, and 1000 hPa would fall below the actual surface—these non-physical locations are termed "ghost levels" ($P_{\text{level}} > P_{\text{surface}}$). To address this,

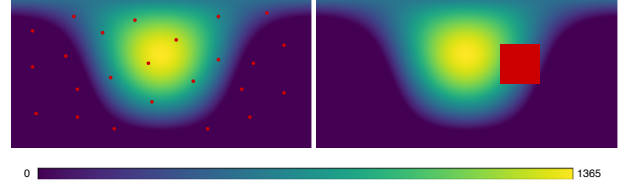


Figure 2. **Visualized sampling strategy.** Left: Global random sampling. Right: The proposed SRC sampling. The dark regions (0 values) indicating areas without direct sunlight, which are excluded as potential SRC center points. Red points or box represents the sampling locations.

DLRTM implements dynamic masking:

$$M(i, j, k) = \begin{cases} 1 & \text{if } P_{\text{level}}(k) \leq P_{\text{surface}}(i, j) \\ 0 & \text{if } P_{\text{level}}(k) > P_{\text{surface}}(i, j) \end{cases} \quad (1)$$

where (i, j) represents a grid point location and k indexes the pressure level. During forward propagation, DLRTM uses this mask to exclude non-physical levels from calculations, ensuring that radiative fluxes are only computed for atmospheric layers that physically exist.

Loss designs. To train the DLRTM model effectively, we employ a mean squared error loss function:

$$L_{\text{reg}} = \frac{1}{H \times W \times R} \sum_{i=1}^H \sum_{j=1}^W M(i, j) \times \left(\sum_{r=1}^R \left((\hat{\mathbf{Y}}_{r,i,j}^{\text{DLRTM}} - \mathbf{Y}_{r,i,j}^{\text{DLRTM}})^2 + \epsilon \right) \right) \quad (2)$$

where $\hat{\mathbf{Y}}^{\text{DLRTM}}$ represents the radiative fluxes predicted by DLRTM based on atmospheric state inputs, and $\mathbf{Y}^{\text{DLRTM}}$ denotes the ground truth radiative fluxes generated by the RRTMG model. The index r runs over the four radiative flux variables (SWUFLX, SWDFLX, LWUFLX, LWDFLX) across all 13 pressure levels.

3.4. FuXi-RTM

Model designs. FuXi-RTM integrates data-driven weather forecasts with physics-aware constraints by combining two core components: a trainable primary forecasting model (FuXi [10]) and a pre-trained DLRTM, as illustrated in Fig.1. Given sequential atmospheric states $\mathbf{X}_{t-1}$ and $\mathbf{X}_t$, the primary model predicts the next state $\hat{\mathbf{X}}_{t+1}$, while DLRTM serves as a differentiable physics regularizer that enforces radiative transfer consistency and outputs the corresponding radiative flux $\hat{\mathbf{Y}}_{t+1}^{\text{DLRTM}}$. During FuXi-RTM's training, DLRTM's parameters remain frozen while it processes outputs from the primary model to generate radiative fluxes. These fluxes are then incorporated into the loss function, creating a physics-guided training signal that enhances forecast accuracy while maintaining computational efficiency. Our implementation focuses specifically on surface-level shortwave upward and downward fluxes (SWUFLX and SWDFLX), as these have shown the greatest influence on forecast improvement. During training, we propose a novel sunlit region-centered (SRC) sampling strategy for

loss calculation, which dynamically selects points from a 250×250 grid centered on randomly chosen sunlit locations, as illustrated in Fig.2. This spatial sampling strategy captures meaningful radiative interactions in areas where solar radiation is most impactful.

Loss designs. The training process for FuXi-RTM balances direct forecasting accuracy with physical consistency through a composite loss function. To minimize discrepancies between model outputs and ground truth, we employ a latitude-weighted Charbonnier L1 loss [7]:

$$L_{forecast} = \frac{1}{C \times H \times W} \sum_{c=1}^C \sum_{i=1}^H \sum_{j=1}^W \alpha_i (\sqrt{(\hat{\mathbf{X}}_{c,i,j} - \mathbf{X}_{c,i,j})^2 + \epsilon^2}) \quad (3)$$

where $\hat{X}$ and X represent FuXi forecast values and the ground truth respectively. The term $\alpha_i = H \times \frac{\cos \Phi_i}{\sum_{i=1}^H \cos \Phi_i}$ is a latitude-specific weighting factor that accounts for the varying grid cell areas at different latitudes, ensuring proper global representation.

