Supporting Information for “A Generative Deep Learning Approach to Stochastic Downscaling of Precipitation Forecasts”
Lucy Harris¹, Andrew T. T. McRae¹, Matthew Chantry², Peter D. Dueben²,
Tim N. Palmer¹

¹Department of Physics, University of Oxford
²European Centre for Medium-Range Weather Forecasts

Contents of this file

1. Introduction
2. Description of case studies
3. Figures S1 to S40

1. Introduction

1.1. Rank histogram plots for increasing lead times

Figures S1 to S6 show rank histogram plots for the GAN, VAE-GAN and ecPoint models for increasing lead time.

These plots are assessed on all available 2020 00Z forecasts, around 350 events in total. An ensemble size of 100 is used in each case. When evaluated on all events, all 3 methods

Corresponding author: L. J. Harris, Department of Atmospheric, Ocean and Planetary Physics, University of Oxford. (lucy.harris2@physics.ox.ac.uk)
(GAN, VAE-GAN, and ecPoint approach baseline) show only small differences between 24h, 48h and 72h lead-time. However, all three methods show very slight worsening at increasing lead times. We would like to note that all 3 were trained on 7–17hr lead time data. When evaluated on the top 0.01% of IFS predictions, all methods perform noticeably worse at 48h and 72h lead times than at 24h lead time. Our neural network approaches are under-dispersive, whereas the ecPoint approach is over-dispersive.

1.2. Precision-Recall curves

Precision-recall curves (PRC) for the GAN, VAE-GAN and ecPoint part-correlation and no-correlation models are plotted here. These were generated by evaluating over 256 different images, with an ensemble size of 100.

Precision is the number of true positives divided by the sum of the true positives and false positives. Recall is calculated as the ratio of the number of true positives to the sum of the true positives plus false negatives. Recall is the same as sensitivity. P-R curves are useful in cases where there is an imbalance in the observations between the two classes, such as in our dataset where there are many more light-rain events than heavy-rain events.

Whilst the baseline is fixed with ROC curves, the baseline of P-R curves is determined by the ratio of positives (P) and negatives (N) as \[ y = \frac{P}{P+N}, \] indicated by the dashed line. A model with perfect skill is depicted as a point at (1,1). A skilful model is represented by a curve that tends towards (1,1) above the flat line of no skill. The area under each line is given, as a proxy for overall performance. However, this should be interpreted with caution, particularly in cases where the straight-line portion between (0,1) or (1, \( \frac{P}{P+N} \)) and the initial/final data values contribute significantly to the area.

September 26, 2022, 10:08am
1.3. Further RAPSD plot examples

Figures S19 to S22 show further RAPSD plot examples for the GAN and the VAE-GAN. These plots are included here to show that the RAPSD results are consistently good across all the example images. Figures S19 and S20 show plots for the second standard example image case for the GAN and the VAE-GAN, respectively, and Figures S21 and S22 show RAPSD plots for the third example image. The GAN plots also include the Lanczos and RainFARM model RAPSD for comparison. The RAPSD plots are generally fairly consistent: both the GAN and VAE-GAN models clearly outperform the Lanczos and RainFARM methods. These plots clearly demonstrate the added value of using the GAN and VAE-GAN models over interpolation of the IFS forecast (e.g. using Lanczos interpolation). It is also interesting to note the spread between individual model ensemble members in both the GAN and the VAE-GAN.

1.4. Further ROC curve examples

The ROC curves for the GAN and VAE-GAN models shown here were generated by evaluating over 256 different images, with a batch size of 1 (for memory reasons) and an ensemble size of 100. A full set of ROC curves are shown here for 0.1 mm/hr and 2 mm/hr in the appendix, in addition to the average pooling and max pooling plots for the 0.5 mm/hr and 5 mm/hr thresholds shown in the main body of the paper.

