Deep Learning Methods for Modeling Precipitation in Cloud Resolving Simulations

Data Science Institute
COLUMBIA UNIVERSITY

Aashna Kanuga, Akshata Patel, Arjun Dhillon,
Dhananjay Deshpande, Hrishikesh Telang, Somendra Tripathi
Faculty Mentor: Pierre Gentine

Data Science Capstone Project
Pierre Gentine - Associate Professor, Earth Institute

Introduction

Current DNN based climate models suffer from generalization issues due to the inherent stochasticity of clouds and its highly sparse distribution. Our goal is to explore methods like CVAEs, GANs & LSTMs to alleviate these issues. The task is to predict the Precipitation based on Humidity (QBP), Temperature (TBP), Surface Pressure (PS), Solar Insolation (SOLIN), Sensible Heat Flux (SHFLX) and the Latent Heat Flux (LHFLX).

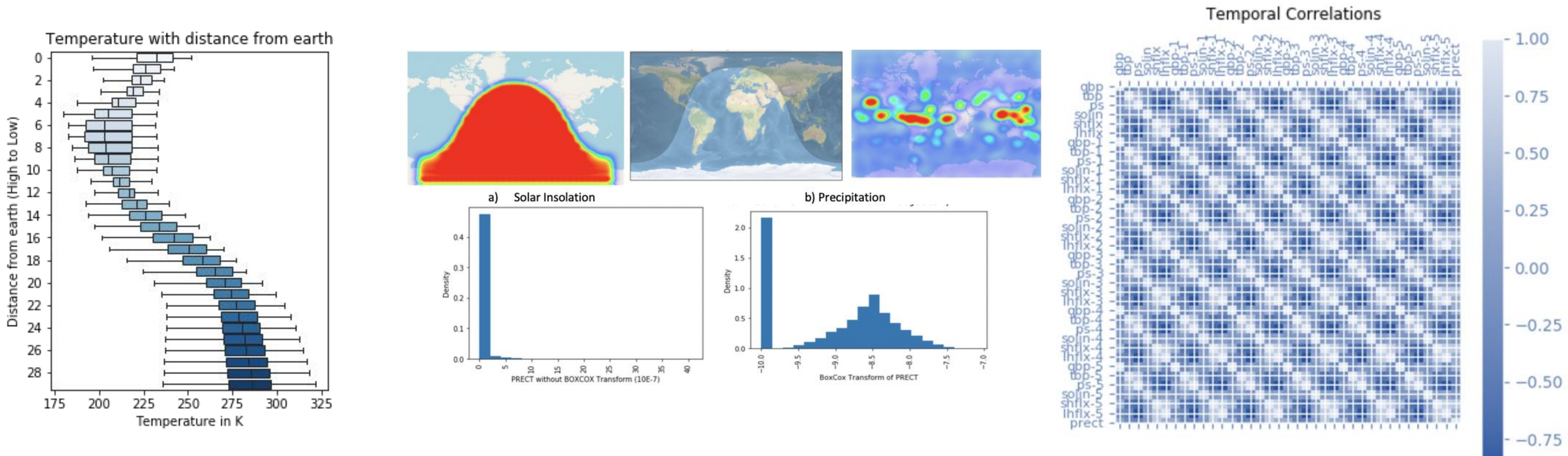


Figure 1. Exploratory analysis of Climate Variables

Model Design

We tackle the above issues by implementing the following three approaches:

- Generative Adversarial Networks to be able to better represent the distribution of precipitation.
- Conditional Variational Autoencoder to add stochasticity to precipitations predictions.
- Long Short Term Memory Models to capture temporal dependencies between features that contribute to PRECT.