Additionally, to enforce radiative transfer consistency, we incorporate a physics regularization term:

$$L_{reg} = \frac{1}{R' \times H' \times W'} \sum_{r=1}^{R'} \sum_{i=1}^{H'} \sum_{j=1}^{W'} \alpha_i (\lambda \sqrt{(\hat{\mathbf{Y}}_{r,i,j}^{DLRTM} - \mathbf{Y}_{r,i,j}^{DLRTM})^2 + \epsilon^2}) \quad (4)$$

where $\mathbf{Y}^{DLRTM}$ and $\hat{\mathbf{Y}}^{DLRTM}$ denote radiative fluxes generated using ERA5 and FuXi forecasts respectively. Here, H' and W' represent the dimensions of the 250×250 grid sampled from sunlit regions, and R' represents the surface-level shortwave fluxes. The parameter $\lambda = 10^{-3}$ balances the contribution of radiative physics constraints against direct meteorological variable prediction.

The combined loss function $L_{total} = L_{forecast} + L_{reg}$ drives the model to simultaneously optimize for forecast accuracy and physical consistency, resulting in predictions that better respect the underlying physical processes governing atmospheric dynamics.

4. Experiment

We evaluate our approach through experiments measuring forecast accuracy and physical consistency. This section presents our experimental configuration, performance comparisons with the baseline model, ablation studies to validate design choices, and analysis of physical conservation properties.

4.1. Experimental Setup

4.1.1. Dataset Configuration

We conduct experiments using the ERA5, which represents the state-of-the-art in global atmospheric reanalysis. The model training follows a rigorous temporal split: we use 15 years of data (2002-2016) for training, 1 year (2017) for

validation, and a comprehensive 5-year period (2018-2022) for testing. For the test period, forecasts are initialized twice daily at 00:00 UTC and 12:00 UTC, generating predictions at 6-hour intervals up to a 10-day horizon. This extensive temporal coverage ensures robust evaluation across diverse atmospheric conditions and seasonal variations.

4.1.2. Implementation Details

Since ERA5 does not include radiative fluxes, we first use the RRTMG to generate two years (2017-2018) of radiative fluxes based on other ERA5 variables. Using this data, we train our DLRTM surrogate model on NVIDIA A100 GPUs for 5 epochs with a batch size of 41,529. The DLRTM employs the PyTorch framework [36] with the AdamW [23, 32] optimizer, configured with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and a cosine annealing learning rate schedule [33] that decays from 10^{-3} to 10^{-8} . The training set consists of one year of 6-hourly data from 2017, while the validation set includes one year of 6-hourly data from 2018.

For the main FuXi-RTM model, we train on a cluster of 4 NVIDIA H100 GPUs for 60,000 iterations with a batch size of 1 per GPU, requiring approximately 81 hours to complete. We use the AdamW optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.95$, an initial learning rate of 2.5×10^{-4} , and a weight decay coefficient of 0.1.

4.1.3. Metrics

Unlike conventional video generation tasks that emphasize diversity in predicted futures, weather forecasting has a unique ground truth for each future state, making deterministic evaluation possible and necessary. Following standard practices in operational weather forecast evaluation [10, 15], we use latitude-weighted root mean square error (RMSE) and normalized RMSE ($nRMSE = \frac{RMSE - RMSE_{baseline}}{RMSE_{baseline}} \times 100\%$) as our primary evaluation metrics to account for the varying grid cell sizes across different latitudes.

Additionally, to evaluate the physical consistency of radiation predictions, we introduce instantaneous surface net solar radiation downwards (ISSRD) as a specific metric. ISSRD represents the effective solar energy received at Earth's surface and is calculated as:

$$ISSRD_{i,j} = \frac{SWDFLX_{i,j}^{surface} - SWUFLX_{i,j}^{surface}}{1 - FAL_{i,j}} \quad (5)$$

where $SWDFLX_{i,j}^{surface}$ and $SWUFLX_{i,j}^{surface}$ represent downward and upward shortwave radiative fluxes at surface level, respectively. ISSRD is crucial for weather forecasting, renewable energy planning, and precision agriculture, directly impacting temperature patterns and cloud formation.

4.1.4. Comparative Framework

We compare our proposed FuXi-RTM against FuXi-base, a pure data-driven weather forecasting model (FuXi) with-

