Towards Climate Variable Prediction with Conditioned Spatio-Temporal Normalizing Flows

Christina Winkler
Mila - Quebec AI Institute
christina.winkler@mila.quebec

David Rolnick
McGill University and Mila - Quebec AI Institute
drolnick@mila.quebec

Abstract

This study investigates how conditional normalizing flows can be applied to remote sensing data products in climate science for spatio-temporal prediction. The method is chosen due to its desired properties such as exact likelihood computation, predictive uncertainty estimation and efficient inference and sampling which facilitates faster exploration of climate scenarios. Experimental findings reveal that the conditioned spatio-temporal flow surpasses both deterministic and stochastic baselines in prolonged rollout scenarios. It exhibits stable extrapolation beyond the training time horizon for extended rollout durations. These findings contribute valuable insights to the field of spatio-temporal modeling, with potential applications spanning diverse scientific disciplines.

1 Introduction

Global Climate Models (GCMs) describe physical processes by partial differential equations. These are often computationally expensive due to coarse grid discretization and solving over large spatial and temporal domains. Yet, there's a growing need for precise forecasts of Earth's climate, not just globally but also locally. This is crucial for guiding local responses to extreme weather events and implement adaptive strategies for e.g. water management or agricultural practices. To bridge this disparity, the machine learning community contributed vastly with deep learning based methods for climate variable super resolution to obtain high-resolution simulation data (Wang et al., 2018; Watson et al., 2020; Groenke et al., 2020; Singh et al., 2019; Harder, 2022; Watson et al., 2020; Chaudhuri & Robertson, 2020). Moreover, researchers are drawing inspiration from video prediction techniques utilizing deep learning, including recurrent neural networks (RNNs) and convolutional neural networks (CNNs), as well as stochastic approaches to accelerate the emulation of future states of remote sensing variables (Rasp & Lerch, 2018; Reichstein et al., 2019; Rasp et al., 2018; Ravuri et al., 2021a; Klemmer et al., 2021; Saxena & Cao, 2019; Zhu et al., 2020; Ben-Bouallegue et al., 2023; Ali et al., 2021; Amato et al., 2020; Wang et al., 2022; Ferchichi et al., 2022; Pisl et al., 2023; Drees et al., 2022; Rußwurm & Körner, 2018; Gao et al., 2020; Ji et al., 2023; Ravuri et al., 2021b). This approach can facilitate faster exploration of climate scenarios and potentially enhance the efficiency of GCM-based climate research and decision-making processes. These contributions could allow researchers to incorporate subgrid convective processes in climate models, allowing for more accurate predictions of thunderstorms, heavy rainfall events and tropical cyclones.

In this work, we propose conditional normalizing flows for spatio-temporal modelling of remote sensing data products for climate science. Specifically, we investigate the capabilities of conditional normalizing flows for climate variable forecasting of different temporal and spatial resolutions. Invertible stochastic models allow for exact likelihood computation, predictive uncertainty estimation and fast inference and sampling. Flows for video modelling have yet been explored (Kumar et al., 2019; Zand et al., 2022; Davtyan et al., 2022). Different from the literature, in our approach temporal correlations are learned via conditioning on a compressed memory state produced by a convolutional gated LSTM (Dauphin et al., 2016) of the input frame. Our main contribution can be summarized as follows:

- We introduce a novel method for efficient spatio-temporal modelling with conditioned normalizing flows.
- We employ ST-Flow on ERA5 data, revealing its consistent and stable performance over long rollout durations, as measured by the RMSE metric, when compared against both deterministic and stochastic baselines.

2 METHOD

A Normalizing Flow in continuous space is based on the change of variables formula. Given two spaces of equal dimension $\mathcal{Z}$ and $\mathcal{X}$; a diffeomorphism $f_\phi : \mathcal{X} \to \mathcal{Z}$ and a prior distribution $p_z(\mathbf{z})$ which is easy to sample from, we can model a complicated distribution $p_X(\mathbf{x})$ as

$$p_X(\mathbf{x}) = p_Z(f_\phi(\mathbf{x})) \left| \frac{\partial f_\phi(\mathbf{x})}{\partial \mathbf{x}} \right|. \tag{1}$$

Here $|\partial f_\phi(\mathbf{x})/\partial \mathbf{x}|$ is the determinant of the Jacobian of the transformation between $\mathbf{z}$ and $\mathbf{x}$, evaluated at $\mathbf{x}$, accounting for volume changes in $p_z(\mathbf{z})$ induced by f_ϕ. Typically, the transformation f_ϕ is defined as composition of diffeomorphisms, i.e. $f = f_0 \circ f_1 \circ ... \circ f_K$.

