
Additional results for “Probabilistic Forecasting with Generative Networks via Scoring Rule Minimization”

Lorenzo Pacchiardi

Rilwan A. Adewoyin

Peter Dueben

Ritabrata Dutta

We provide here additional figures for the different forecasting methods on the WeatherBench dataset.

Figure 1 reports realization and prediction with the deterministic regression methods; instead, in Figures 2, 3, and 4 we report realization and 5 different forecasts obtained with all probabilistic methods discussed in the main text (Section 5.2). In Figures 5, 6, and 7, we show the deviation of the same realizations and forecasts from the forecast mean (obtained empirically from 100 forecasts). If the forecast distribution is calibrated, the realization should look similar to the forecasts themselves. You can see how this is roughly the case for the best performing SRs (as for instance the Patched Energy Score in Figure 6) as well as for WGAN-GP, but it clearly not the case for GAN (Figure 7).

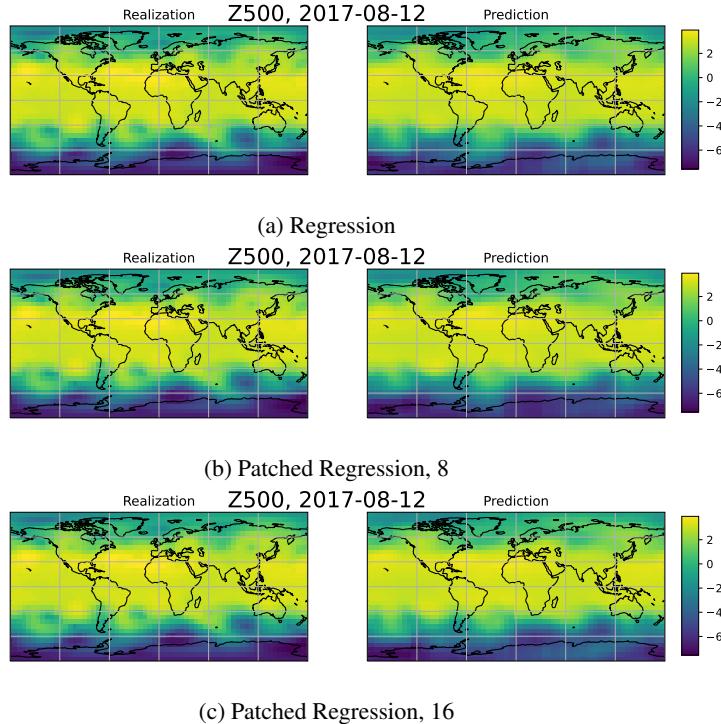


Figure 1: Realization and prediction obtained with the Regression and Patched Regressions for a specific date in the test set for the WeatherBench dataset.

