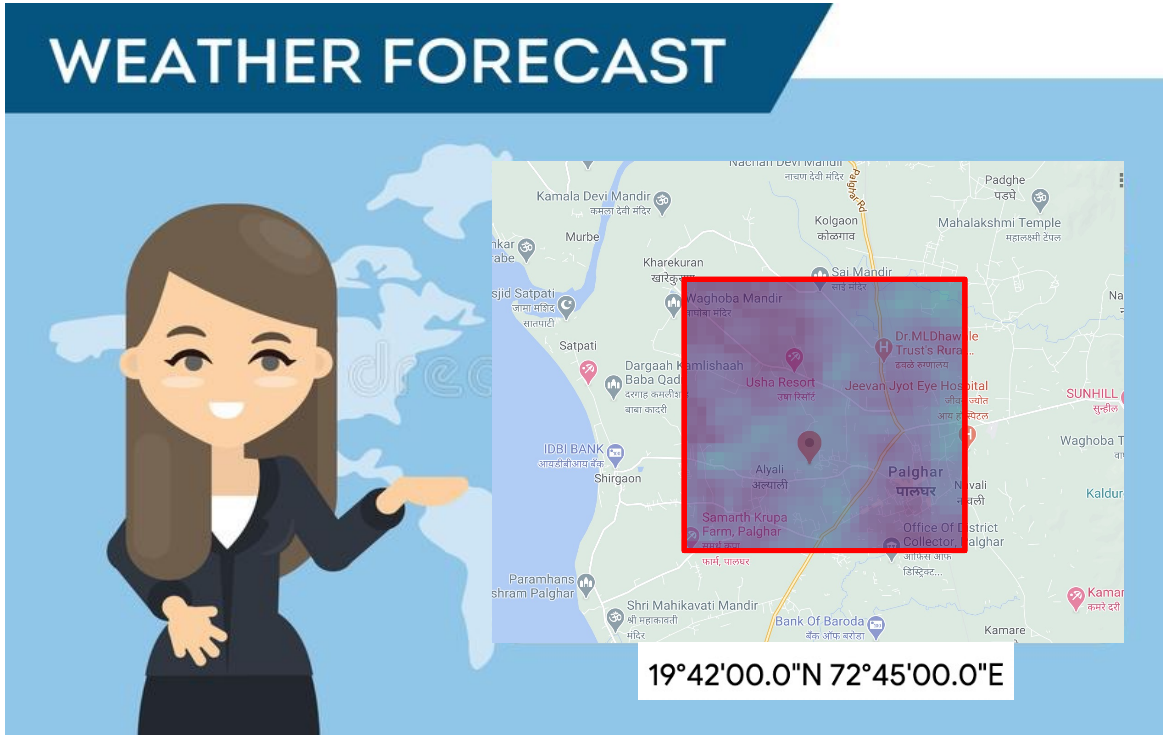
Dario Oliveira

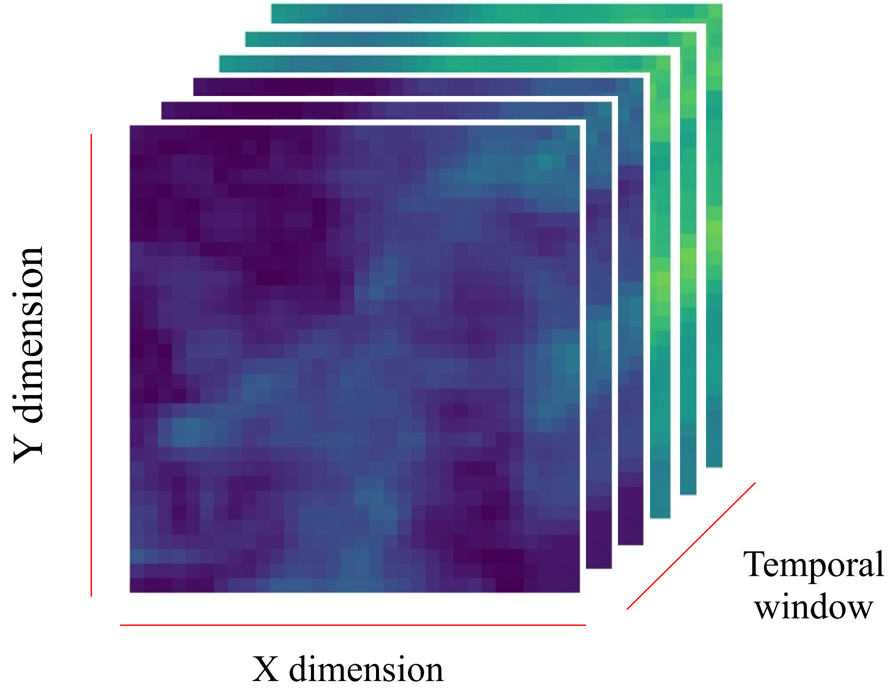
Controlling Weather Field Synthesis Using Variational Autoencoders

Dario Oliveira, Jorge Guevara, Bianca Zadrozny, Campbell Watson

Weather Fields Data

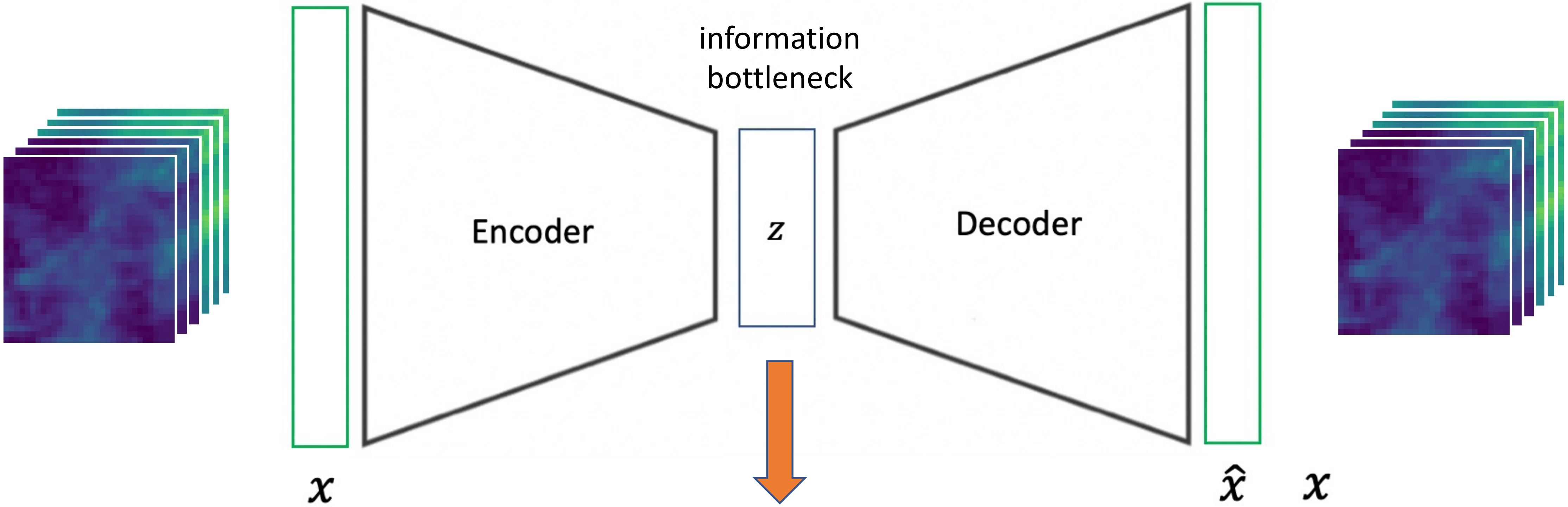
- Weather fields are the multidimensional representations of spatially distributed weather variables, like temperature, wind, precipitation, etc.