2.1 CONDITIONED SPATIO-TEMPORAL NORMALIZING FLOW

We propose using conditional normalizing flows (Winkler et al., 2019; Lugmayr et al., 2021) for spatio-temporal sequence prediction. Invertible stochastic methods are particularly desirable for the task of climate variable prediction due to their ability to capture predictive uncertainty, ability to generate synthetic data for scenario analysis and handling missing data and imputation. Assume we have a target frame $\mathbf{x}_t$ and a context frame which can be arbitrarily long $(\mathbf{x}_{t-1}, \ldots, \mathbf{x}_T)$ at the t-th time-step such that we have the training tuple $(\mathbf{x}_t; \mathbf{x}_{<t}) \in \{X_t\}$. We learn a complicated distribution $p_\mathbf{x}(\mathbf{x}_t|\mathbf{x}_{<t})$ using a conditional prior $p_\theta(\mathbf{z}_t|\mathbf{z}_{<t})$ and a mapping $f_\phi : \mathcal{X}^t \times \mathcal{X}^{<t} \to \mathcal{Z}^t$, which is bijective in $\mathcal{X}^t$ and $\mathcal{Z}^t$. The likelihood of this model can then be written as:

$$p_\mathbf{x}(\mathbf{x}_t|\mathbf{x}_{<t}) = p_\theta(\mathbf{z}_t|\mathbf{z}_{<t}) \left| \det \frac{\partial \mathbf{z}_t}{\partial \mathbf{x}_t} \right| = p_\theta(f_\phi(\mathbf{x}_t, \mathbf{x}_{<t})|\, \mathbf{z}_{<t}) \left| \det \frac{\partial f_\phi(\mathbf{x}_t, \mathbf{x}_{<t})}{\partial \mathbf{x}_t} \right|. \tag{2}$$

Over the temporal axis, the latent prior can then be factorized as:

$$p_\theta(\mathbf{z}|\mathbf{h}) = \prod_{t=1}^{T} p_\theta(\mathbf{z}_t|\mathbf{h}_t), \tag{3}$$

where $\mathbf{h} = GatedConvLSTM(x_{t-1}, x_{t-2}, ...)$ represents the hidden state at time t. We chose a gated convolutional LSTM due to its ability to efficiently capture long-range dependencies in sequential data while addressing the vanishing gradient problem, which is crucial for tasks requiring robust representation of past states and inputs. Factorization of hierarchical latent variables $\{\mathbf{z}_t^l\}_{l=1}^{L} = f_\phi(\mathbf{x}_t, \mathbf{x}_{<t})$ over the scales can then be written as:

$$p_\theta(\mathbf{z}_t|\mathbf{h}_t) = \prod_{l=1}^{L} p_\theta(\mathbf{z}_t^{(l)}|\mathbf{z}_t^{(>l)}, \mathbf{h}_t), \tag{4}$$

where $\mathbf{z}_t^{(>l)}$ is the set of latent variables at the same time step at higher scales. At each scale l, the split prior $p_\theta(\mathbf{z}_t^{(l)}|\mathbf{z}_t^{(>l)}, \mathbf{h}_t)$ takes on the form of a conditionally factorized Gaussian density:

$$p_\theta(\mathbf{z}_t^{(l)}|\mathbf{z}_t^{(\geq l)}, \mathbf{h}_t) = \mathcal{N}(\mathbf{z}_t^{(l)}; \mu, \sigma)$$
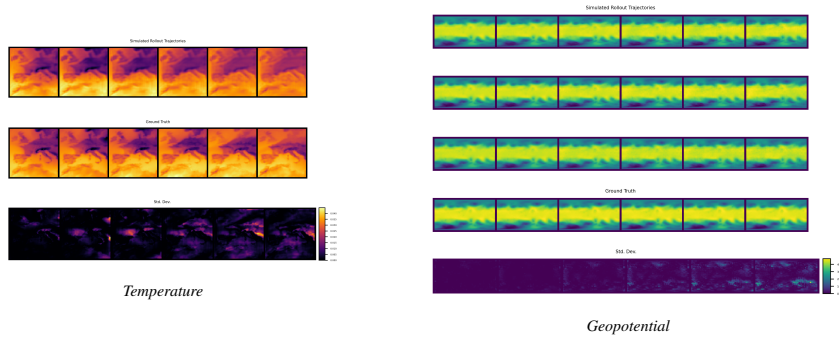$$\text{where } \mu, \log(\sigma) = \text{NN}_\theta(\mathbf{z}_t^{(\geq l)}, \mathbf{h}_t), \tag{5}$$