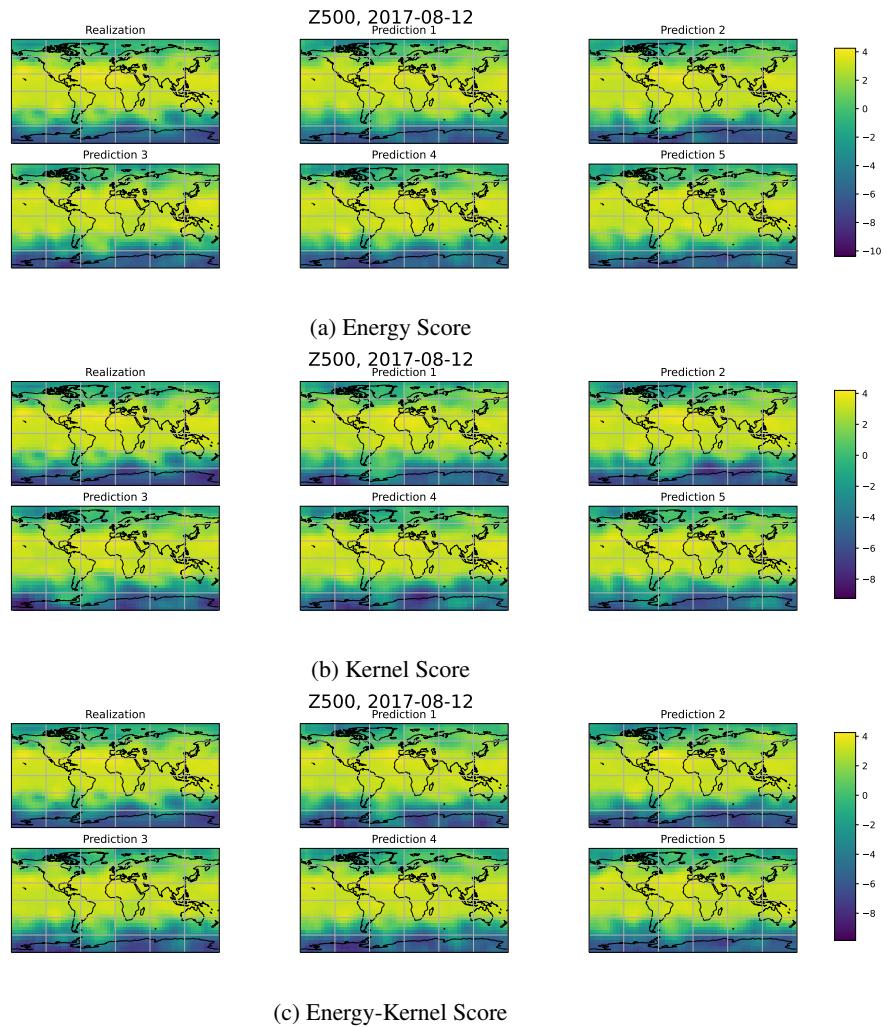


Figure 2: Realization and example of predictions obtained with the Energy, Kernel and Energy-Kernel Scores for a specific date in the test set for the WeatherBench dataset.