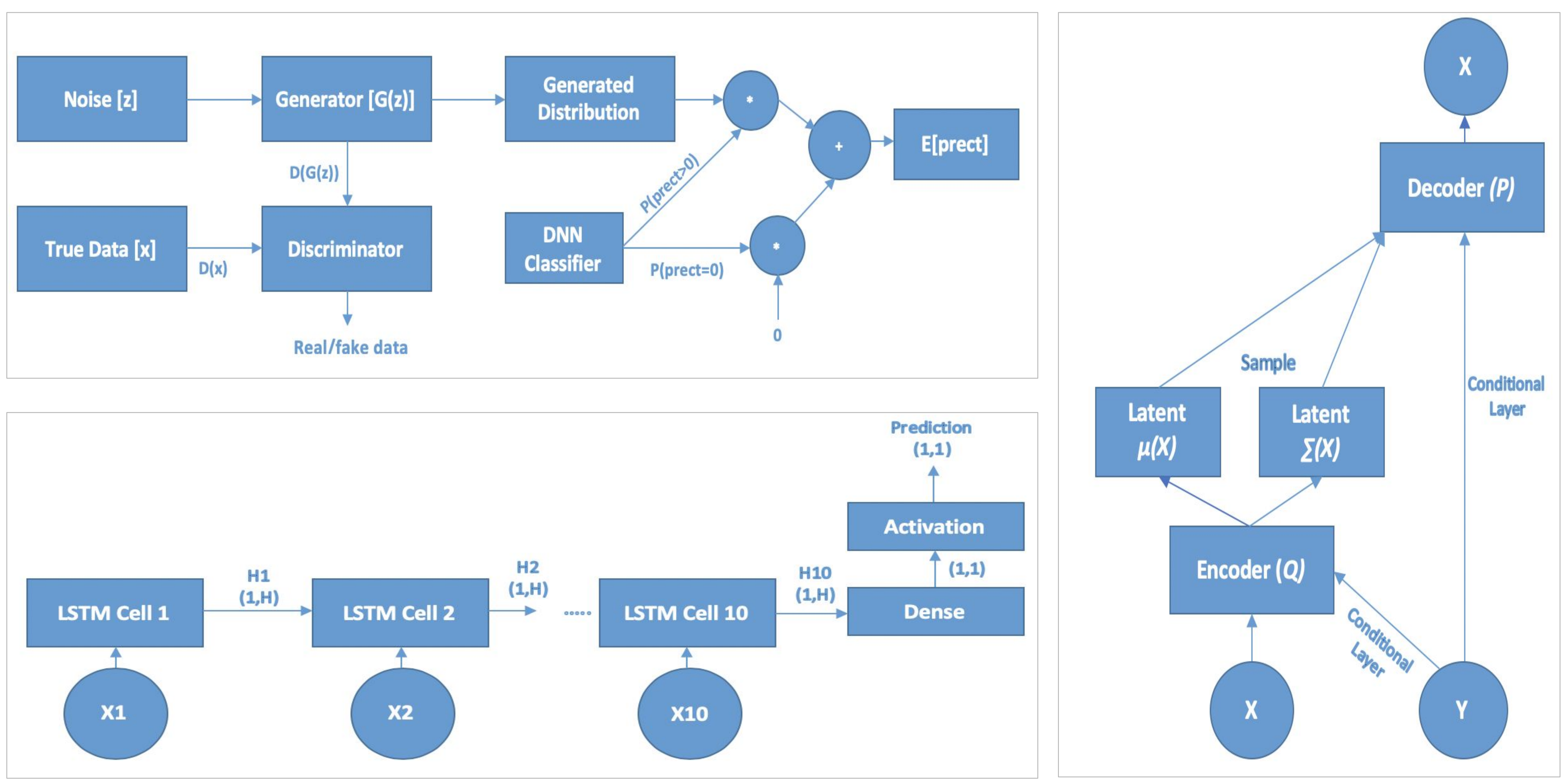
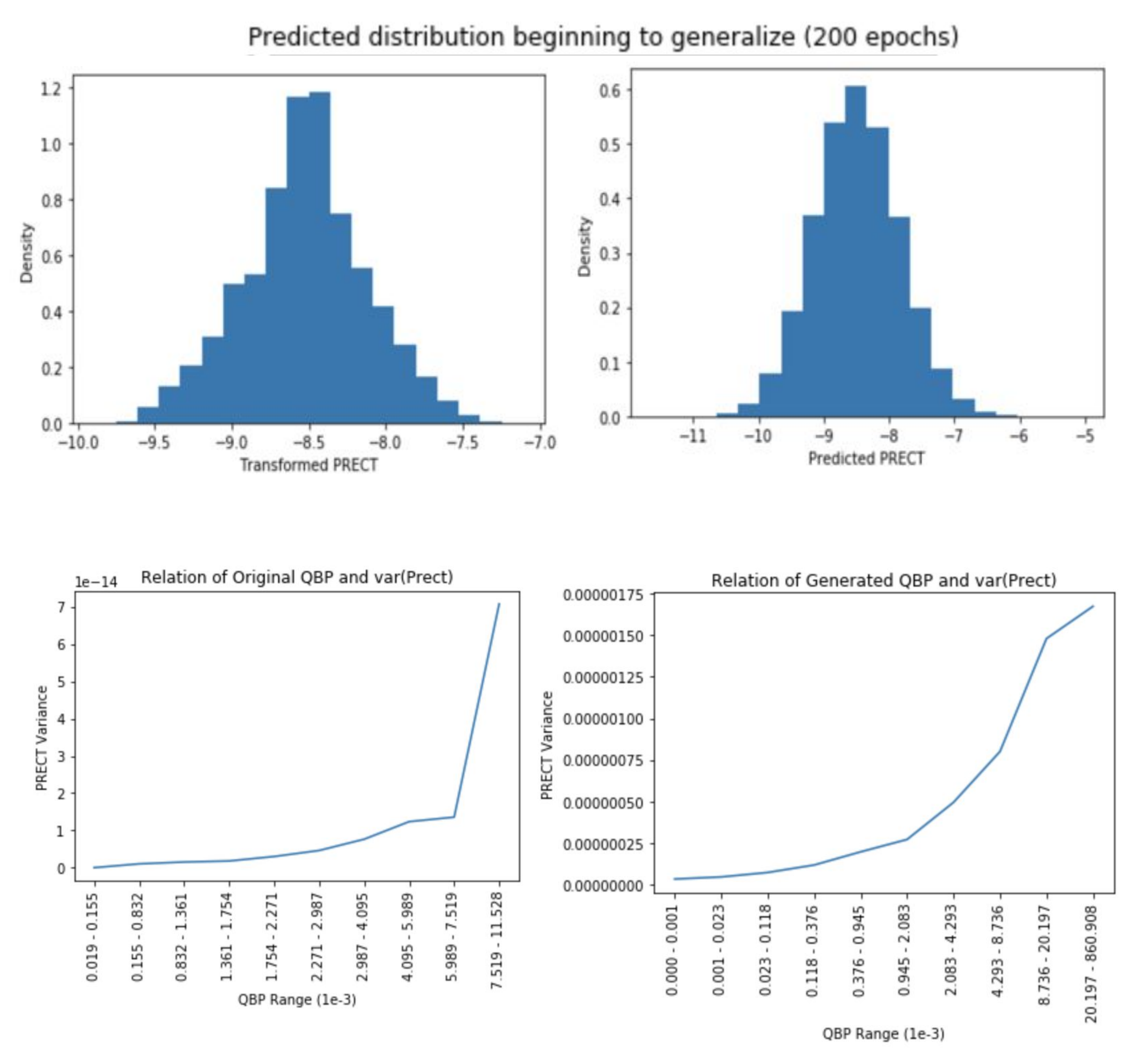
