Skilful precipitation nowcasting using deep generative models of radar

Abstract

This abstract presents a novel depth-generating model for radar-based probabilistic short-term forecasting of precipitation (i.e., precipitation nowcasting). Traditional nowcasting methods advance the precipitation field mainly from the radar-estimated wind field, but have difficulties in capturing some nonlinear events (e.g., convective onset). In contrast, recently introduced deep learning methods, although capable of directly predicting future rainfall, produce ambiguous nowcasts due to a lack of constraints in long time forecasts, and perform poorly in predicting moderate to heavy rainfall events. This deep generative model proposed by the authors aims to address the above challenges. The model is able to provide realistic and spatio-temporally consistent forecasts for areas up to 1536 km × 1280 km, with forecast times ranging from 5 to 90 minutes. Through statistical, economic and cognitive measures, the study shows that this approach improves in terms of forecast quality, consistency and value. Systematically evaluated by over 50 expert meteorologists, this generative model ranked 1 in accuracy and utility in 89% of cases against two competing methods. Quantitative validation demonstrated the skill of these nowcasts without the use of fuzzy processing. The study also demonstrates how generative nowcasting can provide probabilistic forecasts that improve forecast value and support operational utility, especially at resolutions and forecast times that are difficult for other methods to cope with.

Keywords: Depth-generating model; Short-term forecast of precipitation; Probabilistic prediction.

1 Introduction

This study addresses the limitations of traditional short-term forecasting methods for precipitation by introducing a novel depth-generating model. This model utilizes radar data for probabilistic forecasting, which improves the accuracy, consistency, and realism of forecasts, especially when dealing with moderate to heavy rainfall events. This has important implications for a number of areas that rely on weather decision-making, such as emergency services, energy management, and transportation control. The model's performance in terms of accuracy and utility has been validated through evaluation by over 50 expert meteorologists, demonstrating its significant progress in improving the quality and utility of precipitation nowcasting.

2 Related works

The related work involved in this study on short-term forecasting (nowcasting) of precipitation can be categorized into a few main groups: firstly, traditional precipitation forecasting methods, which rely on radar and satellite data to project precipitation fields, but have challenges in predicting nonlinear weather events. Second is the application of deep learning in precipitation forecasting; these methods predict future rainfall by directly using radar data, but have accuracy issues in long-term forecasting. Third are probabilistic precipitation forecasting methods that provide a more flexible and wide range of forecasts. Finally, there is the application of deep generative modeling in precipitation forecasting, an emerging field that aims to improve the accuracy and utility of forecasts by combining traditional and modern deep learning techniques. This research centers on exploring and enhancing the efficiency and reliability of precipitation nowcasting to provide a more robust support tool for weather-dependent decision making.

2.1 Traditional methods

Traditional methods for short-term forecasting of precipitation, also known as nowcasting, typically include the following main techniques: 1. Radar data-based precipitation field imputation: This is the core of existing operational nowcasting methods, in which radar-based wind field estimates are used to advance the precipitation field. However, this approach has challenges in capturing important nonlinear weather events such as convective onset. 2. Numerical Weather Prediction (NWP) Systems: NWP systems generate multiple realistic precipitation forecasts by modeling the coupled physical equations of the atmosphere. These systems are natural candidates for nowcasting because of their ability to derive probabilistic forecasts and uncertainty estimates from ensembles of future forecasts. However, for zero- to two-hour precipitation forecasts, NWP systems tend to provide poor forecasts due to the fact that this time period is less than the model's start-up time and because of difficulties with non-Gaussian distributions in the data assimilation process. 3. Radar Observations Forecasting Methods: Due to limitations in the NWP, alternative methods of forecasting using radar composite observations have been developed. In the UK, radar data are updated every five minutes with a resolution of 1 km x 1 km. 4. Probabilistic nowcasting methods: Examples include STEPS and PySTEPS, which follow the NWP approach of using ensembles to account for uncertainty, but where simulated precipitation is based on convective equations and radar source terms. In these models, the kinematic field is estimated from optical flow, smoothing penalties are used to approximate convective forecasts, and stochastic perturbations are added to the kinematic field and intensity models. These stochastic simulations allow both probabilistic and deterministic forecasts to be derived from them and are applicable at multiple spatial scales ranging from the kilometer scale to the watershed size.