Auto-Encoders

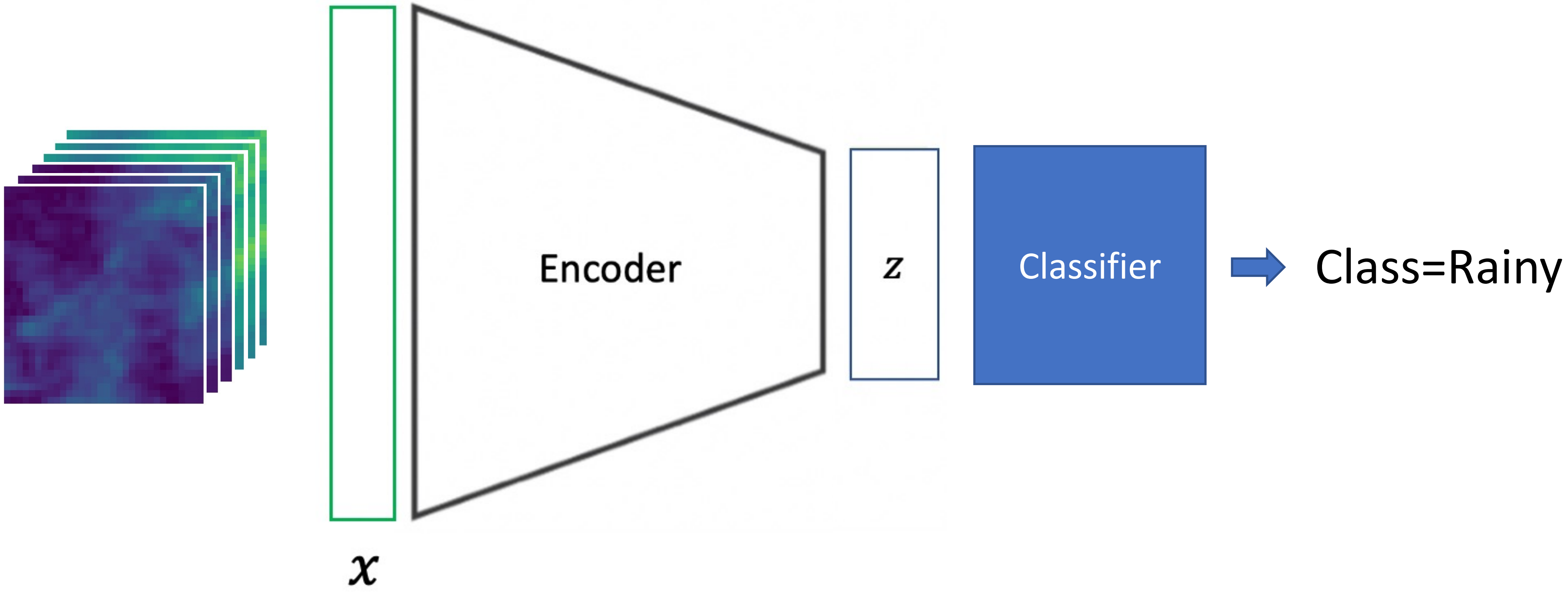
Auto-encoders: models that learn to reconstruct the input data



Z is a very compact, efficient and
discriminant representation of the input

Variational Auto-Encoders - VAE

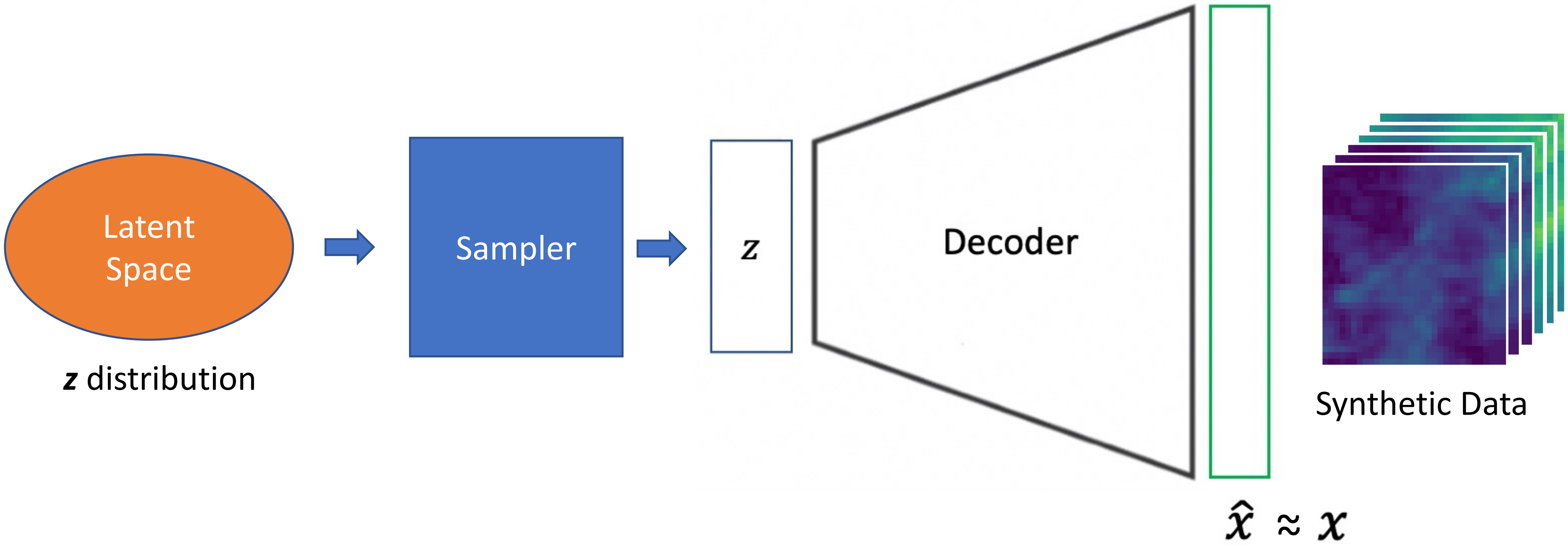
Auto-encoders: models that learn to reconstruct the input data



z can be used for efficient feature extraction

Variational Auto-Encoders - VAE

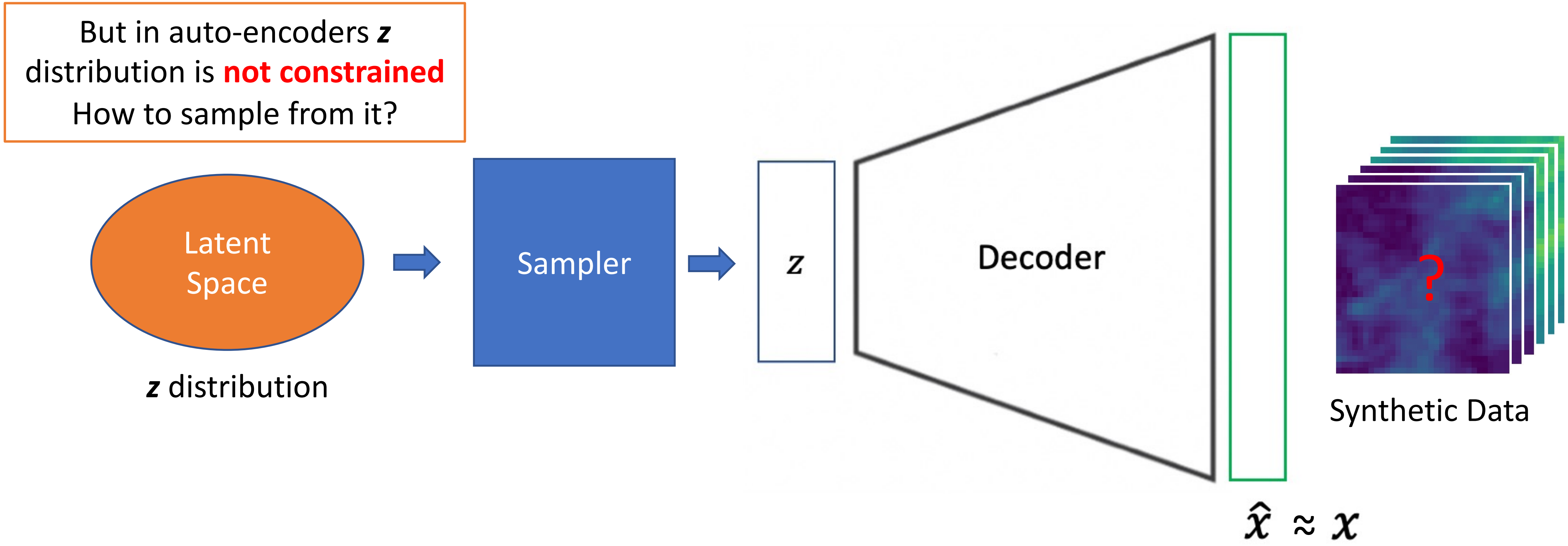
Auto-encoders: models that learn to reconstruct the input data