Figure 2: *Visualization of rollout trajectories starting from the same initial conditions on the ERA5 temperature and geopotential dataset from the conditional normalizing flow model. The last row shows the squared absolute error. For rollouts from other methods, see appendix.*

where NN_θ is a CNN to predicting μ and $\log \sigma$. The learning objective of the full model with memory-based latent dynamics can then be formulated as:

$$\log p_\mathbf{x}(\mathbf{x}|\mathbf{h}) = \log p_\theta(\mathbf{z}|\mathbf{h}) + \sum_{k=1}^{K} \log \left| \det \frac{\partial f_k(\mathbf{z}_k; \mathbf{h})}{\partial f_{k-1}(\mathbf{z}_{k-1}, \mathbf{h})} \right|, \tag{6}$$

where K represents the number of flow steps. An overview of the model architecture is provided in Figure 1.

3 EXPERIMENTS

We conduct experiments on ERA5 remote sensing data products, which are produced by the European Centre for Medium-Range Weather Forecasts (ECMWF). ERA5 data is generated using state-of-the-art numerical weather prediction models, assimilating a wide range of observational data sources, such as satellites, weather stations, and ocean buoys. Figure 3 describes a summary of the datasets used in this study. For preprocessing, we transform values Z by $X = \frac{Z - \min Z}{\max Z - \min Z}$ such that they lie within range [0,1]. We employ a learning rate of 2e-4 using a Step-Wise learning rate scheduler with a decay rate of 0.5 after every 200000th parameter update step. We used the Adam optimizer (Kingma & Ba, 2014) with exponential moving average and coefficients of running averages of gradients and its square are set to $\beta = (0.9, 0.99)$.
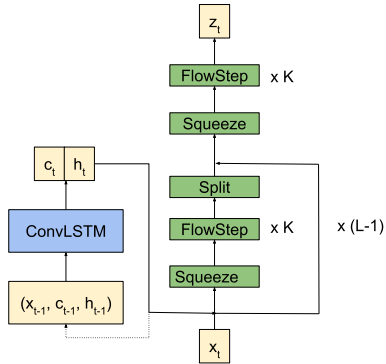

Figure 1: *ST-Flow architecture with K flow steps and L scales. For further explanation, see Appendix A.*

Dataset	Resolution	Time Span	Training/Validation/Test Samples
T2M	0.25°	1979 - 2020	1193 / 340 / 170
500 hPa	5.625°	1979 - 2017	298,043 / 17,532 / 35,064

Figure 3: *Dataset summary with their resolutions, time spans, and training/validation/test samples.*
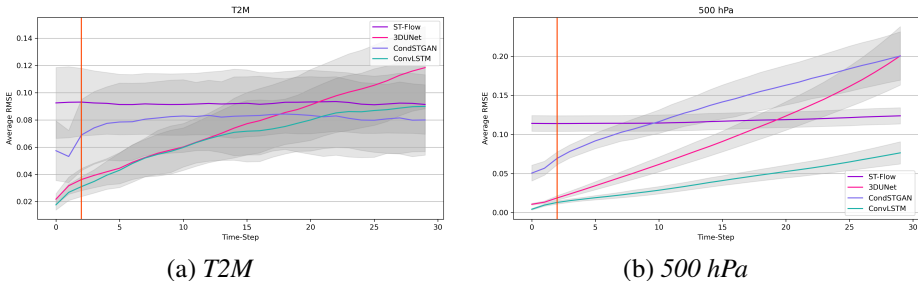
(a) *T2M* (b) *500 hPa*