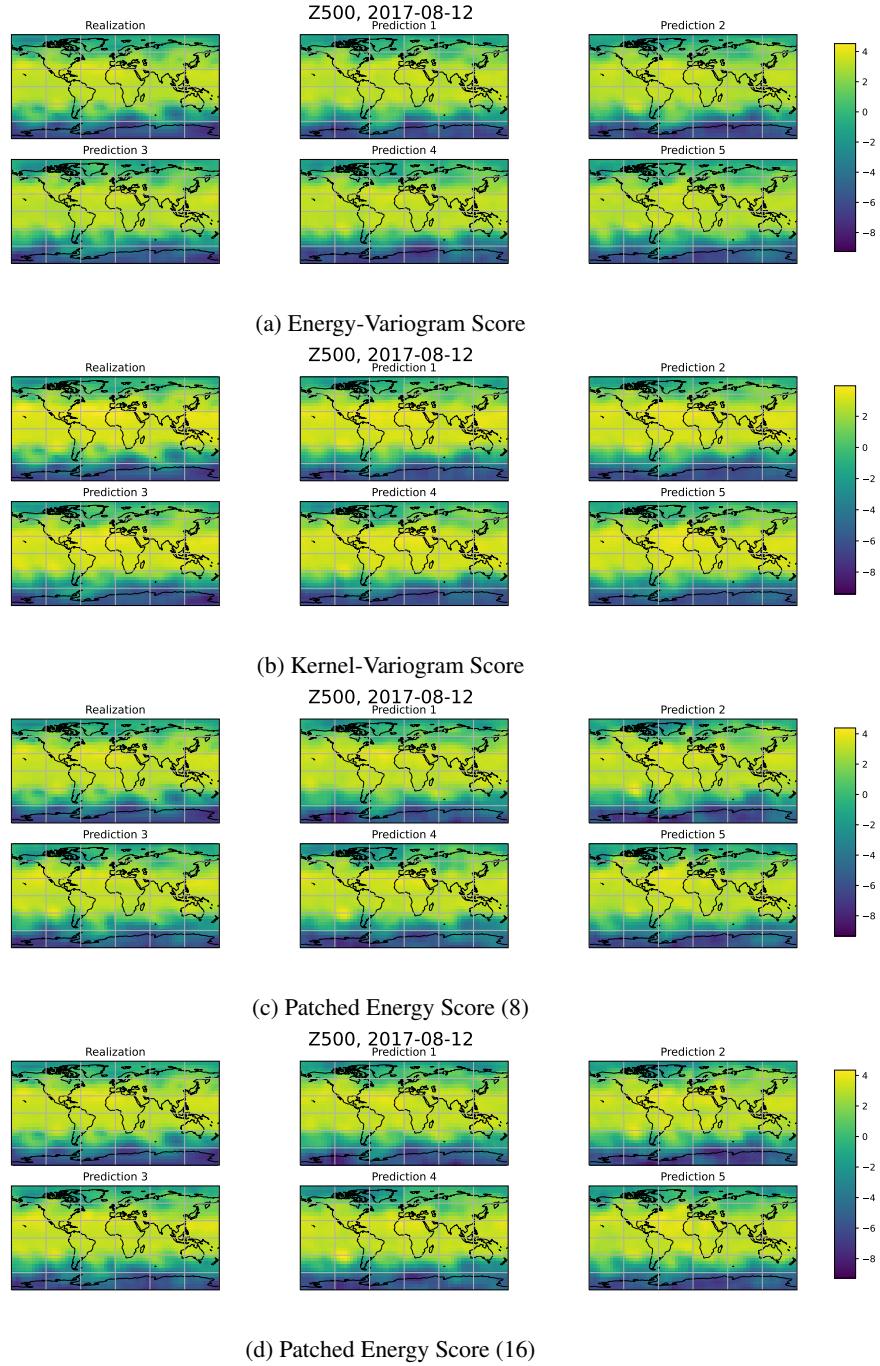


Figure 3: Realization and example of predictions obtained with the Energy-Variogram, Kernel-Variogram and Patched Energy Score (with patch size 8 and 16) for a specific date in the test set for the WeatherBench dataset.