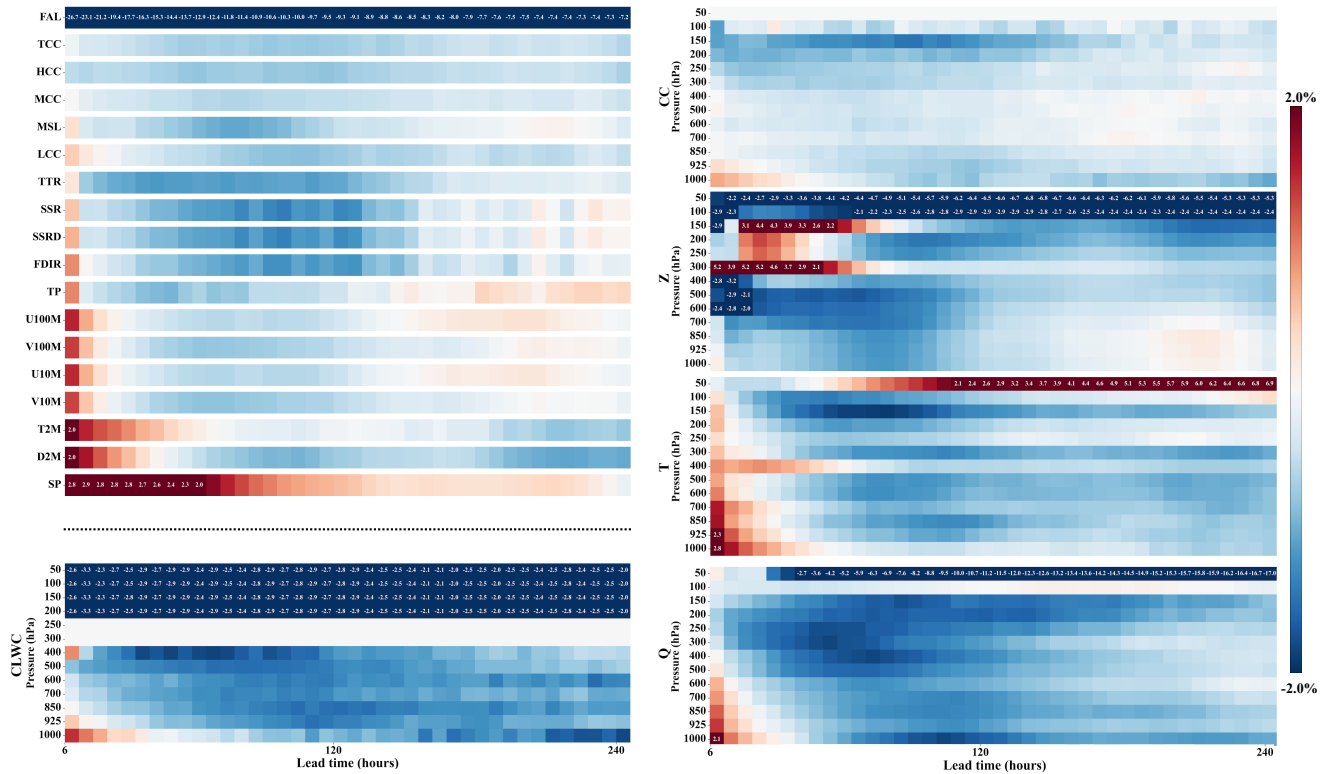


Figure 3. **Scorecard of nRMSE differences in globally-averaged, latitude-weighted RMSE between FuXi-base and FuXi-RTM.** Each subplot corresponds to one of the FuXi direct output variable: 18 surface variables and 5 upper-air variables. For upper-air variables, the rows of each heatmap represent 13 pressure levels. The columns correspond to 40 forecast lead times at 6-hour intervals, spanning from 6 hours to 10 days. The color of each cell indicates the nRMSE differences, with **blue** denoting negative values (FuXi-RTM outperforms FuXi-base) and red indicating positive values (FuXi-base outperforms FuXi-RTM). The nRMSE difference ranges between -2 and 2, with numeric values overlaid on cells that fall outside this range.

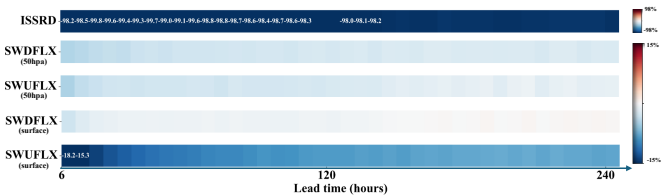


Figure 4. **Scorecard of nRMSE differences in globally-averaged, latitude-weighted RTM RMSE between FuXi-base and FuXi-RTM.** Each subplot corresponds to one of the DLRTM output variable: ISSRD, SWDFLX and SWUFLX at the surface level and 50 hPa. **Blue** denoting negative values (FuXi-RTM outperforms FuXi-base).

out physical constraints. Additionally, we systematically evaluate FuXi-RTM through controlled ablation studies. We create five model variants, each modifying exactly one aspect of the FuXi-RTM configuration while keeping all other components fixed. These ablations investigate two key design dimensions: (1) **sampling strategy** - comparing globally random sampling (FuXi-RTM-Random) versus

our SRC sampling strategy. (2) **radiative flux optimization targets** - evaluating four alternatives to surface-level SW fluxes: both SW and LW fluxes across all pressure levels (FuXi-RTM-13level), SW-only fluxes across all levels (FuXi-RTM-13levelSW), net surface SW radiation (FuXi-RTM-GSW), and instantaneous surface net solar radiation (FuXi-RTM-ISSRD).