Figure 4: *RMSE curves over 100 test samples on the ERA5 daily temperature and hourly geopotential dataset computed for different models. The vertical line indicates the length of the context window size during training, which we set to 2.*

3.1 QUANTITATIVE RESULTS

In Figure 4 we present a comparative analysis of the ST-Flow against two deterministic approaches (3DUNet and ConvLSTM), as well as a spatio-temporal conditional GAN serving as stochastic baseline using the RMSE metric. All models were trained with the same parameter budget and on the original input resolutions of the respective datasets for a lead time of 30 time-steps. Examining the results on the hourly geopotential dataset (Fig. 4 a)) reveals an initial advantage for the deterministic methods over the stochastic ones in terms of performance. However, as the rollout trajectory progresses, the ST-Flow model consistently maintains the most stable RMSE scores throughout the entire duration. Conversely, the conditional GAN architecture exhibits the poorest performance among the models evaluated. On the daily temperature dataset (Fig. 4 b)), results indicate that the deterministic methods perform better initially, but decrease in performance for longer rollout durations. The stochastic methods provide more stable results over the whole rollout period. On the hourly dataset, the variability in performance is lower than for the coarser temporally resolved temperature dataset for all models.

The initial advantage for deterministic methods over stochastic ones, such as observed in the hourly geopotential dataset, can be attributed to their ability to effectively capture short-term patterns and dependencies within the data. Deterministic methods excel at directly predicting the next value in a sequence based on the current and past observations, which can be advantageous in scenarios where short-term trends dominate the data. However, as the rollout trajectory progresses, deterministic methods may struggle to maintain accurate predictions over longer horizons due to their inherent limitations in capturing complex, evolving patterns and dependencies. These methods may become increasingly susceptible to errors accumulating over time, leading to a degradation in performance.

The conditional GAN provides superior performance over the ST-Flow on the temperature dataset, but not on the geopotential dataset. This may be attributed to the simpler temporal and spatial dynamics in the temperature dataset as compared to the geopotential dataset.

In contrast, the ST-Flow model, demonstrates greater stability in RMSE scores over the entire duration of the rollout trajectory on both datasets. By leveraging its capacity to capture spatio-temporal long-term relationships, the ST-Flow model can make more robust and consistent predictions over extended horizons compared to deterministic methods.

Overall, the observed trends underscore the importance of considering the modeling capabilities of different approaches, as well as the specific characteristics and dynamics of the dataset, when assessing performance over varying prediction horizons.

3.2 EFFICIENT SPATIAL REPRESENTATION LEARNING FOR CLIMATE VARIABLE FORECASTING WITH CONDITIONAL NORMALIZING FLOWS

In this experiment, we leverage the power of conditional normalizing flows as auto-encoders to efficiently reduce the spatial dimensionality of input frames. Subsequently, we employ simulations based on these condensed representations, providing a streamlined method for handling large-scale climate datasets. The outcomes, depicted in Figure 5, present the RMSE results of the spatio-
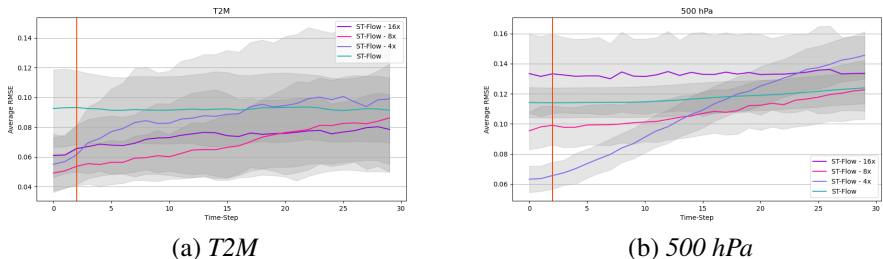
(a) *T2M*   (b) *500 hPa*

Figure 5: *RMSE curves over 100 test samples on the ERA5 daily temperature and hourly geopotential dataset for different input resolutions of the data (orig,4x,8x,16x). The vertical line indicates the length of the context window size during training, which we set to 2.*

temporal normalizing flow across a spectrum of downsampling resolutions (4x, 8x, 16x) applied during the simulation. These RMSE values are computed within the reconstructed original image space, achieved through the training of a conditional normalizing flow as post-processing operation.