z can be used for <u>stochastic synthesis</u>

Variational Auto-Encoders - VAE

Auto-encoders: models that learn to reconstruct the input data



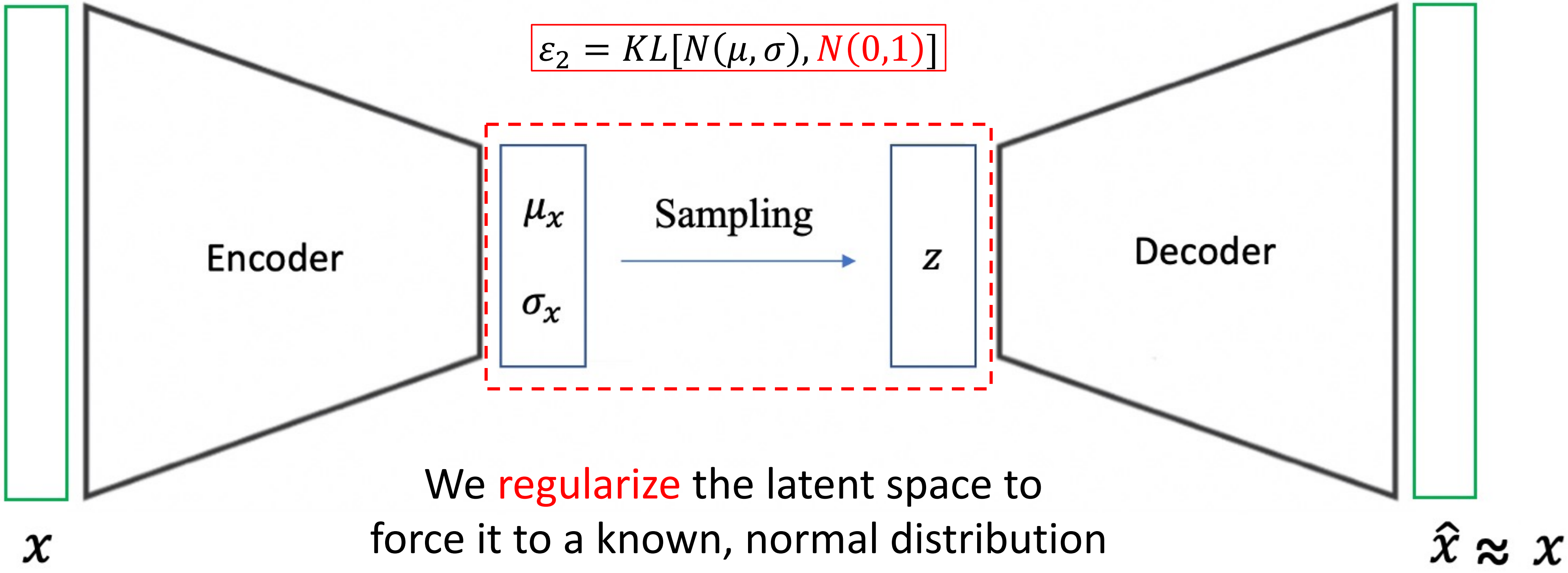
z can be used for <u>stochastic synthesis</u>

Variational Auto-Encoders - VAE

Variational Auto-Encoders: constraining latent space distribution

$\varepsilon_1 = \|x - \hat{x}\|$

$\varepsilon_2 = KL[N(\mu, \sigma), N(0,1)]$

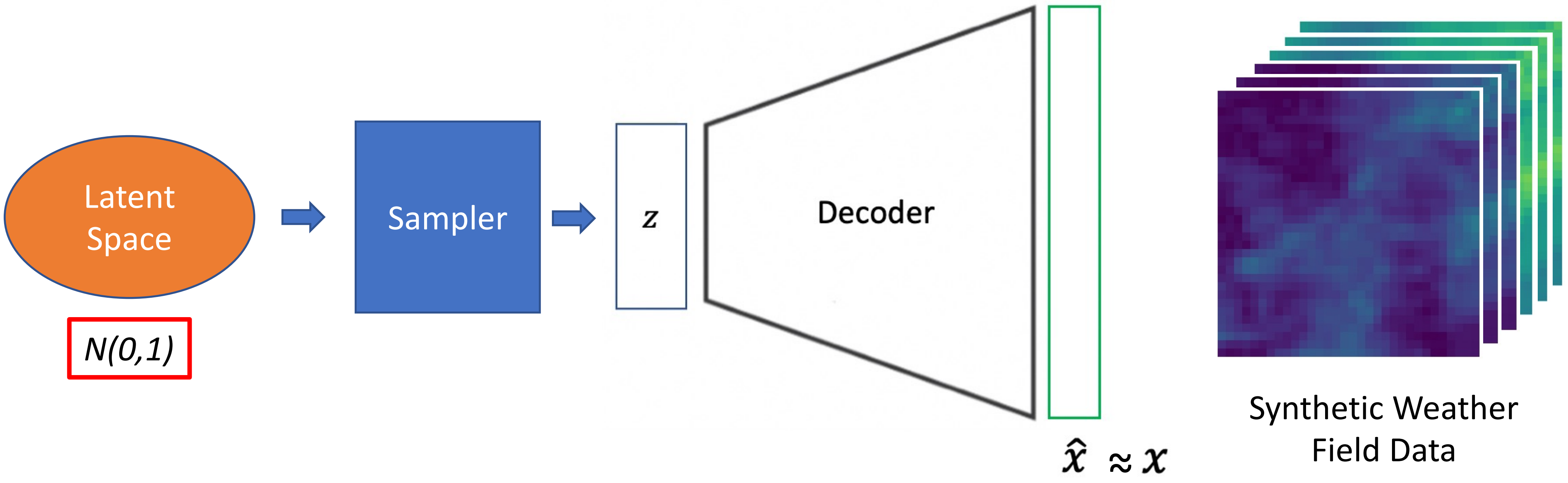


We regularize the latent space to force it to a known, normal distribution

Now, we can sample from *N(0,1)* for realistic synthesis!