4.2. Main Results

Fig.3 presents the normalized differences in globally-averaged, latitude-weighted RMSE between FuXi-base and FuXi-RTM across all 3320 variables and lead time combinations. Results demonstrate that the hybrid physics-guided architecture, which enforces radiation constraints, significantly improves forecast accuracy. Specifically, FuXi-RTM outperforms FuXi-base in 88.51% of all 3320 combinations. For variables highly related with radiation, such as CC and Q, the percentage of FuXi-RTM's superior performance increases to 95.38% and 93.46%, respectively. Notably, for CLWC, improvements exceed 2% in nRMSE differences at pressure levels above 200 hPa. Given the crit-

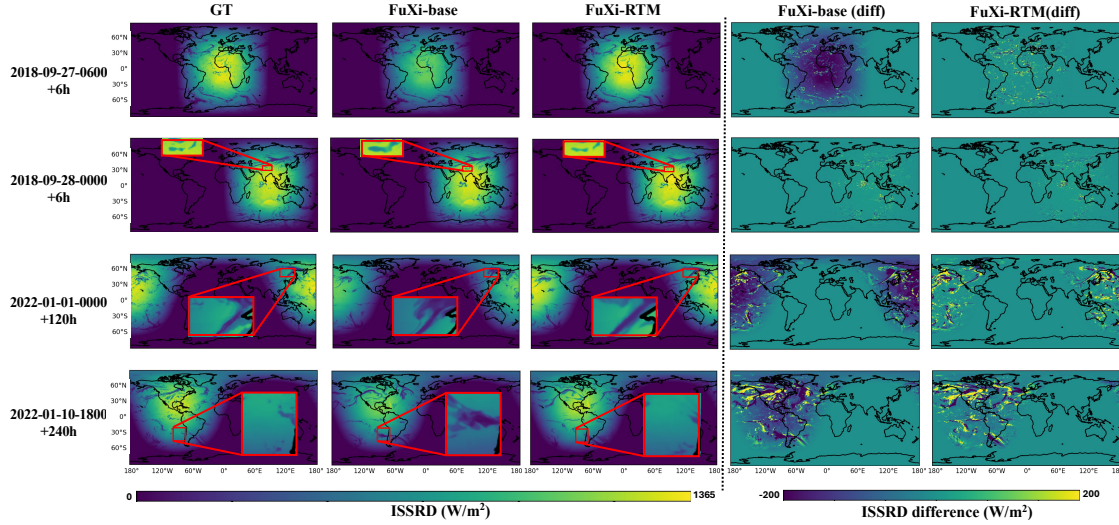


Figure 5. **Snapshot examples of ISSRD.** From left to right: GT (ground truth), FuXi-base (model predictions), FuXi-RTM (model predictions), FuXi-base (diff) (difference between FuXi-base and GT), and FuXi-RTM (diff) (difference between FuXi-RTM and GT). The forecasts are initialized at four different times: 06 UTC on September 27, 2018; 00 UTC on September 28, 2018; 00 UTC on January 1, 2022; and 18 UTC on January 10, 2022. The corresponding forecast horizons are 6, 6, 120, and 240 hours, respectively.

ical role of clouds and radiation in regulating Earth’s energy balance, these enhancements underscore FuXi-RTM’s potential to advance predictive capabilities in weather and climate forecasting [29].

However, challenges in training FuXi with DLRTM lead to initial underperformance for certain variables. For instance, FuXi-RTM trails FuXi-base in predicting Q at 1000 hPa for up to 1.25 days before surpassing it. The most significant improvements are observed for FAL, with nRMSE differences exceeding 7% throughout the 10-day forecasts. This is likely due to FAL’s strong coupling with surface SW fluxes, which FuXi-RTM explicitly supervises. Accurate modeling of FAL is crucial, as it governs the partitioning of Earth’s energy between absorption and reflection [46], directly impacting weather dynamics and long-term climate trends [14, 16].

Radiation Prediction Evaluation. Fig.4 illustrates the nRMSE differences for DLRTM output variables, including ISSRD as well as SWDFLX and SWUFLX at the surface level and TOA (50 hPa in this study). Our evaluation reveals that FuXi-RTM achieves modest improvements in predicting SWDFLX and SWUFLX at 50 hPa despite only explicitly constraining surface-level shortwave radiation. Furthermore, FuXi-RTM demonstrates substantially smaller RMSE for ISSRD predictions with improvements approaching 100%, while the performance gains for SWDFLX and SWUFLX remain comparatively modest. Analysis of the ISSRD computation formula 5 reveals that this significant improvement is primarily attributable to the higher accuracy of the Fraction of FAL term, which serves as the denominator in the calculation, thereby amplifying

the relative enhancement in prediction quality.