Generally, findings from both datasets indicate that for longer rollout durations, employing a coarser spatial representation yields more stable outcomes. The phenomenon of employing a coarser spatial representation leading to more stable outcomes for longer rollout durations can be attributed to several factors. Firstly, by simplifying the system's dynamics through coarsening, the complexity of the simulation is reduced, making it more resilient to small-scale fluctuations or uncertainties that may arise over time. Secondly, the smoothing effect of coarsening, achieved through averaging or aggregating spatial information, helps to dampen out noise or variability at smaller scales, resulting in a smoother and more predictable system behavior. Additionally, the computational efficiency gained from using a coarser spatial representation enables longer simulations to be conducted more efficiently, allowing researchers to study the system's behavior over extended time periods without significant computational overhead.

However, performance in relation to the downsampling factor varies across datasets, likely due to discrepancies in spatial and temporal resolutions. For instance, datasets with higher spatial resolutions such as the geopotential dataset may contain more detailed information about the system's dynamics, making it challenging to effectively condense the data without losing critical features. This is also reflected in the results presented in Figure 5 panel b).

The efficacy of dimensionality reduction on input frames lies in its ability to eliminate redundant information while retaining essential data necessary for subsequent prediction tasks. By reducing the dimensionality of the dataset, the model focuses on capturing the most relevant features, leading to improved efficiency and interpretability. This process helps in discarding extraneous details that may introduce noise or complexity without adding significant value to the predictive task at hand. As a result, the dimensionality reduction enhances the model's ability to generalize patterns and make accurate predictions, even when operating on lower-dimensional representations of the input data.

4   CONCLUSION

Experimentally, we have shown that deterministic methods initially outperform stochastic ones due to their proficiency in capturing short-term patterns, notably in the geopotential dataset. However, their performance declines over longer horizons due to limitations in capturing evolving patterns. Overall, experiments with the ST-Flow show that simplifying the spatial representation of the system aids in stabilizing its behavior over longer durations by reducing complexity, smoothing variability and improving computational efficiency. In summary, while the existing ERA5 datasets may be a great proof of concept, it is important to evaluate the trade-offs between computational complexity, data accuracy, and the specific needs of the real-world modeling task. Also, there is a need for exploring model calibration techniques such as incorporating physics-informed neural networks to ensure that predictions of AI models accurately reflect the true underlying data distribution and account for any systematic biases or errors.

REFERENCES

Ahmad Ali, Yanmin Zhu, and Muhammad Zakarya. Exploiting dynamic spatio-temporal graph convolutional neural networks for citywide traffic flows prediction. *Neural networks : the official journal of the International Neural Network Society*, 145:233–247, 2021. URL `https://api.semanticscholar.org/CorpusID:240238949`.

Federico Amato, Fabian Guignard, Sylvain Robert, and Mikhail F. Kanevski. A novel framework for spatio-temporal prediction of environmental data using deep learning. *Scientific Reports*, 10, 2020. URL `https://api.semanticscholar.org/CorpusID:229314791`.

Zied Ben-Bouallegue, Jonathan A. Weyn, Mariana C. A. Clare, Jesper Sören Dramsch, Peter Dueben, and Matthew Chantry. Improving medium-range ensemble weather forecasts with hierarchical ensemble transformers. *Artificial Intelligence for the Earth Systems*, 2023. URL `https://api.semanticscholar.org/CorpusID:257833736`.

Chiranjib Chaudhuri and Colin Robertson. Cligan: A structurally sensitive convolutional neural network model for statistical downscaling of precipitation from multi-model ensembles. *Water*, 12(12), 2020. ISSN 2073-4441. doi: 10.3390/w12123353. URL `https://www.mdpi.com/2073-4441/12/12/3353`.

Yann N. Dauphin, Angela Fan, Michael Auli, and David Grangier. Language modeling with gated convolutional networks. *CoRR*, abs/1612.08083, 2016. URL `http://arxiv.org/abs/1612.08083`.

Aram Davtyan, Sepehr Sameni, and Paolo Favaro. Randomized conditional flow matching for video prediction, 2022. URL `https://arxiv.org/abs/2211.14575`.

Lukas Drees, Immanuel Weber, Marc Rußwurm, and Ribana Roscher. Time dependent image generation of plants from incomplete sequences with cnn-transformer. In *German Conference on Pattern Recognition*, 2022. URL `https://api.semanticscholar.org/CorpusID:252545811`.