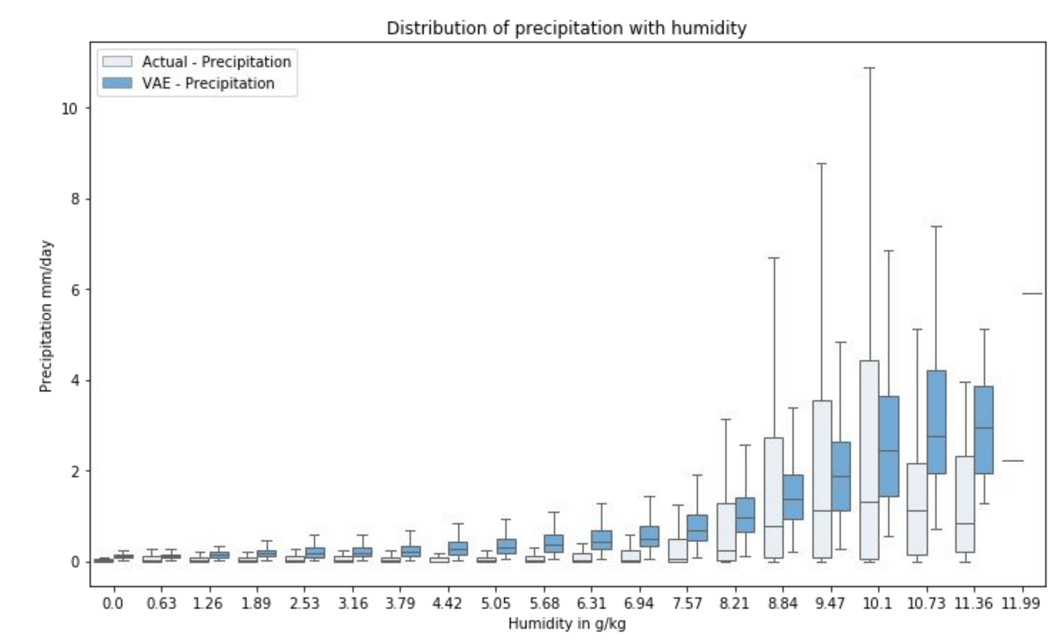