To gain deeper insights into ISSRD performance, Fig.5 compares global spatial distribution of ground truth (GT) with the predicted ISSRD from FuXi-base and FuXi-RTM for four randomly selected forecast initialization dates and lead time. The difference maps between predictions and GT, shown in the right two columns, clearly reveal that FuXi-RTM achieves smaller errors compared to FuXi-base. This improvement is particularly evident in sunlit regions, where FuXi-base exhibits widespread negative ISSRD biases. These spatial patterns align with the nRMSE differences quantified in our earlier analyses, further validating the enhanced performance of physics-constrained approach. **DLRTM Evaluation.** Beyond accuracy metrics (see supplementary material), we evaluate computational efficiency. DLRTM achieves orders of magnitude speedup over the traditional RRTMG model through parallel batch processing, reducing computation time from 22 minutes (8 CPUs) to approximately 3 seconds (1 H100 GPU) for global (721×1440) grid points.

4.3. Ablation Study

Tab.1 compares the RMSE of FuXi-base, FuXi-RTM, and its variants trained under different configurations for key meteorological variables. FuXi-RTM consistently outperforms both its variants and FuXi-base across most variables. Notably, all FuXi-RTM variants demonstrate superior overall performance compared to FuXi-base.

Our ablation studies reveal several key insights. Models employing our sunlit region-centered sampling significantly outperformed random global sampling (FuXi-

Table 1. **Ablation studies on different settings.** Performance of FuXi-base, FuXi-RTM and its variants trained under different configurations for critical and classical variables, such as 50-hPa specific humidity (Q50), 500-hPa specific cloud liquid water content (CLWC500). The best and second-best performing variants are highlighted in **bold** and underlined, respectively. Units for each variable are provided in the last row.

Model	Q50	Q500	CLWC500	CC150	TCC	TTR	TP	FAL
FuXi-RTM-Random	0.1857	0.7459	0.0226	0.1725	0.3193	159.2536	2.4327	0.02515
FuXi-RTM-13level	0.1697	<u>0.7409</u>	0.0226	0.1732	0.3184	159.2348	2.4418	0.02313
FuXi-RTM-13levelSW	0.1762	0.7409	<u>0.0224</u>	<u>0.1714</u>	<u>0.3183</u>	<u>158.4191</u>	2.4200	0.02335
FuXi-RTM-GSW	<u>0.1668</u>	0.7567	0.0226	0.1735	0.3207	160.9495	2.4564	0.02263
FuXi-RTM-ISSRD	0.1741	0.7441	0.0226	0.1719	0.3191	159.1379	2.4299	0.02379
FuXi-RTM	0.1546	0.7388	0.0223	0.1705	0.3179	158.1290	2.4092	<u>0.02300</u>
FuXi-base	0.1735	0.7453	0.0226	0.1720	0.3196	159.1416	2.4127	0.02553
Unit	10^{-3} g/kg	g/kg	10^{-2} g/kg	10^{-2} (0 - 1)	(0 - 1)	J/m ²	m	(0 - 1)

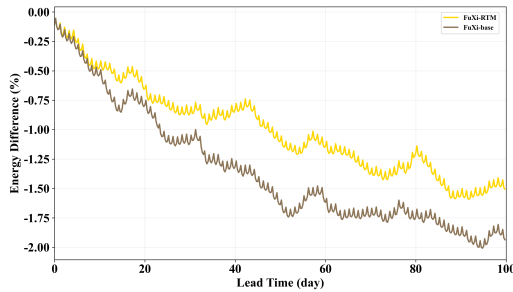


Figure 6. **Verification of global total atmospheric energy conservation.** Normalized differences in global total atmospheric energy of FuXi-RTM (yellow) and FuXi-base (brown) forecasts relative to ERA5-derived reference values. Values represent averages over forecasts initialized at 00 UTC on five random dates in 2018. The horizontal axis represents the forecast duration, and the vertical axis represents the percentage of energy loss. The lower the value on the vertical axis, the more severe the energy loss.

RTM-Random), demonstrating that spatially coherent gradient computation enhances feature learning by preserving local contextual relationships while reducing interference from irrelevant background signals. For variants optimizing all pressure levels (FuXi-RTM-13level and FuXi-RTM-13levelSW), we observed diminished performance likely due to information redundancy in the radiative transfer process—surface radiation constraints inherently capture vertical atmospheric interactions through backpropagation, making explicit multi-level supervision unnecessary and potentially counterproductive. The FuXi-RTM-GSW and FuXi-RTM-ISSRD variants, which optimize derived radiation metrics rather than direct fluxes, underperformed compared to FuXi-RTM. This suggests that constraining fundamental radiative components (surface SW fluxes) provides the model with more comprehensive physical information than derived quantities that inherently contain less information about the underlying radiative processes.

4.4. Global total atmospheric energy conservation

To further validate the physical consistency of our approach, we examine the conservation of global total atmospheric energy (detailed formulation in supplementary material). As shown in Fig.6, FuXi-RTM demonstrates superior energy conservation compared to FuXi-base, with the difference becoming particularly pronounced beyond 10-day forecasts. These enhanced conservation properties directly result from our radiative constraints, confirming FuXi-RTM’s improved physical fidelity, which becomes increasingly important for long-term predictions.