Controling Weather Field Data Synthesis

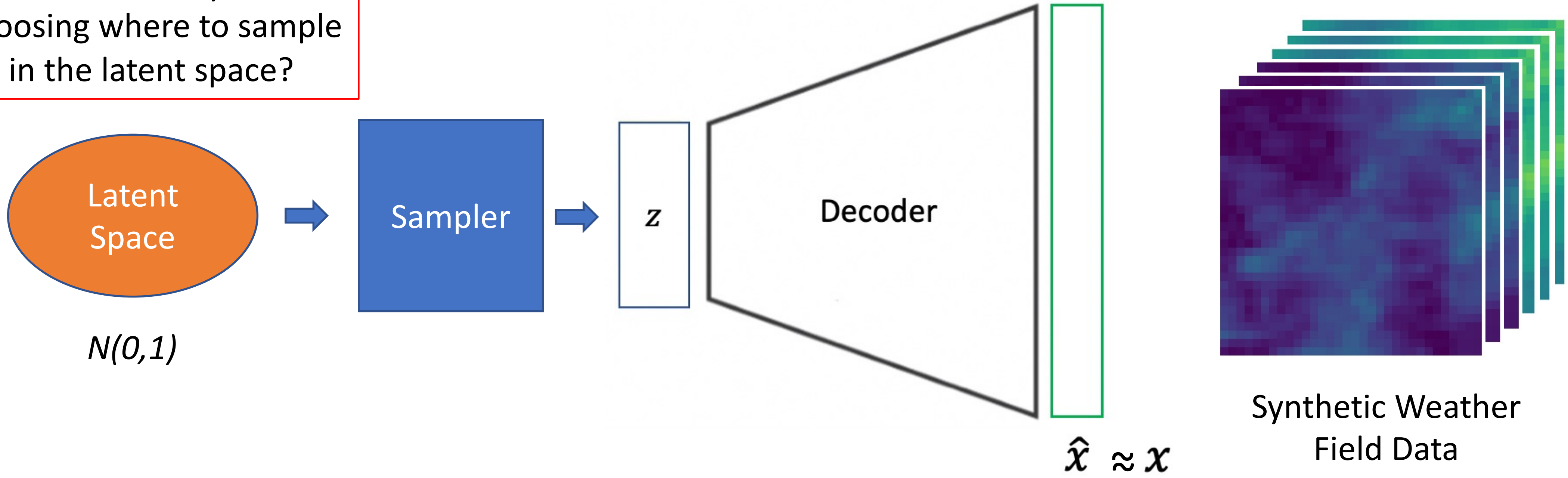
Using trained VAE weather field data synthesis



Controling Weather Field Data Synthesis

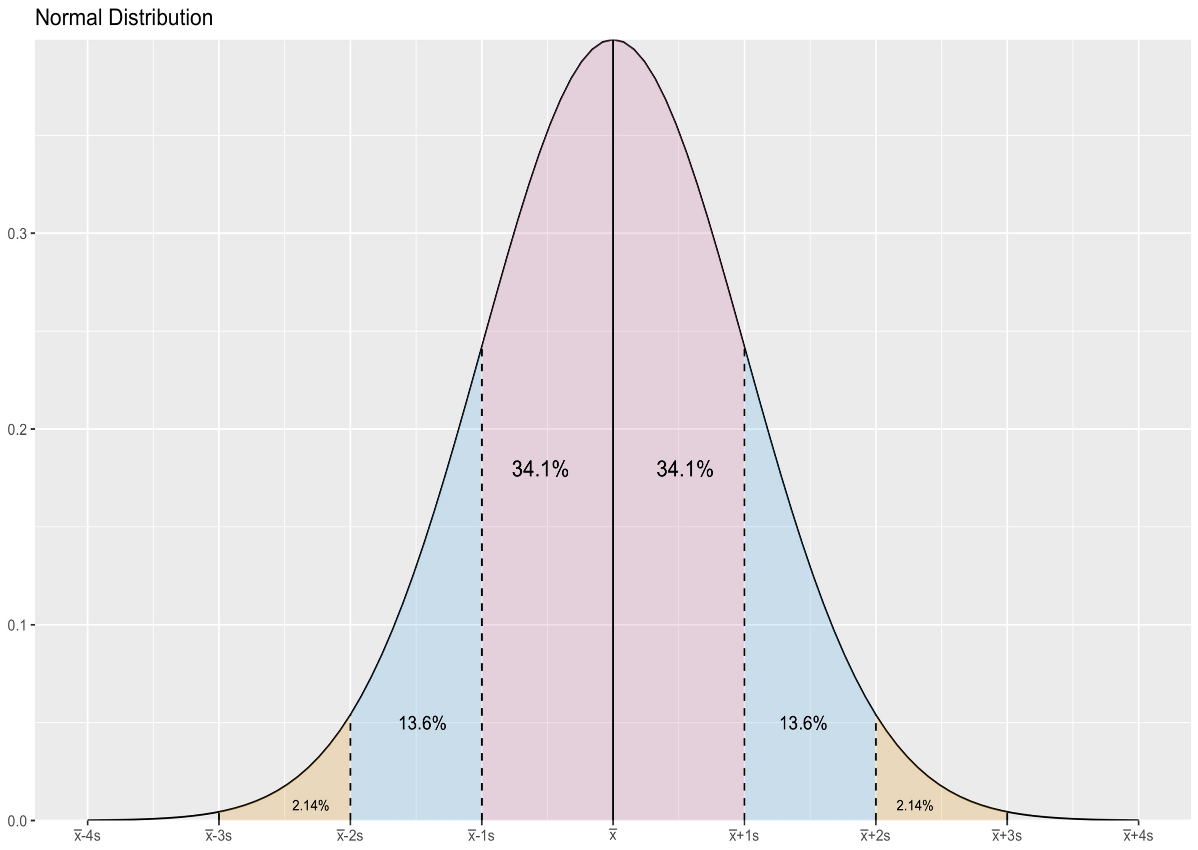
Using trained VAE weather field data synthesis

Can we control synthesis choosing where to sample in the latent space?