Aya Ferchichi, Ali Ben Abbes, Vincent Barra, and Imed Riadh Farah. Forecasting vegetation indices from spatio-temporal remotely sensed data using deep learning-based approaches: A systematic literature review. *Ecological Informatics*, 68:101552, 2022. ISSN 1574-9541. doi: https://doi.org/10.1016/j.ecoinf.2022.101552. URL `https://www.sciencedirect.com/science/article/pii/S1574954122000012`.

Nan Gao, Hao Xue, Wei Shao, Sichen Zhao, Kyle Kai Qin, Arian Prabowo, Mohammad Saiedur Rahaman, and Flora D. Salim. Generative adversarial networks for spatio-temporal data: A survey. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 13:1 – 25, 2020. URL `https://api.semanticscholar.org/CorpusID:221186900`.

Brian Groenke, Luke Madaus, and Claire Monteleoni. ClimAlign: Unsupervised statistical downscaling of climate variables via normalizing flows. August 2020.

Paula Harder. Generating physically-consistent high-resolution climate data with hard-constrained neural networks. In *AAAI 2022 Fall Symposium: The Role of AI in Responding to Climate Challenges*, 2022. URL `https://www.climatechange.ai/papers/aaaifss2022/5`.

Y. Ji, B. Gong, M. Langguth, A. Mozaffari, and X. Zhi. Clgan: a generative adversarial network (gan)-based video prediction model for precipitation nowcasting. *Geoscientific Model Development*, 16(10):2737–2752, 2023. doi: 10.5194/gmd-16-2737-2023. URL `https://gmd.copernicus.org/articles/16/2737/2023/`.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2014. URL `http://arxiv.org/abs/1412.6980`. cite arxiv:1412.6980Comment: Published as a conference paper at the 3rd International Conference for Learning Representations, San Diego, 2015.

Konstantin Klemmer, Tianlin Xu, Beatrice Acciaio, and Daniel B. Neill. SPATE-GAN: improved generative modeling of dynamic spatio-temporal patterns with an autoregressive embedding loss. *CoRR*, abs/2109.15044, 2021. URL `https://arxiv.org/abs/2109.15044`.

Manoj Kumar, Mohammad Babaeizadeh, Dumitru Erhan, Chelsea Finn, Sergey Levine, Laurent Dinh, and Durk Kingma. Videoflow: A flow-based generative model for video. *CoRR*, abs/1903.01434, 2019. URL `http://arxiv.org/abs/1903.01434`.

Andreas Lugmayr, Martin Danelljan, Fisher Yu, Luc Van Gool, and Radu Timofte. Normalizing flow as a flexible fidelity objective for photo-realistic super-resolution. *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 874–883, 2021. URL `https://api.semanticscholar.org/CorpusID:243833072`.

Jan Pisl, Lloyd Haydn Hughes, Marc Rußwurm, and Devis Tuia. Classification of tropical deforestation drivers with machine learning and satellite image time series. In *IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium*, pp. 911–914, 2023. doi: 10.1109/IGARSS52108.2023.10281472.

Stephan Rasp and Sebastian Lerch. Neural networks for post-processing ensemble weather forecasts. *ArXiv*, abs/1805.09091, 2018. URL https://api.semanticscholar.org/CorpusID:43927264.

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. doi: 10.1073/pnas.1810286115. URL https://www.pnas.org/doi/abs/10.1073/pnas.1810286115.

Suman Ravuri, Karel Lenc, Matthew Willson, Dmitry Kangin, Remi Lam, Piotr Mirowski, Megan Fitzsimons, Maria Athanassiadou, Sheleem Kashem, Sam Madge, Rachel Prudden, Amol Mandhane, Aidan Clark, Andrew Brock, Karen Simonyan, Raia Hadsell, Niall Robinson, Ellen Clancy, Alberto Arribas, and Shakir Mohamed. Skilful precipitation nowcasting using deep generative models of radar. *Nature*, 597:672–677, 09 2021a. doi: 10.1038/s41586-021-03854-z.