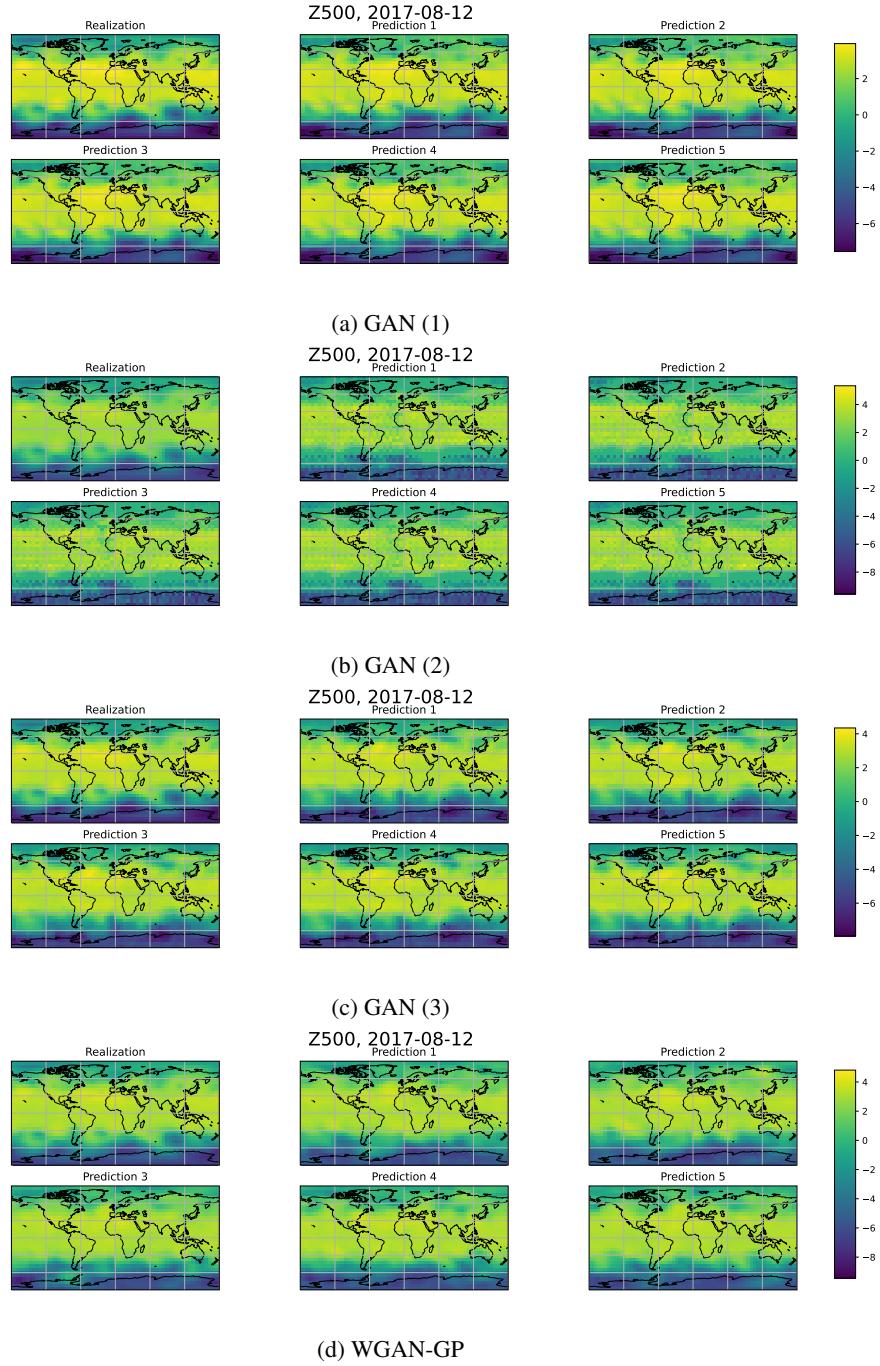


Figure 4: Realization and example of predictions obtained with the three considered GAN setups and WGAN-GP for a specific date in the test set for the WeatherBench dataset. Notice how the second GAN setup leads to unphysical features.

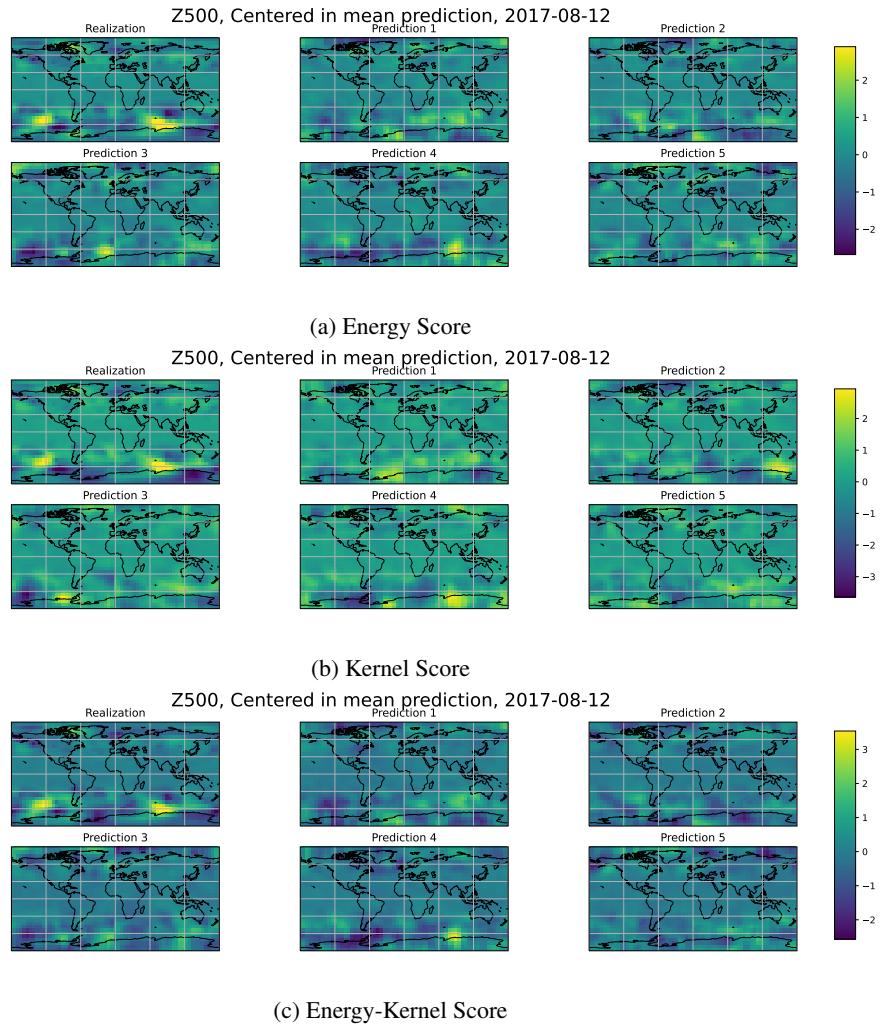


Figure 5: Deviations of the realization and forecasts from the forecast mean (obtained empirically from 100 forecasts) for the Energy, Kernel and Energy-Kernel Scores for a specific date in the test set for the WeatherBench dataset. The absolute values of the forecasts used here are shown in Figure 2.