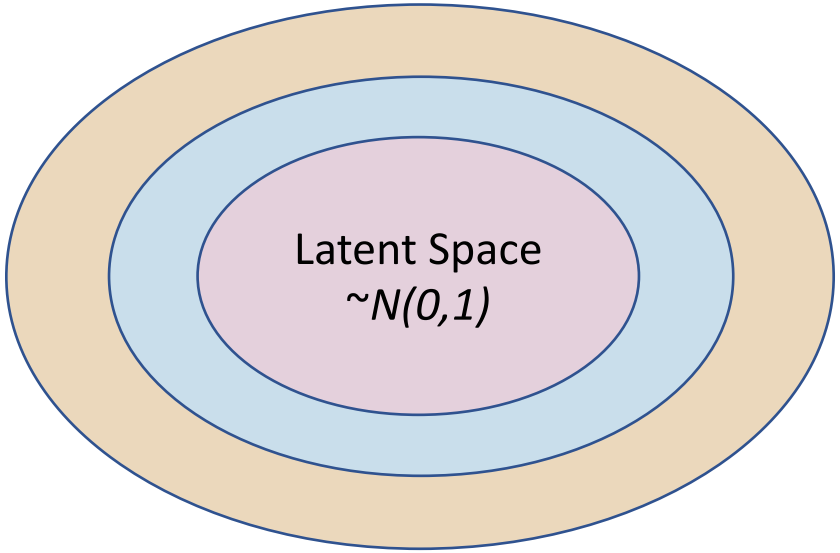


Controling Weather Field Data Synthesis

How is the VAE latent space distribution? ~ *N(0,1)*

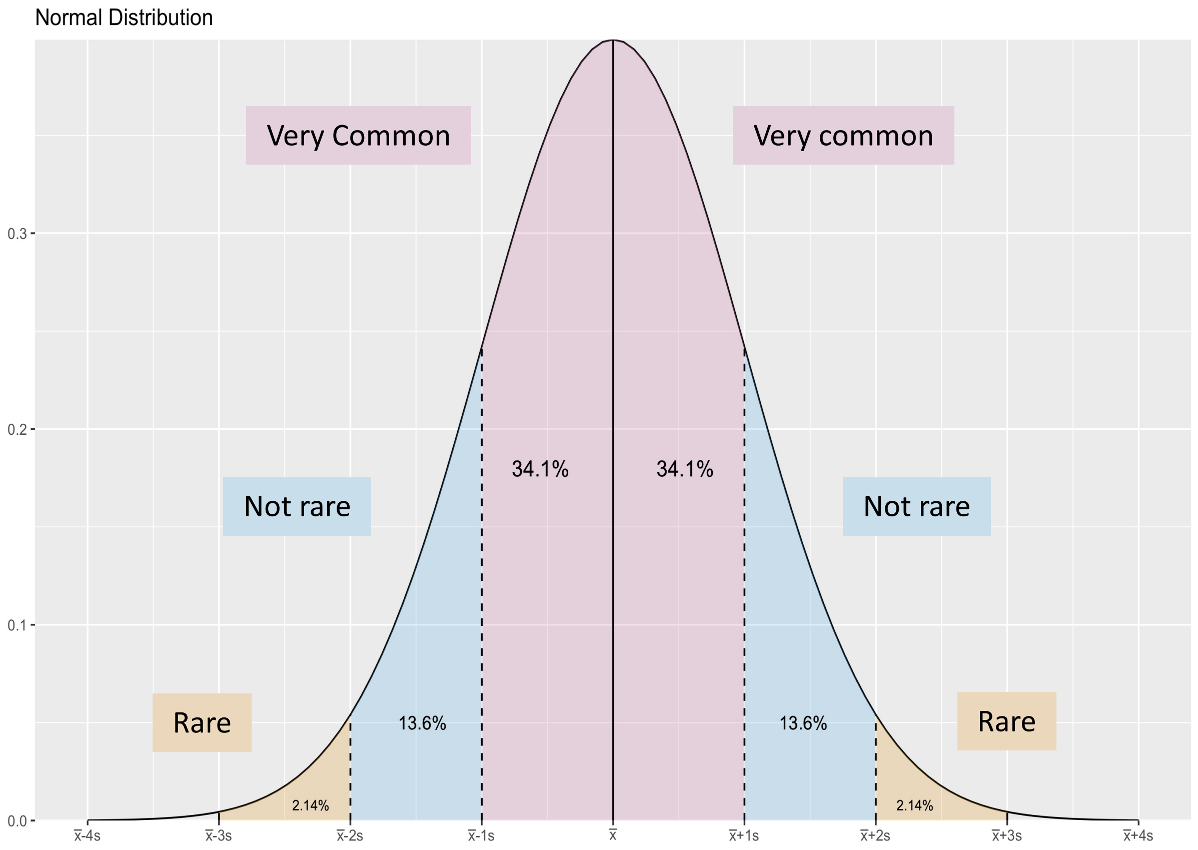


But it is N-dimensional

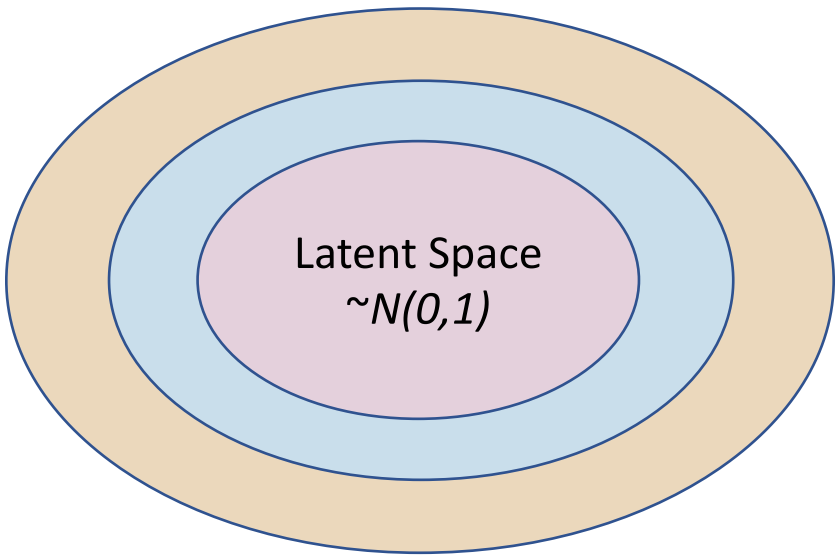


Controling Weather Field Data Synthesis

How are climate events distributed in the latent space ~ *N(0,1)*?

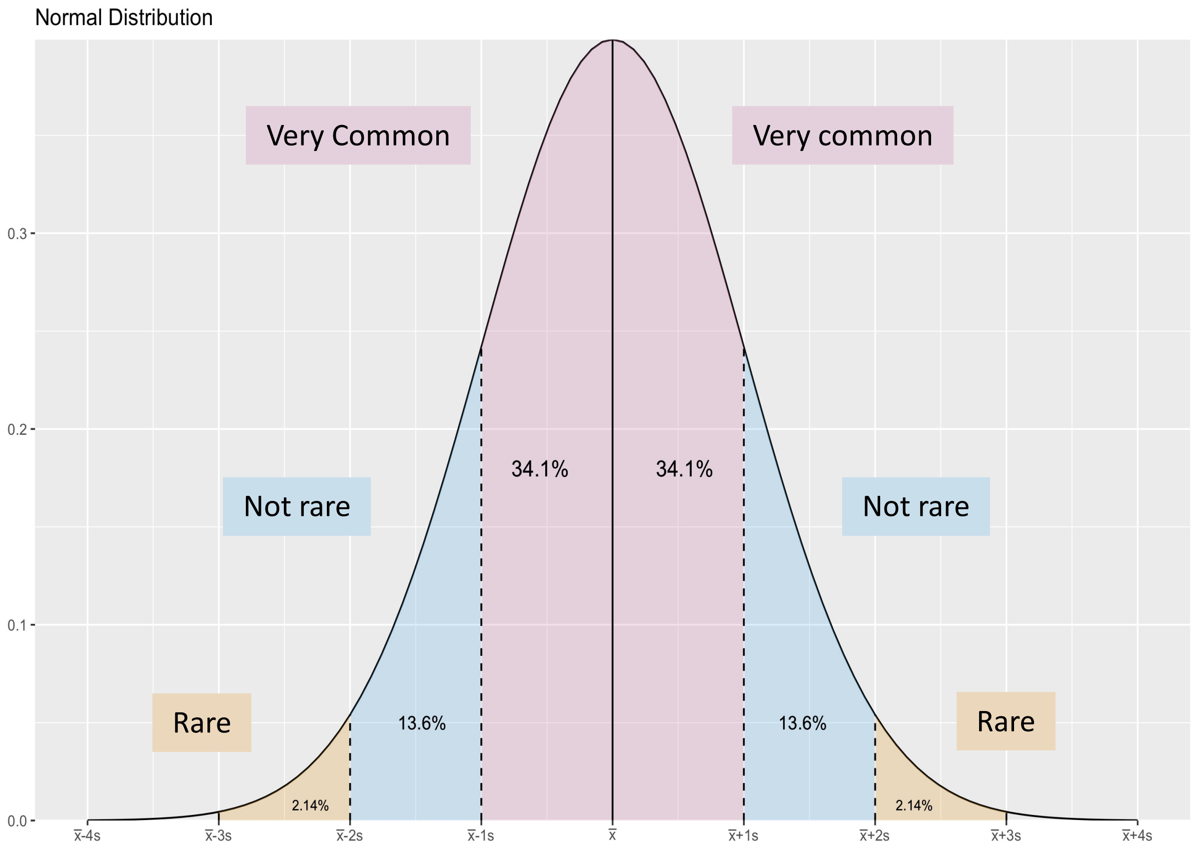


Very common events will necessarily be located near the distribution mean and rare events will be located far from it