3 Method

3.1 Overview of the methodology of this paper

The method proposed in this paper is a deep generative model (DGMR) with the following main design and architecture: 1. architectural design: this nowcasting model is a generator that is trained using two discriminators and an additional regularization term. This generator contains a conditional stack that processes the last four radar fields that are used as contexts. Effectively utilizing such contexts is often a challenge for conditional generative models, and this stack structure allows for information processing at multiple resolutions using contextual data. The contextual representations generated by the stack are used as inputs to the sampler. One potential conditional stack is sampled from an N(0, 1) Gaussian distribution and transformed into a second potential representation. The sampler is a recursive network of convolutionally gated recurrent units (GRUs) that use the context and latent representations as inputs. The sampler predicts 18 future radar fields over the next 90 minutes. 2. Adversarial learning: two discriminators are used in the paper for spatial and temporal adversarial learning. The spatial and temporal discriminators share the same structure, the difference being that the temporal discriminator uses 3D convolution to account for the temporal dimension. The spatial discriminator uses only 8 of the 18 prediction times and uses randomized 128×128 cropping for the temporal discriminator. These choices allow the model to adapt to memory constraints. Spatial attention blocks were added to the latent condition stack to enhance the robustness of the model to different regions and event types, and implicit regularization was provided to prevent overfitting, especially for the US dataset. 3. Optimization and regularization: the generator and discriminator use spectrally normalized convolution throughout the network, an approach widely recognized to provide improvements in optimization. During model development, researchers initially found that adding a batch normalization layer (excluding variance scaling) before the linear layer of both discriminators improved training stability. Although the final results used batch normalization, the researchers were later able to obtain nearly identical quantitative and qualitative results without using it. [1]Figure 1 rough sketch:

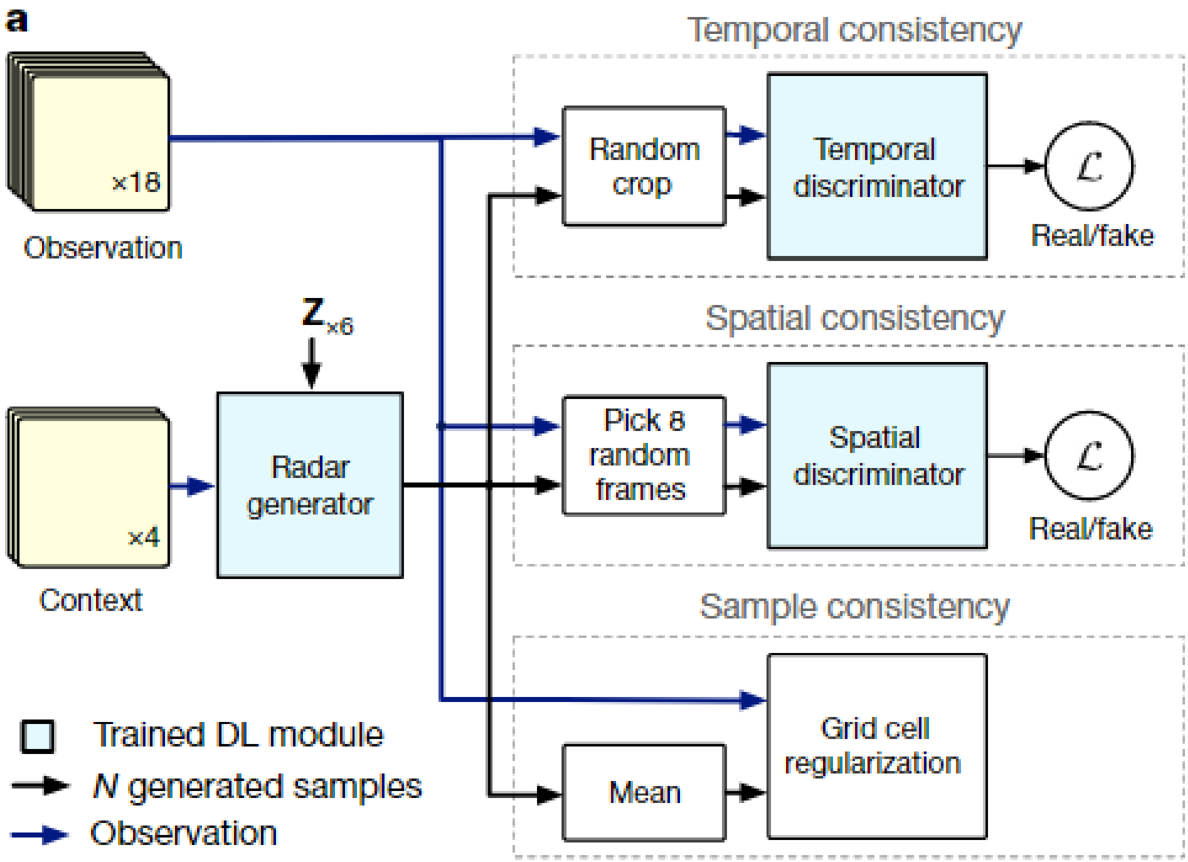

Figure 1. Overview of the method

3.2 Feature extraction

The feature extraction module in this paper's approach includes a conditioning stack that processes the past four radar fields as context. Effectively utilizing such a context is often a challenge for condition generation models, but this stack structure allows information from the context data to be used at multiple resolutions. This stack produces a contextual representation that serves as an input to the sampler. Another latent conditioning stack (latent conditioning stack) samples from the N(0,1) Gaussian distribution and transforms it into a second latent representation. The sampler is a recursive network of convolutionally gated recurrent units (GRUs) that use the context and latent representations as inputs. The sampler is responsible for predicting 18 future radar fields over the next 90 minutes. Feature Extraction Module Figure 2:

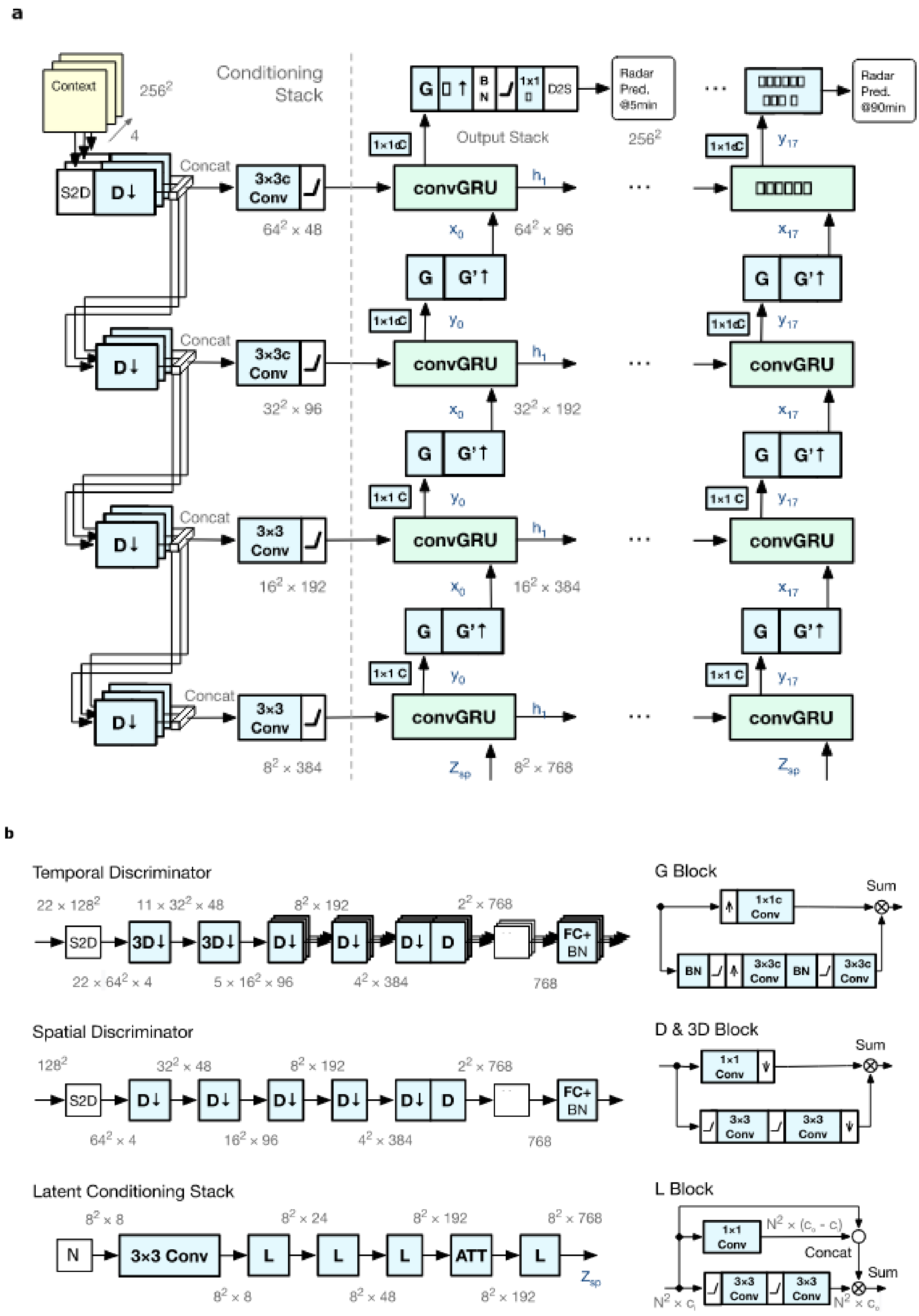

Figure 2. Feature extraction

3.3 Loss

The generator is trained with losses from the two discriminators and a grid cell regularization term (denoted $\mathcal{L}_R(\theta)$). The spatial discriminator D_ϕ has parameters ϕ, the temporal discriminator T_ψ has parameters ψ, and the generator G_θ has parameters θ. We indicate the concatenation of two fields using the notation $\{X; G\}$. The generator objective that is maximized is

$$\mathcal{L}_G(\theta) = \mathbb{E}_{X_{1:M},Z}\left[\mathbb{E}_{X_{M+1:N}}[D(G_\theta(Z; X_{1:M}))] + T(X_{1:M}; G_\theta(Z; X_{1:M}))\right] - \lambda_R(\theta). \tag{1}$$

$$\mathcal{L}_R(\theta) = \frac{1}{HWN}\left\|(E[Z, G_\theta(Z; X_{1:M})] - X_{M+1:M+N}) \odot W(X_{M+1:M+N})\right\|_1. \tag{2}$$

We use Monte Carlo estimates for expectations over the latent Z in equations (2) and (3). These are calculated using six samples per input $X_{1:M}$, which comprises $M = 4$ radar observations. The grid cell regularizer ensures that the mean prediction remains close to the ground truth, and is averaged across all grid cells along the height H, width W and lead-time N axes. It is weighted towards heavier rainfall targets using the function $w(y) = \max(y + 1, 24)$, which operate element-wise for input vectors, and is clipped at 24 for robustness to spuriously large values in the radar. The GAN spatial discriminator loss $\mathcal{L}_D(\phi)$ and temporal discriminator loss $\mathcal{L}_T(\psi)$ are minimized with respect to parameters ϕ and ψ, respectively; $\text{ReLU}(x) = \max(0, x)$. The discriminator losses use a hinge loss formulation:

$$\mathcal{L}_D(\phi) = \mathbb{E}_{X_{1:M},Z}\left[\text{ReLU}(1 - D_\phi(X_{M+1:M+N})) + \text{ReLU}(1 + D_\phi(G(Z; X_{1:M})))\right]. \tag{3}$$

$$\mathcal{L}_T(\psi) = \mathbb{E}_{X_{1:M},Z}\left[\text{ReLU}(1 - T_\psi(X_{1:M+N})) + \text{ReLU}(1 + T_\psi(X_{1:M}; G(Z; X_{1:M})))\right]. \tag{4}$$

4 Implementation details

4.1 Comparing with the released source codes

The improvements I added to the existing code focus on the following areas:

1. performance evaluation enhancements: - I introduced TensorFlow functions for calculating the Critical Success Index (CSI) and the Continuous Ranking Probability Score (CRPS). These are important metrics for evaluating the accuracy of precipitation forecasts. - I also define auxiliary functions to handle NaN values, which are important to ensure the stability and accuracy of the computations.

2. Integration of model prediction and evaluation: - I loaded a model and made predictions for the input frames, and then calculated the CSI and CRPS for each sample and each frame. this approach allows you to evaluate the performance of the model in more detail and specificity. - I also initialized the arrays used to store the CSI and CRPS values and performed traversal calculations, which provided a comprehensive view of the model's performance.

3. Visualization enhancements: - I implemented a function to display the prediction results and evaluation metrics in an animation, which not only provides an intuitive understanding of the model's prediction results, but also shows its performance metrics at the same time, enhancing the informativeness of the visualization. - I

also realized the function of outputting the animation to GIF format, which makes it easier to share and display the results.

4. Save function for vector graphics: - I provided the function to save each frame of each sample as SVG format. the advantages of SVG format are its scalability and high quality graphical representation, which is ideal for printing and precise display.

These improvements make the original code significantly better for performance evaluation, result visualization and data presentation, enhancing the usefulness of the model and the presentation of results.

5 Results and analysis

The result of this experiment shows a radar precipitation map, from which we can see the distribution and intensity of precipitation. In the image, the colors may represent the magnitude of precipitation, usually red and yellow indicate higher precipitation while blue indicates lower precipitation. The text above the image provides a quantitative assessment of precipitation prediction performance for this particular sample and frame:

- CSI 2mm = 0.57: This represents a Critical Success Index (CSI) of 0.57 when using 2mm as the threshold.CSI is a metric of prediction performance, with a value of 1 indicating a perfect prediction and 0 indicating no skill.0.57 is a relatively moderate score, indicating that at the 2mm precipitation level, the prediction makes a relatively accurate distinction.

- CSI 8mm = 0.38: When the threshold is raised to 8 mm, the CSI decreases to 0.38. This generally implies a decrease in the model's performance in predicting heavier precipitation events. Because there are fewer heavy magnitude precipitation events, predicting them is more challenging, hence the lower CSI.

- CRPS = 0.35: The Continuous Ranked Probability Score (CRPS) is another metric that assesses forecasting performance and measures how well the forecast distribution matches the observed data.CRPS values range from 0 (perfect forecast) to 1 (worst forecast), so 0.35 indicates that the model provides skillful forecasts overall.

In summary, the results of this experiment indicate that the model shows some skill in predicting precipitation, especially performing relatively well on lower precipitation magnitudes (e.g., 2 mm), but there are some challenges in predicting higher magnitudes (e.g., 8 mm).