Suman V. Ravuri, Karel Lenc, Matthew Willson, Dmitry Kangin, Rémi Lam, Piotr Mirowski, Megan Fitzsimons, Maria Athanassiadou, Sheleem Kashem, Sam Madge, Rachel Prudden, Amol Mandhane, Aidan Clark, Andrew Brock, Karen Simonyan, Raia Hadsell, Niall H. Robinson, Ellen Clancy, Alberto Arribas, and Shakir Mohamed. Skillful precipitation nowcasting using deep generative models of radar. *CoRR*, abs/2104.00954, 2021b. URL https://arxiv.org/abs/2104.00954.

Markus Reichstein, Gustau Camps-Valls, Bjorn Stevens, Martin Jung, Joachim Denzler, Nuno Carvalhais, and Prabhat. Deep learning and process understanding for data-driven earth system science. *Nature*, 566:195 – 204, 2019. URL https://api.semanticscholar.org/CorpusID:61156451.

Marc Rußwurm and Marco Körner. Convolutional lstms for cloud-robust segmentation of remote sensing imagery. *ArXiv*, abs/1811.02471, 2018. URL https://api.semanticscholar.org/CorpusID:53249546.

Divya Saxena and Jiannong Cao. D-gan: Deep generative adversarial nets for spatio-temporal prediction. *ArXiv*, abs/1907.08556, 2019. URL https://api.semanticscholar.org/CorpusID:197935426.

Ashutosh Kumar Singh, Adrian Albert, and Brian White. Downscaling numerical weather models with gans. 2019. URL https://api.semanticscholar.org/CorpusID:226785468.

Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, September 2018.

Yanheng Wang, Danfeng Hong, Jianjun Sha, Lianru Gao, Lian Liu, Yonggang Zhang, and Xianhui Rong. Spectral-spatial-temporal transformers for hyperspectral image change detection. *IEEE Transactions on Geoscience and Remote Sensing*, PP:1–1, 01 2022. doi: 10.1109/TGRS.2022.3203075.

Campbell D. Watson, Chulin Wang, Timothy Lynar, and Komminist Weldemariam. Investigating two super-resolution methods for downscaling precipitation: Esrgan and car, 2020.

Christina Winkler, Daniel E. Worrall, Emiel Hoogeboom, and Max Welling. Learning likelihoods with conditional normalizing flows. *CoRR*, abs/1912.00042, 2019. URL http://arxiv.org/abs/1912.00042.

Mohsen Zand, Ali Etemad, and Michael Greenspan. Flow-based spatio-temporal structured prediction of dynamics, 2022.

Shixiang Zhu, H. Milton, Shuang Li, Zhigang Peng, and Yao Xie. Interpretable deep generative spatio-temporal point processes. 2020. URL https://api.semanticscholar.org/CorpusID:251439773.
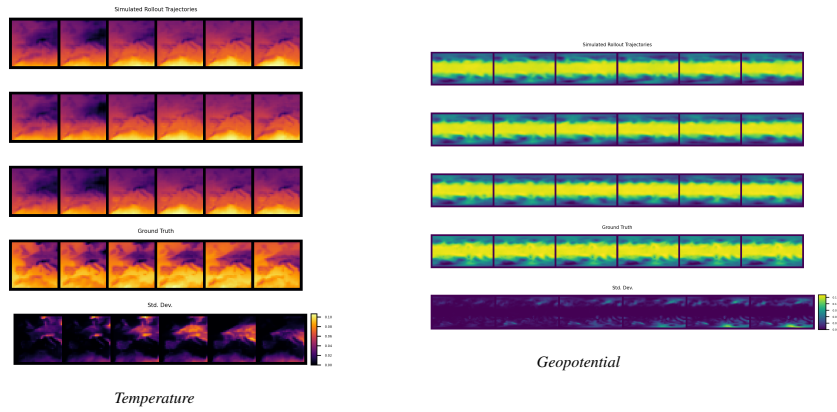
Figure 6: *Visualization of rollout trajectories starting from the same initial conditions on the ERA5 temperature and geopotential dataset from the **conditional GAN** model. The last row shows the squared absolute error. For rollouts from other methods, see appendix.*
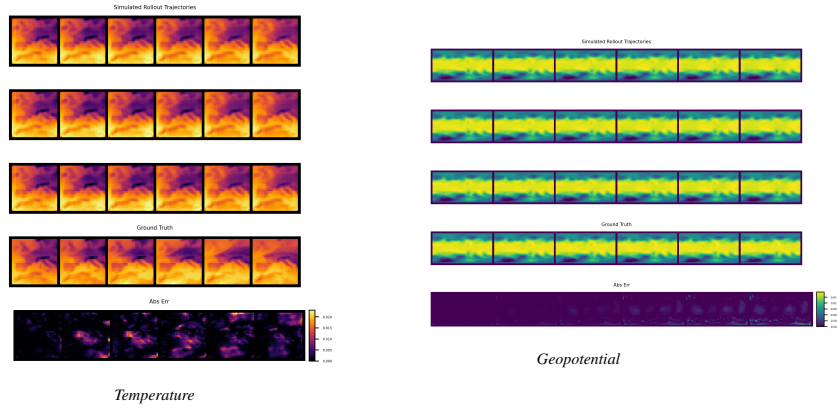