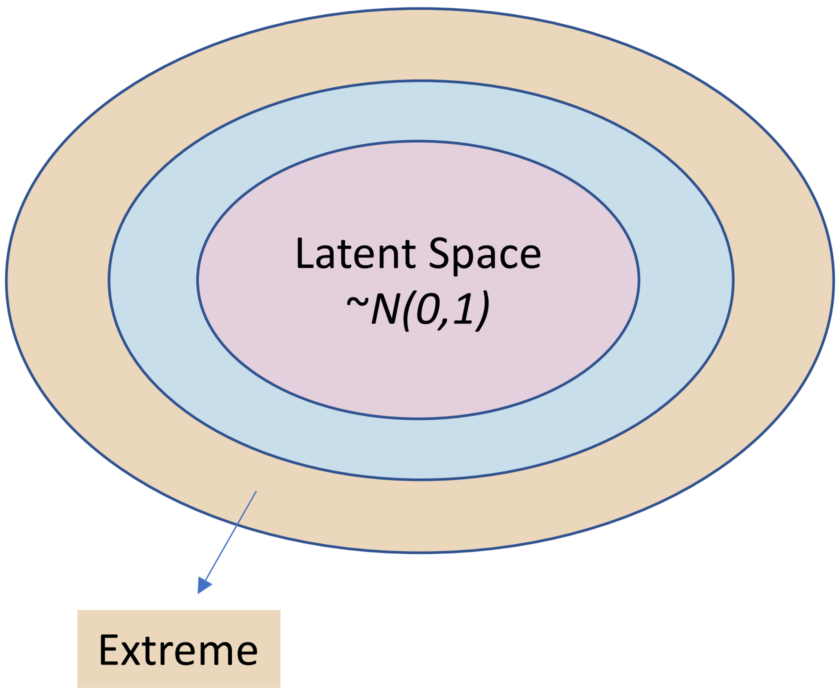


Controling Weather Field Data Synthesis

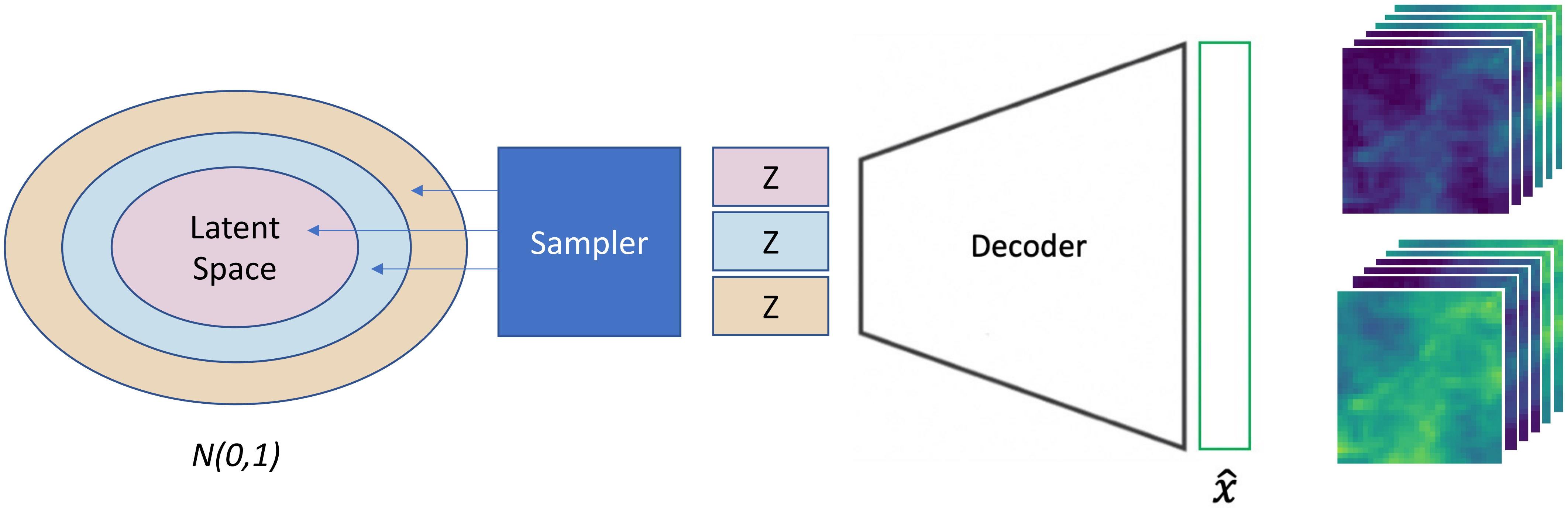
How are climate events distributed in the latent space ~ *N(0,1)*?



But extreme weather events
are also usually rare!