1.5. Further FSS plot examples

FSS curves for the GAN and VAE-GAN models are plotted in Figures S33 and S34. These were generated by evaluating over 256 different images, with an ensemble size of 100. FSS curves are shown here for the additional 0.1 mm/hr and 2 mm/hr thresholds.
1.6. Rank histogram plots for ablation studies

Figures S35 and S36 show rank histograms for the “no content-loss” and “no geographic fields” ablation studies, compared with the original GAN, on all events and on the top 0.01% of forecast predictions.

1.7. Further model prediction examples

These figures show the GAN and VAE-GAN model predictions for a further four different weather scenarios, on top of the four used in the main paper, and are included here to demonstrate the wide-ranging capabilities of the models.

2. Description of case studies used in the main paper

The first example is from 21:00-22:00 UTC of the 7th June 2019, and shows a highly-structured pattern of rainfall, with a curved band of moderate precipitation across the centre of the image surrounded by lighter rainfall, with additional structure in the bottom right-hand corner. All GAN and VAE-GAN model predictions produce a similar banded structure to the NIMROD ground truth image, with locally sharply varying structure within the rainfall band. The models also make bold predictions of extremely localised intense rainfall, whereas the IFS forecast fails to capture the peak intensities of the rainfall. There is fine-scale variation between the different predictions, but the large-scale structure remains consistent, suggesting a high level of confidence in this forecast.

The second example is from 18:00-19:00 UTC of the 1st May 2019, and shows lighter, scattered rainfall across the country. The IFS forecast does not capture the full extent of the rainfall and under-predicts the intensity in places, whilst over-predicting the area of
light rainfall over Northern Ireland. The GAN and VAE-GAN predictions show a much more realistic, finely detailed structure within the precipitation patches, whilst maintaining the same, broadly correct, large-scale structure. The models successfully remove the over-prediction of light rain over Northern Ireland. There is significant variation between the different predictions, corresponding to significant uncertainty in the scenario.

The third example is from 01:00-02:00 UTC of the 5\textsuperscript{th} August 2019, and shows two distinct bands of rain across the image, with the large-scale structure captured reasonably well by the IFS prediction. However, the western band of rain has a much more defined spatial structure than displayed in the IFS forecast, which predicts light rainfall over a greater area than in reality. The north-eastern band of rain has a fairly simple structure, but the intensity is significantly under-predicted by the IFS forecast. Firstly, and perhaps most importantly, the GAN model predictions capture the full intensity of the rainfall. The VAE-GAN corrects the intensity somewhat, but not enough. Secondly, unlike the IFS forecast, the GAN and VAE-GAN do not over-predict light rainfall for a large area for the western rainfall band. There is significant variation in the location of the precipitation due to the inherent uncertainty, but there is evidence the models are using the orographic information to produce sensible predictions, as the rainfall is consistently predicted over the Highlands, the Brecon Beacons, and the western Pennines.

The fourth example is from 04:00-05:00 UTC of the 27\textsuperscript{th} January 2019, and shows scattered light rainfall clustered into two groups, one in the north and one in the south. The NIMROD data shows lots of fine-scale structure and some medium-intensity rainfall over the Highlands. The IFS forecast only predicts light rainfall, and whilst the large-scale structure is broadly correct, there is an area of rainfall forecast to the north-east of the
UK over the North Sea that is not present in the NIMROD data. Two of the three GAN model predictions also show this area of rainfall forecast, with the third showing some rainfall forecast in the same area. The VAE-GAN predictions seem to largely eliminate this spurious prediction. However, overall, the GAN and VAE-GAN predictions closely mirror the NIMROD data, showing patchy, light rainfall across much of the country with a more realistic-looking structure than the blurrier IFS forecast. The patch of rain over the south of the UK is well-calibrated with location and intensity matching the NIMROD data closely for all three example GAN model predictions, while the VAE-GAN predictions perhaps under-predict this slightly. There is again evidence of the orography informing the model, with a consistent concentration of rainfall centred on the Highlands. The NIMROD data shows bands of rain aligned north-west to south-east with the wind. We are pleased to see similar features existing in the GAN predictions.
Figure S1. Calibration plot for the GAN model at increasing lead times: (a) shows the frequency of per-pixel normalised ranks for the trained GAN model evaluated on the hold-out dataset (2020). The dotted grey line shows the ideal distribution for comparison. (b) shows the same as panel a, except displaying the CDFs of the distributions.