Figure 7: *Visualization of rollout trajectories starting from the same initial conditions on the ERA5 temperature and geopotential dataset from the **convLSTM** model. The last row shows the squared absolute error. For rollouts from other methods, see appendix.*
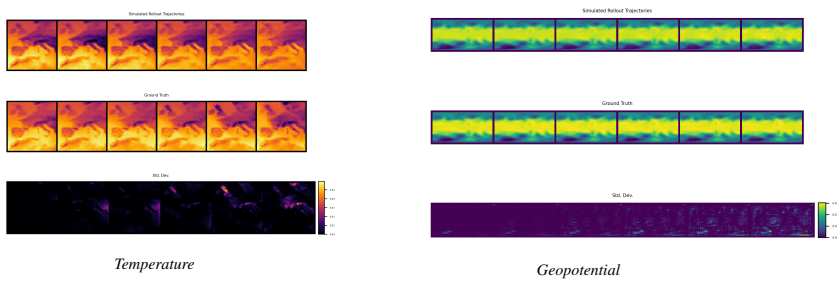

Figure 8: *Visualization of rollout trajectories starting from the same initial conditions on the ERA5 temperature and geopotential dataset from the **3DUNet** model. The last row shows the squared absolute error. For rollouts from other methods, see appendix.*

Table 1: *Architecture details for a single coupling layer in the spatio-temporal flow architecture. The variable c_{out} denotes the number of output channels. All convolutional layers are followed by a ReLU activation.*

Layer	Hidden Channels	Kernel size
Upsample Conv3d	1	$1 \times 1 \times 1$
Conv3d	512	3×3
Conv3d	512	3×3
Conv3d	c_{out}	1×1

A   APPENDIX: SIMULATED ROLLOUT TRAJECTORIES

B   APPENDIX: ARCHITECTURE

B.1   SPATIO-TEMPORAL CONDITIONAL NORMALIZING FLOW

B.2   GATED CONVOLUTIONAL LSTM

Table 2: *Architecture details for the full gated convolutional neural network, as described in Figure 9. We chose n=6 layers for all experiments.*

Layer	Hidden Channels	Kernel size	Padding
First Conv3d	128	3x3	1
Last Conv3d	256	3x3	1

Table 3: *Architecture details for the gated convolutional layer, as described in Figure 10. The first convolutional layer is followed by a concatenated ReLU activation.*

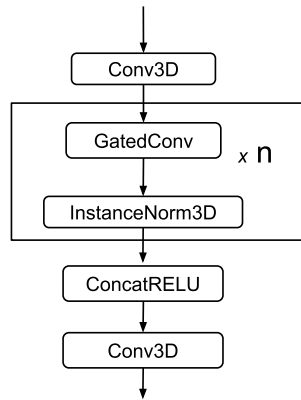| Layer	Hidden Channels	Kernel size	Padding
Conv3d	128	3x3	1
Conv3d	256	3x3	1

Figure 9: *Schematic of full gated convolutional neural network architecture with n layers.*
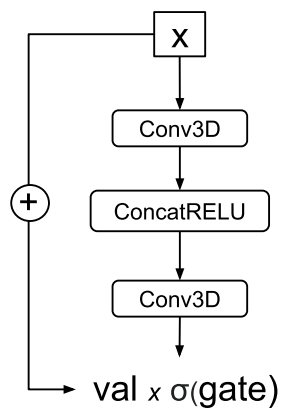

Figure 10: *Schematic of gated convolutional layer. 'val' stands for output value of the layer.*