5. Conclusion

FuXi-RTM is a hybrid physics-guided deep learning model that innovatively enforces radiation constraints by employing a deep learning surrogate for radiation parameterization. To the best of our knowledge, FuXi-RTM is the first deep learning-based weather forecasting model to explicitly model physical processes, as opposed to integrating a dynamical core with a NN serving as a holistic parameterization. Overall, FuXi-RTM demonstrates that incorporating radiation constraints not only enhances forecasts for radiation and clouds but also improves predictions for conventional meteorological variables. Several potential avenues exist for further improving FuXi-RTM. First, the model currently excludes the u and v wind components at all pressure levels. Scaling up to include the uv wind and additional components could further improve the model’s performance, though this would necessitate balancing the trade-off with increased computational resource requirements. Second, the hybrid physics-guided architecture of FuXi-RTM is adaptable to the incorporation of additional physical processes, such as convection, PBL, and cloud microphysics, into deep learning-based weather forecasting framework. By addressing the limitations, this hybrid architecture paves the way for next-generation weather forecasting systems that are accurate, efficient, and trustworthy.

References

- [1] Peter Bauer, Alan Thorpe, and Gilbert Brunet. The quiet revolution of numerical weather prediction. *Nature*, 525(7567): 47–55, 2015. 1
- [2] Kaifeng Bi et al. Accurate medium-range global weather forecasting with 3d neural networks. *Nature*, 2023. 2
- [3] B N Biswas, Somnath Chatterjee, SP Mukherjee, and Subhradeep Pal. A discussion on euler method: A review. *Electronic Journal of Mathematical Analysis and Applications*, 1(2):2090–2792, 2013. 2
- [4] Andreas Blattmann, Robin Rombach, Huan Ling, Tim Dockhorn, Seung Wook Kim, Sanja Fidler, and Karsten Kreis. Align your latents: High-resolution video synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 22563–22575, 2023. 2
- [5] Zied Ben Bouallègue et al. The rise of data-driven weather forecasting: A first statistical assessment of machine learning-based weather forecasts in an operational-like context. *Bulletin of the American Meteorological Society*, pages 1520–0477, 2024. 2
- [6] Tim Brooks, Bill Peebles, Connor Holmes, Will DePue, Yufei Guo, Li Jing, David Schnurr, Joe Taylor, Troy Luhman, Eric Luhman, Clarence Ng, Ricky Wang, and Aditya Ramesh. Video generation models as world simulators, 2024. 2
- [7] P. Charbonnier, L. Blanc-Feraud, G. Aubert, and M. Barlaud. Two deterministic half-quadratic regularization algorithms for computed imaging. In *Proceedings of 1st International Conference on Image Processing*, pages 168–172 vol.2, 1994. 5
- [8] J. G. Charney, R. Fjörtoft, and J. von Neumann. Numerical integration of the barotropic vorticity equation. *Tellus A: Dynamic Meteorology and Oceanography*, 1950. 1
- [9] Kang Chen et al. Fengwu: Pushing the skillful global medium-range weather forecast beyond 10 days lead, 2023. Preprint at <https://arxiv.org/abs/2304.02948>. 2
- [10] Lei Chen et al. Fuxi: A cascade machine learning forecasting system for 15-day global weather forecast. *npj Climate and Atmospheric Science*, pages 1–11, 2023. 2, 4, 5
- [11] Lei Chen et al. A machine learning model that outperforms conventional global subseasonal forecast models. *Nature Communications*, 15(1):6425, 2024. 2