Controling Weather Field Data Synthesis



Controlling Weather Field Synthesis

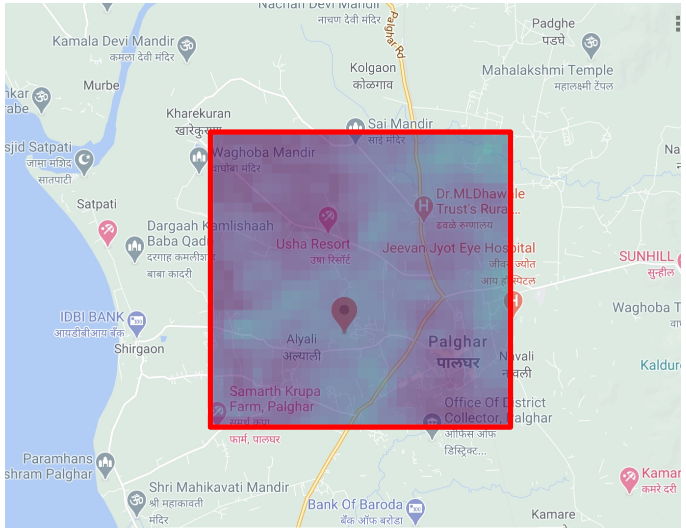
Z
Z
Z

more extreme (rare) weather fields

We don not need to define exactly what extreme means!
extreme=rare

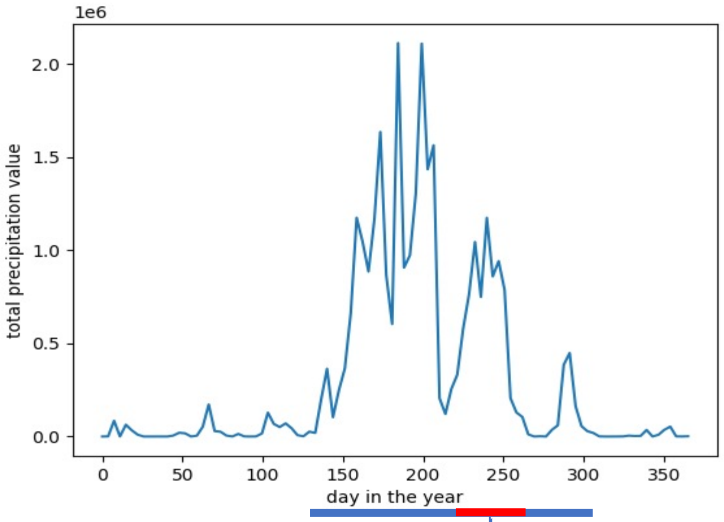
Use Case: Palghar Monsoons

Monsoons in Palghar, Southwest India

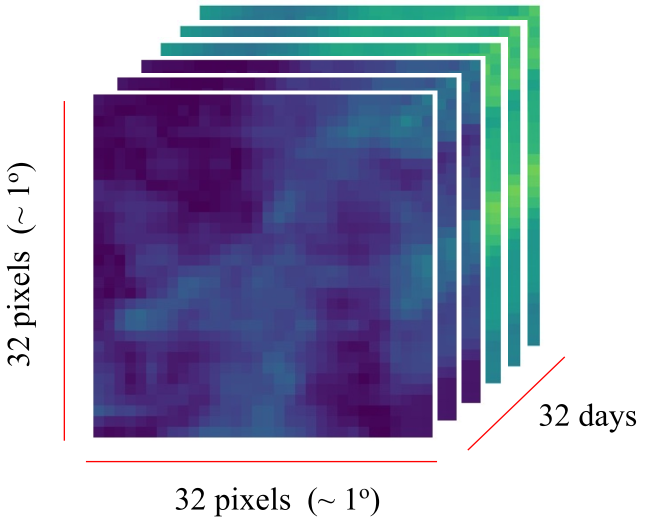


19°42'00.0"N 72°45'00.0"E

Precipitation in a random year



Weather Field Sample



CHIRPS dataset[1] from Palghar, India

39 years of data: 1981-01-01 to 2020-01-01

Training: 1981-2010

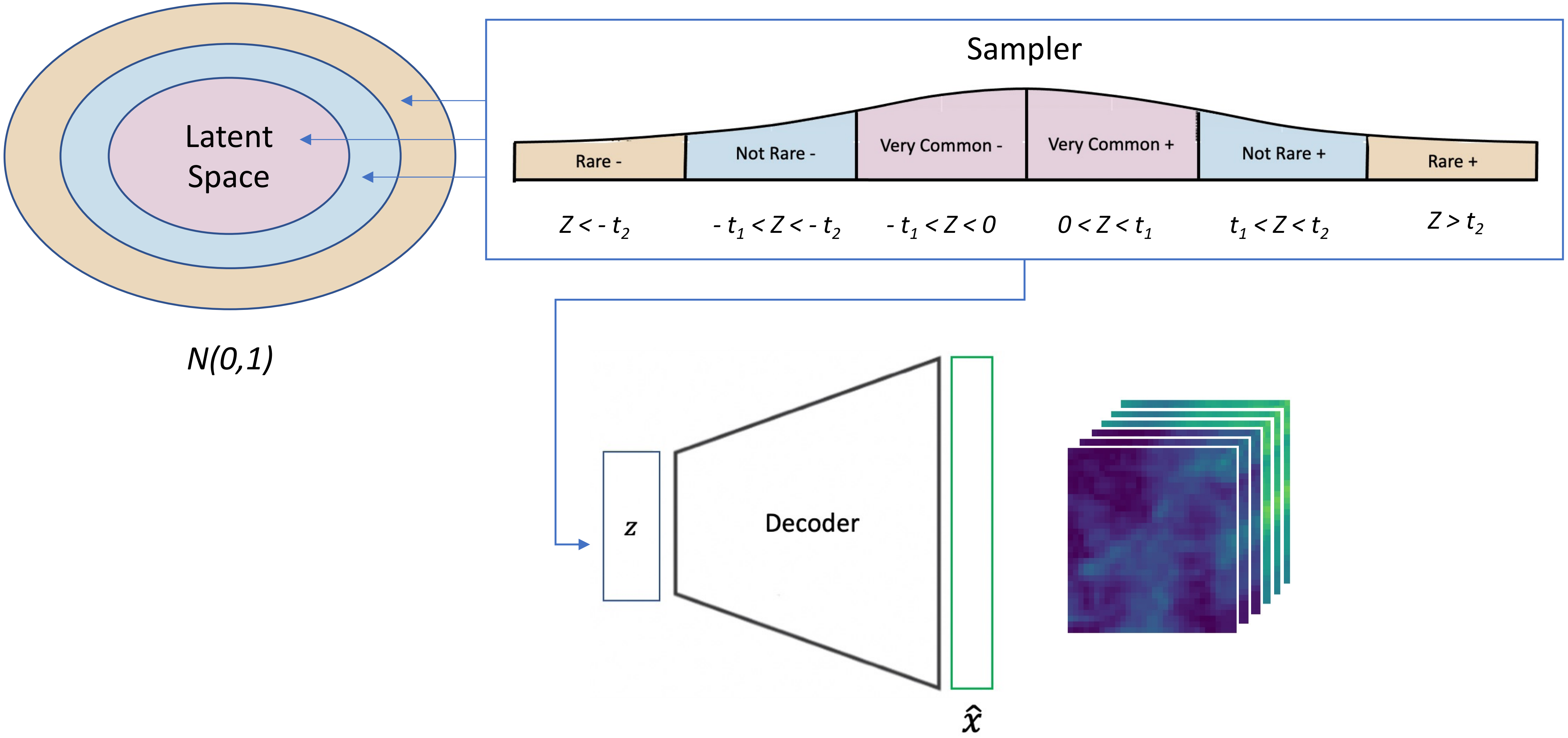
Testing: 2010-2020

sampling sequences of 32 days

[1] Chris Funk, Pete Peterson, Martin Landsfeld, Diego Pedreros, James Verdin, Shraddhanand Shukla, Gregory Husak, James Rowland, Laura Harrison, Andrew Hoell, et al. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific data*, 2(1):1–21, 2015.

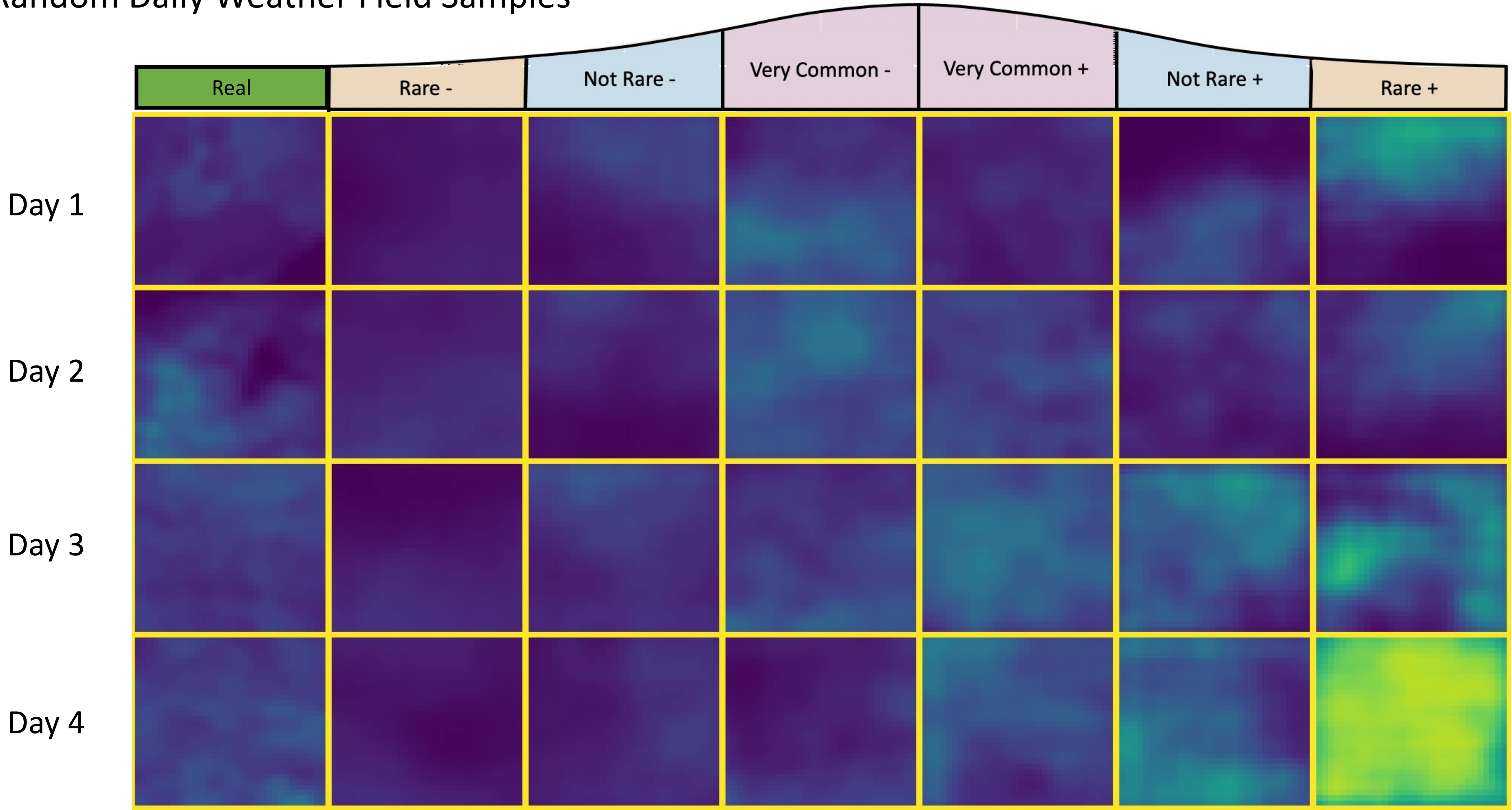
Use Case: Palghar Monsoons - Results

Experiment: with a trained variational autoencoder, control the sampling of *Z* based on *N(0,1)* quantiles