Figure 3 predicted results of the first sample of the first frame (fifth frame) of the fifth minute:

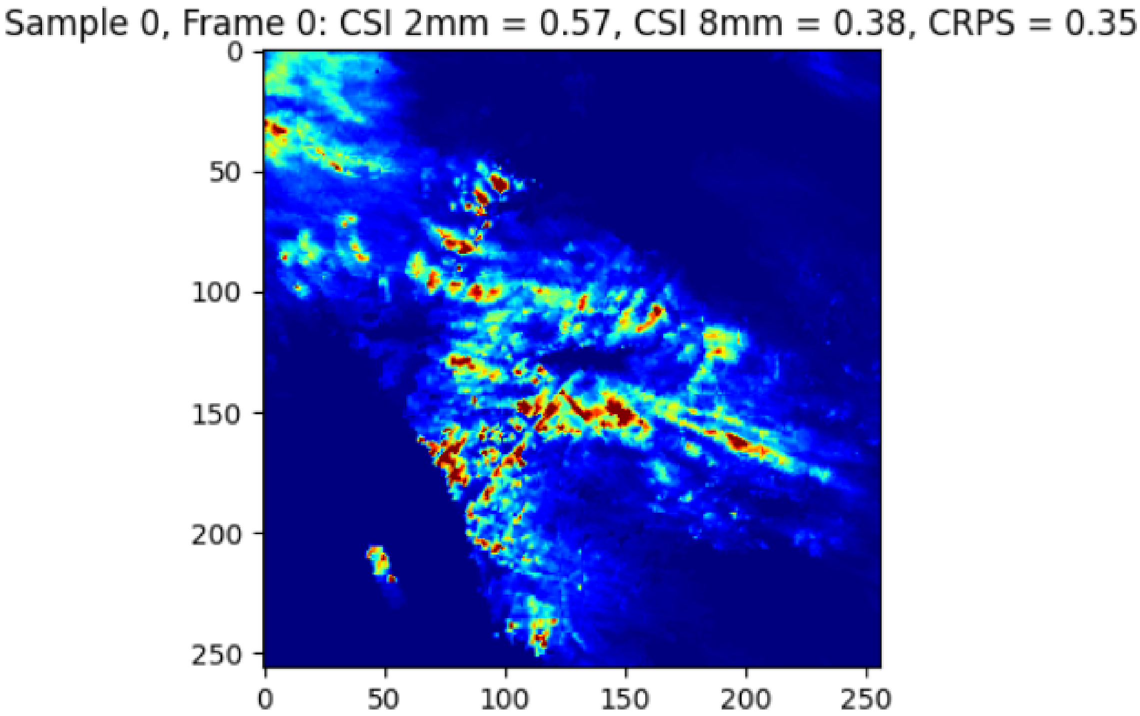

Figure 3. Experimental results

6 Conclusion and future work

The conclusions and future work section of the literature states:

- IMPROVEMENTS TO EXISTING METHODS: This research directly addresses the long-standing problem of accurate short-term weather forecasting (nowcasting) by using Deep Generative Models (DGMs) and improves upon existing solutions. This provides needed insights for decision makers in the real world. Using statistical, economic, and cognitive measures, the authors demonstrate that their generative nowcasting methodology has made progress in improving the quality, consistency, and value of forecasts, providing fast and accurate short-term forecasts at lead times where existing methods have difficulty.

- FUTURE CHALLENGES: Although generative methods provide skillful forecasts compared to other solutions, forecasting large amounts of precipitation over long lead times remains a challenge. This study highlights that standard validation metrics and expert judgments are not always indicative of each other's value, underscoring the need for new quantitative measures that are better aligned with operational utility when evaluating models with less inductive bias and high capacity.

- DIRECTIONS FOR FUTURE WORK: The authors hope that their work will form the basis for new data, code and validation methods, as well as greater convergence between machine learning and environmental science in predicting larger sets of environmental variables. This will make it possible to provide both competitive validation and operational utility.

References

[1] 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. Skillful precipitation nowcasting using deep generative models of radar. *Nature*, 597:672–677, 2021.