- [12] SA Clough, MW Shephard, EJ Mlawer, JS Delamere, MJ Iacono, K Cady-Pereira, S Boukabara, and PD Brown. Atmospheric radiative transfer modeling: A summary of the aer codes. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 91(2):233–244, 2005. 3
- [13] Shepard A Clough, Michael J Iacono, and Jean-Luc Moncet. Line-by-line calculations of atmospheric fluxes and cooling rates: Application to water vapor. *Journal of Geophysical Research: Atmospheres*, 97(D14):15761–15785, 1992. 3
- [14] Robert E. Dickinson. Land surface processes and climate—surface albedos and energy balance. In *Theory of Climate*, pages 305–353. Elsevier, 1983. 7
- [15] Thomas Haiden et al. Evaluation of ECMWF forecasts, including the 2021 upgrade, 2021. 2, 5
- [16] Alex Hall. The role of surface albedo feedback in climate. *Journal of Climate*, 17(7):1550 – 1568, 2004. 7
- [17] Yilun Han, Guang J. Zhang, Xiaomeng Huang, and Yong Wang. A moist physics parameterization based on deep learning. *Journal of Advances in Modeling Earth Systems*, 12(9):e2020MS002076, 2020. e2020MS002076 2020MS002076. 3
- [18] Hans Hersbach et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.*, 146(730):1999–2049, 2020. 3
- [19] Jonathan Ho, Tim Salimans, Alexey Gritsenko, William Chan, Mohammad Norouzi, and David J Fleet. Video diffusion models. *Advances in Neural Information Processing Systems*, 35:8633–8646, 2022. 2
- [20] Robin Hogan et al. Radiation in numerical weather prediction, 2017. 1
- [21] Bingyi Kang, Yang Yue, Rui Lu, Zhijie Lin, Yang Zhao, Kaixin Wang, Gao Huang, and Jiashi Feng. How far is video generation from world model: A physical law perspective. *arXiv preprint arXiv:2411.02385*, 2024. 2
- [22] Levon Khachatryan, Andranik Movsisyan, Vahram Tadevosyan, Roberto Henschel, Zhangyang Wang, Shant Navasardyan, and Humphrey Shi. Text2video-zero: Text-to-image diffusion models are zero-shot video generators. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15954–15964, 2023. 2
- [23] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017. Preprint at <https://arxiv.org/abs/1412.6980>. 5
- [24] Diederik P Kingma, Max Welling, et al. Auto-encoding variational bayes, 2013. 2
- [25] Remi Lam et al. Learning skillful medium-range global weather forecasting. *Science*, 2023. 2
- [26] Simon Lang et al. Aifs – ecmwf’s data-driven forecasting system, 2024. Preprint at <https://arxiv.org/abs/2406.01465>. 2
- [27] Jianan Li, Tao Huang, Qingxu Zhu, and Tien-Tsin Wong. Physics-based scene layout generation from human motion. In *ACM SIGGRAPH 2024 Conference Papers*, pages 1–10, 2024. 2
- [28] Xiaofan Li, Yifu Zhang, and Xiaoqing Ye. Drivingdiffusion: Layout-guided multi-view driving scenarios video generation with latent diffusion model. In *European Conference on Computer Vision*, pages 469–485. Springer, 2024. 2
- [29] Kuo-Nan Liou. Influence of cirrus clouds on weather and climate processes: A global perspective. *Monthly Weather Review*, 114(6):1167 – 1199, 1986. 7
- [30] Peiyuan Liu, Tian Zhou, Liang Sun, and Rong Jin. Mitigating time discretization challenges with weatherode: A sandwich physics-driven neural ode for weather forecasting, 2024. Preprint at <https://arxiv.org/abs/2410.06560>. 2
- [31] Ze Liu et al. Swin transformer v2: Scaling up capacity and resolution. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11999–12009, 2022. 4
- [32] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2017. 5