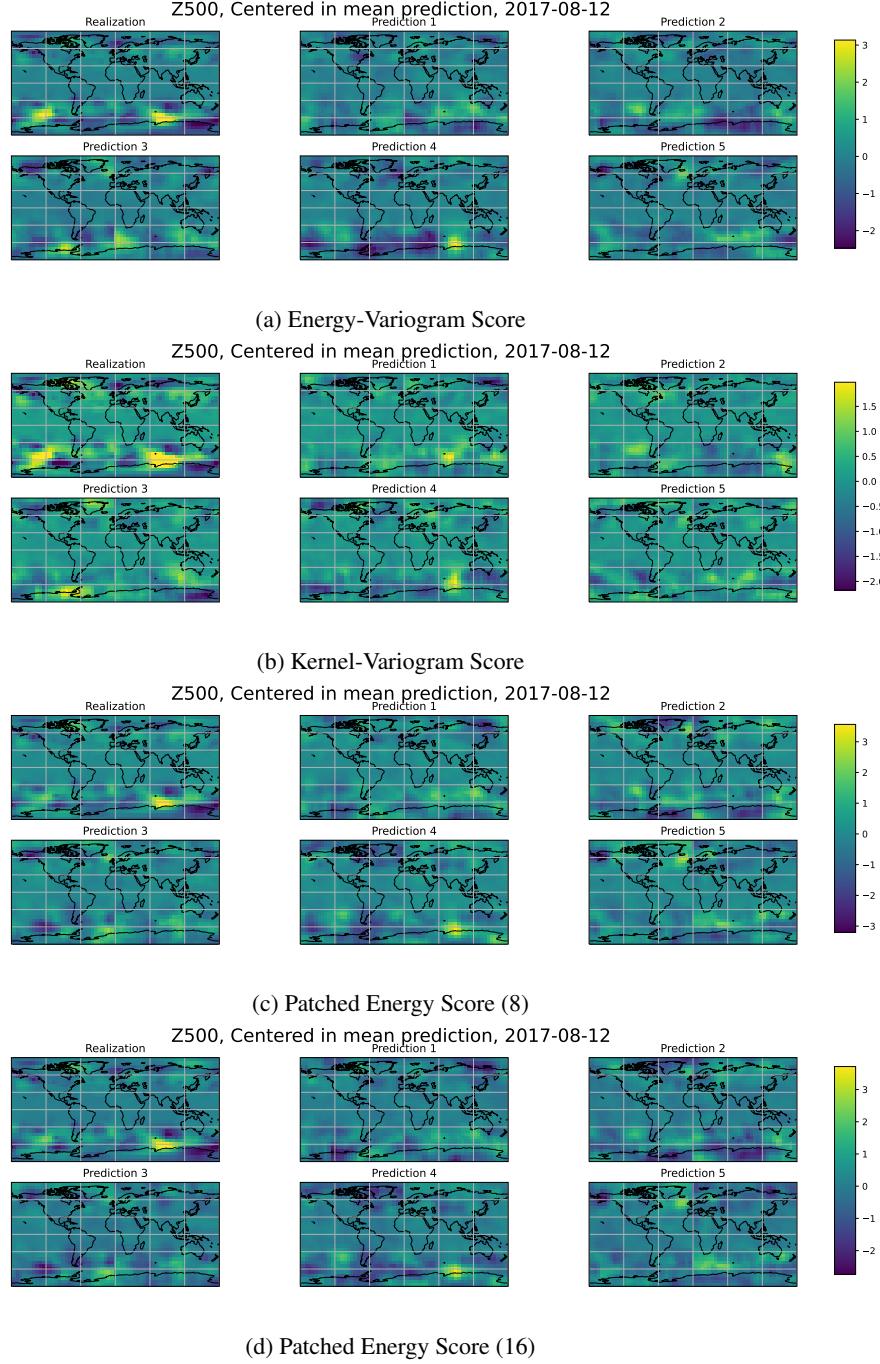


Figure 6: Deviations of the realization and forecasts from the forecast mean (obtained empirically from 100 forecasts) for the Energy-Variogram, Kernel-Variogram and Patched Energy Score (with patch size 8 and 16) for a specific date in the test set for the WeatherBench dataset. The absolute values of the forecasts used here are shown in Figure 3.