Figure S2. Thresholded calibration plot for the GAN model at increasing lead times: (a) shows the frequency of per-pixel normalised ranks over the 0.01% threshold for the trained GAN model evaluated on the hold-out dataset (2020). The dotted grey line shows the ideal distribution for comparison. (b) shows the same as panel a, except displaying the CDFs of the distributions.
Figure S3. Calibration plot for the VAE-GAN model at increasing lead times: (a) shows the frequency of per-pixel normalised ranks for the trained VAE-GAN model evaluated on the hold-out dataset (2020). The dotted grey line shows the ideal distribution for comparison. (b) shows the same as panel a, except displaying the CDFs of the distributions.

Figure S4. Thresholded calibration plot for the VAE-GAN model at increasing lead times: (a) shows the frequency of per-pixel normalised ranks over the 0.01% threshold for the trained VAE-GAN model evaluated on the hold-out dataset (2020). The dotted grey line shows the ideal distribution for comparison. (b) shows the same as panel a, except displaying the CDFs of the distributions.
Figure S5. Calibration plot for the ecPoint model at increasing lead times: (a) shows the frequency of per-pixel normalised ranks for the ecPoint model evaluated on the hold-out dataset (2020). The dotted grey line shows the ideal distribution for comparison. (b) shows the same as panel a, except displaying the CDFs of the distributions.

Figure S6. Thresholded calibration plot for the ecPoint model at increasing lead times: (a) shows the frequency of per-pixel normalised ranks over the 0.01% threshold for the trained ecPoint model evaluated on the hold-out dataset (2020). The dotted grey line shows the ideal distribution for comparison. (b) shows the same as panel a, except displaying the CDFs of the distributions.
Figure S7. PR curves for the GAN, VAE-GAN and ecPoint models for a 0.1 mm/hr precipitation threshold, pixel-wise

Figure S8. PR curves for the GAN, VAE-GAN and ecPoint models for a 0.1 mm/hr precipitation threshold, average pooling over 16×16 pixels
Figure S9. PR curves for the GAN, VAE-GAN and ecPoint models for a 0.1 mm/hr precipitation threshold, max pooling over 16×16 pixels

Figure S10. PR curves for the GAN, VAE-GAN and ecPoint models for a 0.5 mm/hr precipitation threshold, pixel-wise
Figure S11. PR curves for the GAN, VAE-GAN and ecPoint models for a 0.5 mm/hr precipitation threshold, average pooling over 16×16 pixels.

Figure S12. PR curves for the GAN, VAE-GAN and ecPoint models for a 0.5 mm/hr precipitation threshold, max pooling over 16×16 pixels.
Figure S13. PR curves for the GAN, VAE-GAN and ecPoint models for a 2.0 mm/hr precipitation threshold, pixel-wise

Figure S14. PR curves for the GAN, VAE-GAN and ecPoint models for a 2.0 mm/hr precipitation threshold, average pooling over 16×16 pixels
Figure S15. PR curves for the GAN, VAE-GAN and ecPoint models for a 2.0 mm/hr precipitation threshold, max pooling over 16\times16 pixels

Figure S16. PR curves for the GAN, VAE-GAN and ecPoint models for a 5.0 mm/hr precipitation threshold, pixel-wise
Figure S17. PR curves for the GAN, VAE-GAN and ecPoint models for a 5.0 mm/hr precipitation threshold, average pooling over 16×16 pixels.