Figure 2. Model Architectures for GANs (top-left), LSTM (bottom-left) and CVAE (right)

Results

- Training the GAN for 200 epochs, we find that it has shown some limited abilities to learn the characteristics of the data, such as the relationship between the variance of PRECT and QBP, and the approximate range of the variables.
- CVAE learns a generative model capturing the high variability of precipitation at the extremes while maintaining similar performance as baseline DNN.
- LSTM models with different batch sizes and cell units were tested, but while it predicts constant values a with low MSE, it cannot model variability





PRECT	Actual	LSTM (200 units)	LSTM (10 units)	LSTM (1 unit)
Count	5.80E+06	5.80E+06	5.80E+06	5.80E+06
MSE	0.0	2.09E-11	7.71E-10	5.55E-14
Mean	2.81E-08	4.60E-06	2.77E-05	2.43E-07
STD DEV	9.66E-08	0.00E+00	0.00E+00	0.00E+00

Figure 3. Results for GAN (left), CVAE (top-right) and LSTM (bottom-right)

Conclusion

Overall, some models show promising results, while the others still need to be fine-tuned

- GAN needs to be trained for much longer (order of thousands) in order to give sufficiently accurate results.
- Conditional VAE approach showed promise in capturing the higher-order statistics of precipitation in climate simulations.
- LSTM models were not able to find patterns using 5 hours of data inputs per prediction and more experimentation is needed with larger dataset for better results.

Acknowledgments

We would like to extend our sincere gratitude to Prof. Gentine, and his entire team for their continuous guidance and support throughout the course of this project.

References

Stephan Rasp et al (2018) - Deep learning to represent subgrid processes in climate models

P. Gentine et al (2018) - Could Machine Learning Break the Convection Parameterization Deadlock?