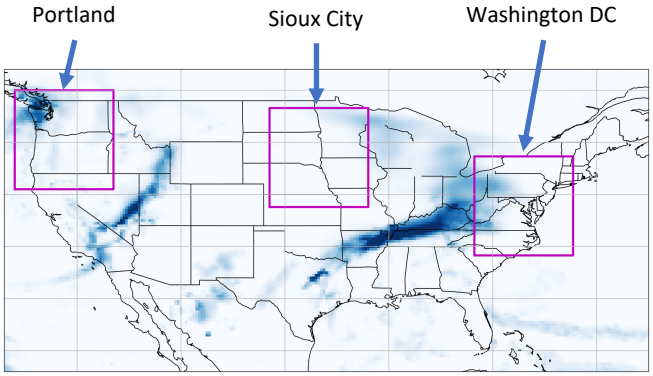
Bias correction and downscaling of rainfall.

Fenwick Cooper, Andrew McRae, Matthew Chantry, Lucy Harris & Tim Palmer, University of Oxford

Application of the cGAN neural network model of Harris, McRae, Chantry, Dueben & Palmer (2022) to **three regions of the USA**.

9.5° x 9.5° regions centered on:



Predicting 4km Radar data (NCEP stage IV) from ¼ degree ERA5

Results

Table 1: Continuous rank probability score (CRPS) of cGAN precipitation trained on ERA5 at each location.

Training location	Correcting ERA5 Testing location				Correcting HIRES
	Portland	Sioux	Washington	UK	UK
Portland	**0.0525**	0.0695	0.1325	0.1216	0.1190
Sioux City	0.0627	**0.0634**	0.2414	0.3419	0.3283
Washington	0.0592	0.0657	**0.1151**	0.1149	0.1034
UK	0.0618	0.0667	0.1259	**0.0931**	0.0984
ERA5 MAE	0.096	0.076	0.148	0.121	0.121

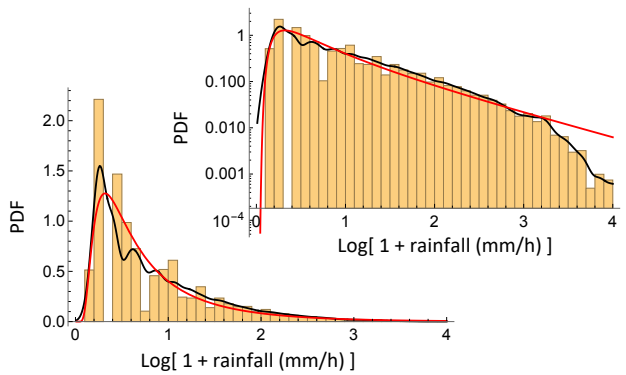
cGAN trained on HIRES IFS correcting HIRES IFS 7-17h forecasts, has a CRPS of **0.0856**, compared to an IFS MAE of 0.1412 (Harris et al. 2022).

Linear regression models

Why try linear regression?
1. Fast
2. Simple

Washington DC Regan international airport

1979-2021 hourly +. Total ~40'000 rain gauge measurements.



Kernel density estimate

PDF fit (inverse gamma, see below)

The histogram has gaps because of the precision of the rain gauge.

Even taking logs of the rainfall data results in a long-tailed distribution. Linear regression is complicated by the extreme nature of rainfall.

We want to understand what it is about precipitation that makes it extreme.
We can do this by fitting a simple model with extreme behavior.

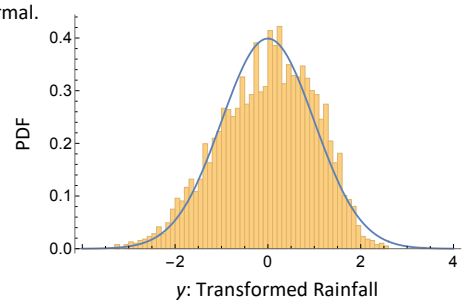
$$\frac{dx}{dt} = Bx + f + \sqrt{2A}\, x\, \xi(t)$$

x – Rainfall
B, f, A – Constants
$\xi(t)$ – White noise

The PDF of this model is the inverse gamma distribution.

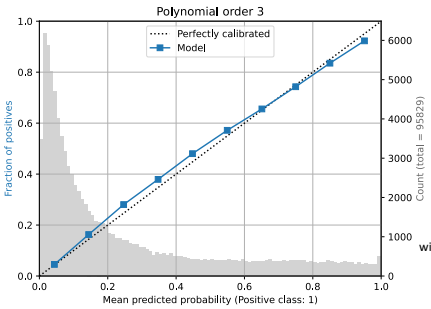
$$\rho(x) = \frac{A}{f\,\Gamma\left(1-\frac{B}{A}\right)} \operatorname{Exp}\left(-\frac{f}{Ax}\right)\left(\frac{f}{Ax}\right)^{2-\frac{B}{A}}$$

Which gives us the transformation of the rainfall to something more Normal.

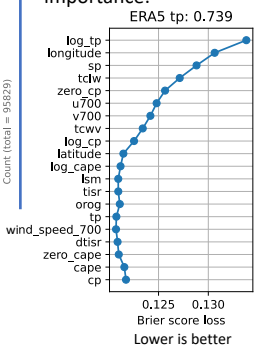


Reliable regression models can be formulated using standard methods.

EG. for probability that rainfall exceeds a threshold:



And cumulative (top-bottom) predictor importance:



(CRPS compared to cGAN to follow …)

Conclusions

Neural network model cGAN:

1. Regional training is necessary.
   - The Sioux City model performs badly elsewhere.
   - The Washington DC model is a good all-rounder.
2. Training using HIRES gives improved predictions compared to training with ERA5.
   - Whether predicting ERA5 or predicting short range (7-17h) HIRES forecasts.

Rainfall distribution:

1. The distribution of rainfall is well modeled by a simple multiplicative noise process.
   - This suggests a transformation that removes the problem of fat tails in the distribution.
   - The distribution is not compatible with measures like the RMSE because it doesn't have a variance.
2. Simple regression models are reliable and improve the Brier score for rainfall exceeding a certain value.