Figure S18. PR curves for the GAN, VAE-GAN and ecPoint models for a 5.0 mm/hr precipitation threshold, max pooling over 16×16 pixels.
Figure S19. Plot displaying the radially-averaged log power spectra of example image 2 comparing our GAN model to both Lanczos interpolation and the RainFARM method.

Figure S20. Plot displaying the radially-averaged log power spectra of example image 2 comparing our VAE-GAN model to both Lanczos interpolation and the RainFARM method.
Figure S21. Plot displaying the radially-averaged log power spectra of example image 3 comparing our GAN model to both Lanczos interpolation and the RainFARM method.

Figure S22. Plot displaying the radially-averaged log power spectra of example image 3 comparing our VAE-GAN model to both Lanczos interpolation and the RainFARM method.
Figure S23. ROC curves for the GAN, VAE-GAN and ecPoint models for a 0.1 mm/hr precipitation threshold, pixel-wise

Figure S24. ROC curves for the GAN, VAE-GAN and ecPoint models for a 0.1 mm/hr precipitation threshold, average pooling (16×16 pixels)
Figure S25. ROC curves for the GAN, VAE-GAN and ecPoint models for a 0.1 mm/hr precipitation threshold, max pooling (16×16 pixels)

Figure S26. ROC curves for the GAN, VAE-GAN and ecPoint models for a 0.5 mm/hr precipitation threshold, average pooling (16×16 pixels)

September 26, 2022, 10:08am
**Figure S27.** ROC curves for the GAN, VAE-GAN and ecPoint models for a 0.5 mm/hr precipitation threshold, max pooling (16×16 pixels)

**Figure S28.** ROC curves for the GAN, VAE-GAN and ecPoint models for a 2.0 mm/hr precipitation threshold, pixel-wise
Figure S29. ROC curves for the GAN, VAE-GAN and ecPoint models for a 2.0 mm/hr precipitation threshold, average pooling (16×16 pixels)

Figure S30. ROC curves for the GAN, VAE-GAN and ecPoint models for a 2.0 mm/hr precipitation threshold, max pooling (16×16 pixels)

September 26, 2022, 10:08am
Figure S31. ROC curves for the GAN, VAE-GAN and ecPoint models for a 5.0 mm/hr precipitation threshold, average pooling (16×16 pixels)

Figure S32. ROC curves for the GAN, VAE-GAN and ecPoint models for a 5.0 mm/hr precipitation threshold, max pooling (16×16 pixels)
Figure S33. FSS curves for the GAN, VAE-GAN and ecPoint models, 0.1 mm/hr precipitation threshold

Figure S34. FSS curves for the GAN, VAE-GAN and ecPoint models, 2.0 mm/hr precipitation threshold
Figure S35.  Calibration plot for various ablated GAN models: (a) shows the frequency of per-pixel normalised ranks for the trained models evaluated on the hold-out dataset (2020). The dotted grey line shows the ideal distribution for comparison. (b) shows the same as panel a, except displaying the CDFs of the distributions.

Figure S36.  Thresholded calibration plot for various ablated GAN models: (a) shows the frequency of per-pixel normalised ranks over the 0.01% threshold for the trained models evaluated on the hold-out dataset (2020). The dotted grey line shows the ideal distribution for comparison. (b) shows the same as panel a, except displaying the CDFs of the distributions.
Figure S37. Model prediction examples and means including input fields of total precipitation, wind speed and direction, and orography, 11:00-12:00 UTC, 05 April 2019
Figure S38. Model prediction examples and means including input fields of total precipitation, wind speed and direction, and orography, 18:00-19:00 UTC, 23 June 2019.
Figure S39. Model prediction examples and means including input fields of total precipitation, wind speed and direction, and orography, 04:00-05:00 UTC, 19 Dec 2019
Figure S40. Model prediction examples and means including input fields of total precipitation, wind speed and direction, and orography, 13:00-14:00 UTC, 08 Feb 2020.