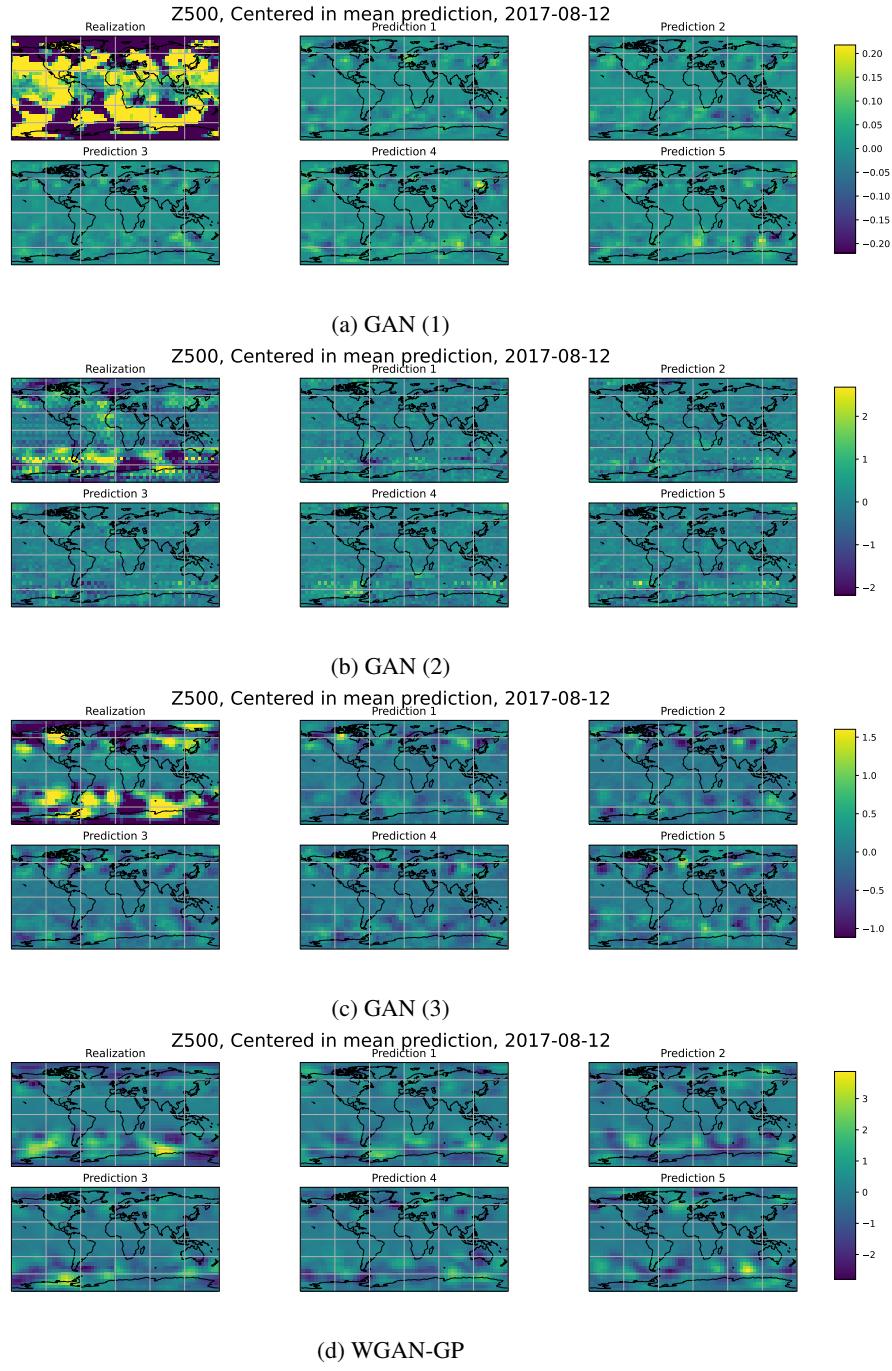


Figure 7: Deviations of the realization and forecasts from the forecast mean (obtained empirically from 100 forecasts) for the three considered GAN setups and WGAN-GP for a specific date in the test set for the WeatherBench dataset. Notice how, for the first and third GAN setups, the scale of variations with respect to the predictive mean to the realization is much larger for the realization than for the predictions. Instead, the second GAN setup leads to unphysical features. The absolute values of the forecasts used here are shown in Figure 4.