- [33] Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts, 2017. Preprint at <https://arxiv.org/abs/1608.03983>. 5
- [34] Zhengxiong Luo, Dayou Chen, Yingya Zhang, Yan Huang, Liang Wang, Yujun Shen, Deli Zhao, Jingren Zhou, and Tieniu Tan. Videofusion: Decomposed diffusion models for high-quality video generation. *arXiv preprint arXiv:2303.08320*, 2023. 2
- [35] Tung Nguyen et al. Scaling transformer neural networks for skillful and reliable medium-range weather forecasting. *Advances in Neural Information Processing Systems*, 37: 68740–68771, 2025. 2
- [36] Adam Paszke et al. Automatic differentiation in pytorch. In *NIPS 2017 Workshop on Autodiff*, 2017. 5
- [37] Jaideep Pathak et al. Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators. Preprint at <https://arxiv.org/abs/2202.11214>, 2022. 2
- [38] Zhaoxia Pu and Eugenia Kalnay. Numerical weather prediction basics: Models, numerical methods, and data assimilation. *Handbook of hydrometeorological ensemble forecasting*, pages 67–97, 2019. 1, 3
- [39] Stephan Rasp, Michael S. Pritchard, and Pierre Gentine. Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences*, 115(39):9684–9689, 2018. 3
- [40] Bodo Ritter and Jean-Francois Geleyn. A comprehensive radiation scheme for numerical weather prediction models with potential applications in climate simulations. *Monthly Weather Review*, 120(2):303–325, 1992. 1, 3
- [41] Carl Runge. Über die numerische auflösung von differentialgleichungen. *Mathematische Annalen*, 46(2):167–178, 1895. 2
- [42] John Schreck et al. Community research earth digital intelligence twin (credit), 2024. Preprint at <https://arxiv.org/abs/2411.07814>. 2
- [43] Uriel Singer, Adam Polyak, Thomas Hayes, Xi Yin, Jie An, Songyang Zhang, Qiyan Hu, Harry Yang, Oron Ashual, Oran Gafni, et al. Make-a-video: Text-to-video generation without text-video data. *arXiv preprint arXiv:2209.14792*, 2022. 2
- [44] David J Stensrud. Parameterization schemes: keys to understanding numerical weather prediction models, 2009. 1, 2
- [45] Graeme L Stephens, Juilin Li, Martin Wild, Carol Anne Clayson, Norman Loeb, Seiji Kato, Tristan L’ecuyer, Paul W Stackhouse Jr, Matthew Lebsock, and Timothy Andrews. An update on earth’s energy balance in light of the latest global observations. *Nature Geoscience*, 5(10):691–696, 2012. 1
- [46] Graeme L. Stephens, Denis O’Brien, Peter J. Webster, Peter Pilewski, Seiji Kato, and Jui-lin Li. The albedo of earth. *Reviews of Geophysics*, 53(1):141–163, 2015. 7
- [47] Kevin E. Trenberth, John T. Fasullo, and Jeffrey Kiehl. Earth’s global energy budget. *Bulletin of the American Meteorological Society*, 90(3):311–324, 2009. 1
- [48] M. Tuononen, E. J. O’Connor, and V. A. Sinclair. Evaluating solar radiation forecast uncertainty. *Atmospheric Chemistry and Physics*, 19(3):1985–2000, 2019. 2
- [49] Aaron Van Den Oord, Oriol Vinyals, et al. Neural discrete representation learning. *Advances in neural information processing systems*, 30, 2017. 2
- [50] Yogesh Verma, Markus Heinonen, and Vikas Garg. Climode: Climate and weather forecasting with physics-informed neural odes, 2024. Preprint at <https://arxiv.org/abs/2404.10024>. 2
- [51] Wenjing Wang, Huan Yang, Zixi Tuo, Huiguo He, Junchen Zhu, Jianlong Fu, and Jiaying Liu. Videofactory: Swap attention in spatiotemporal diffusions for text-to-video generation, 2023. 2
- [52] Xiaofeng Wang, Zheng Zhu, Guan Huang, Boyuan Wang, Xinze Chen, and Jiwen Lu. Worldreamer: Towards general world models for video generation via predicting masked tokens. *arXiv preprint arXiv:2401.09985*, 2024. 2
- [53] Oliver Watt-Meyer, Gideon Dresdner, Jeremy McGibbon, Spencer K. Clark, Brian Henn, James Duncan, Noah D. Brenowitz, Karthik Kashinath, Michael S. Pritchard, Boris Bonev, Matthew E. Peters, and Christopher S. Bretherton. Ace: A fast, skillful learned global atmospheric model for climate prediction, 2023. Preprint at <https://arxiv.org/abs/2310.02074>. 2
- [54] Jinbo Xing, Menghan Xia, Yuxin Liu, Yuechen Zhang, Yong Zhang, Yingqing He, Hanyuan Liu, Haoxin Chen, Xiaodong Cun, Xintao Wang, et al. Make-your-video: Customized video generation using textual and structural guidance. *IEEE Transactions on Visualization and Computer Graphics*, 2024. 2
- [55] Wanghan Xu, Fenghua Ling, Wenlong Zhang, Tao Han, Hao Chen, Wanli Ouyang, and Lei Bai. Generalizing weather forecast to fine-grained temporal scales via physics-ai hybrid modeling, 2025. Preprint at <https://arxiv.org/abs/2405.13796>. 2
- [56] Yichen Yao, Xiaohui Zhong, Yongjun Zheng, and Zhibin Wang. A physics-incorporated deep learning framework for parameterization of atmospheric radiative transfer. *Journal of Advances in Modeling Earth Systems*, 15(5), 2023. 3, 4
- [57] Shijin Yuan, Guansong Wang, Bin Mu, and Feifan Zhou. Tianxing: A linear complexity transformer model with explicit attention decay for global weather forecasting. *Advances in Atmospheric Sciences*, 42(1):9–25, 2025. 2
- [58] Janni Yuval and Paul A O’Gorman. Stable machine-learning parameterization of subgrid processes for climate modeling at a range of resolutions. *Nature communications*, 11(1): 3295, 2020. 3
- [59] Janni Yuval, Paul A O’Gorman, and Chris N Hill. Use of neural networks for stable, accurate and physically consistent parameterization of subgrid atmospheric processes with good performance at reduced precision. *Geophysical Research Letters*, 48(6), 2021. 3
- [60] Zangwei Zheng, Xiangyu Peng, Tianji Yang, Chenhui Shen, Shenggui Li, Hongxin Liu, Yukun Zhou, Tianyi Li, and Yang You. Open-sora: Democratizing efficient video production for all, 2024. 2
- [61] X. Zhong, Z. Ma, Y. Yao, L. Xu, Y. Wu, and Z. Wang. WRF-ML v1.0: a bridge between WRF v4.3 and machine learning parameterizations and its application to atmospheric radia-

tive transfer. Geoscientific Model Development, 16(1):199–209, 2023. [3](#)

- [62] Xiaohui Zhong, Xing Yu, and Hao Li. Machine learning parameterization of the multi-scale Kain–Fritsch (MSKF) convection scheme and stable simulation coupled in the Weather Research and Forecasting (WRF) model using WRF–ML v1. 0. Geoscientific Model Development, 17(9):3667–3685, 2024. [3](#)