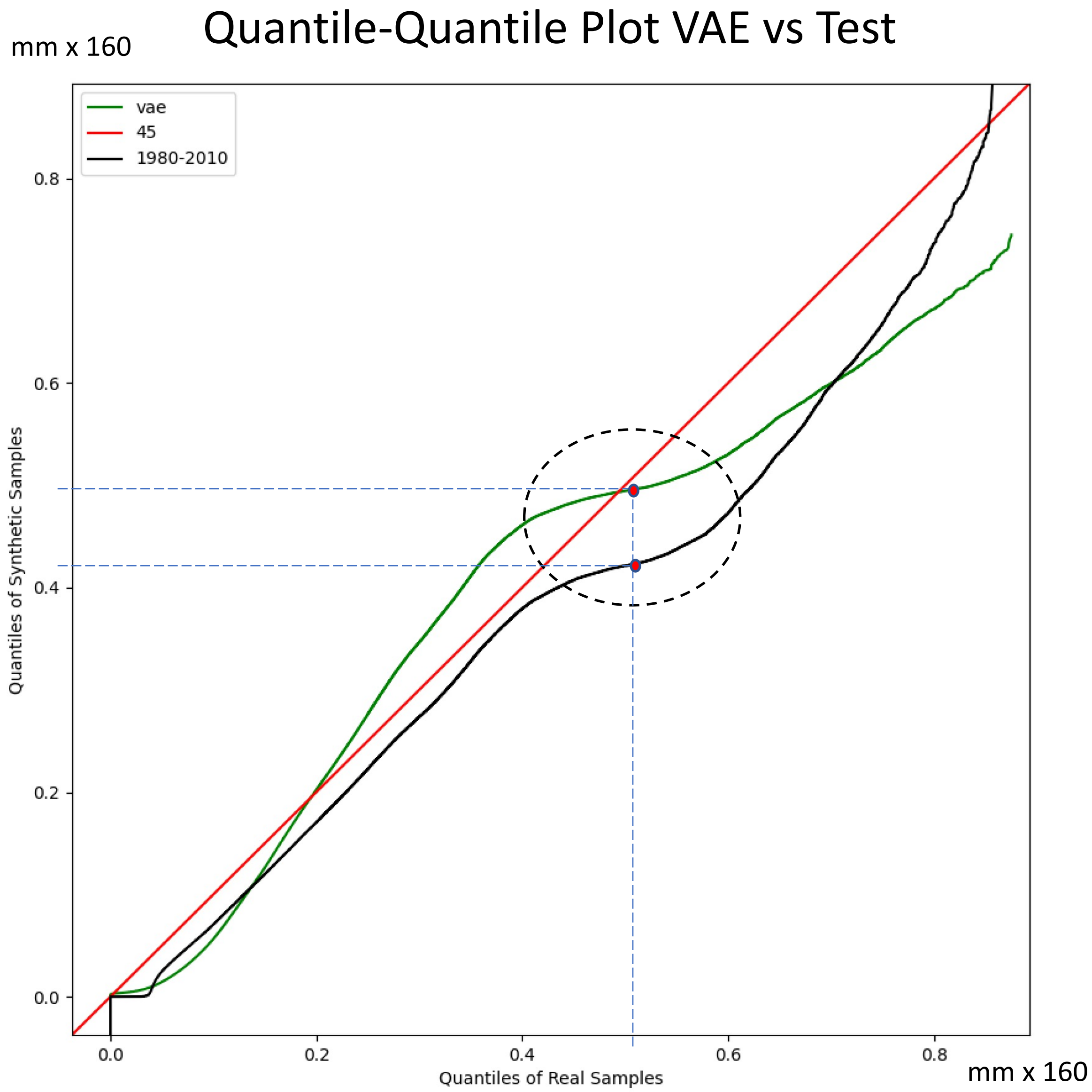


Use Case: Palghar Monsoons - Results

Random Daily Weather Field Samples



Use Case: Palghar Monsoons - Results



QQ-Plots compare distributions, where each point in the curves is the correspondence of their quantiles

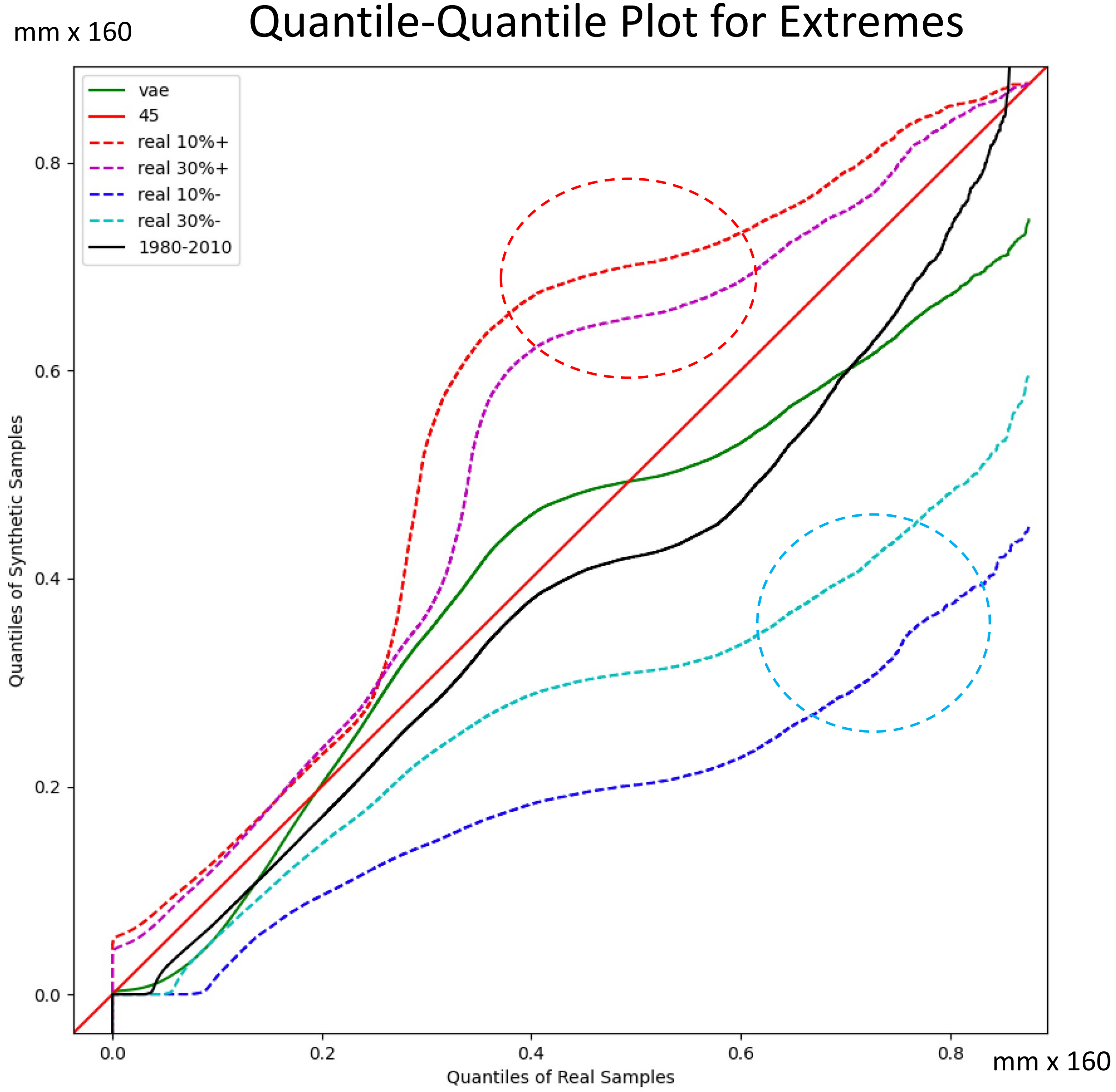
- In the highlighted area, the amount of 80mm rain observed in test set was related to around 65mm in the historical data, and to around 80mm in the synthetic VAE data.

Historical data as a predictor underestimates precipitation for higher quantiles, which means that strong rain events increased comparing 1980-2010 to 2010-2020.

VAE synthetic data as a predictor is a bit better than historical data for above average values and a bit worse for the highest ones.

Use Case: Palghar Monsoons - Results

Quantile-Quantile Plot for Extremes

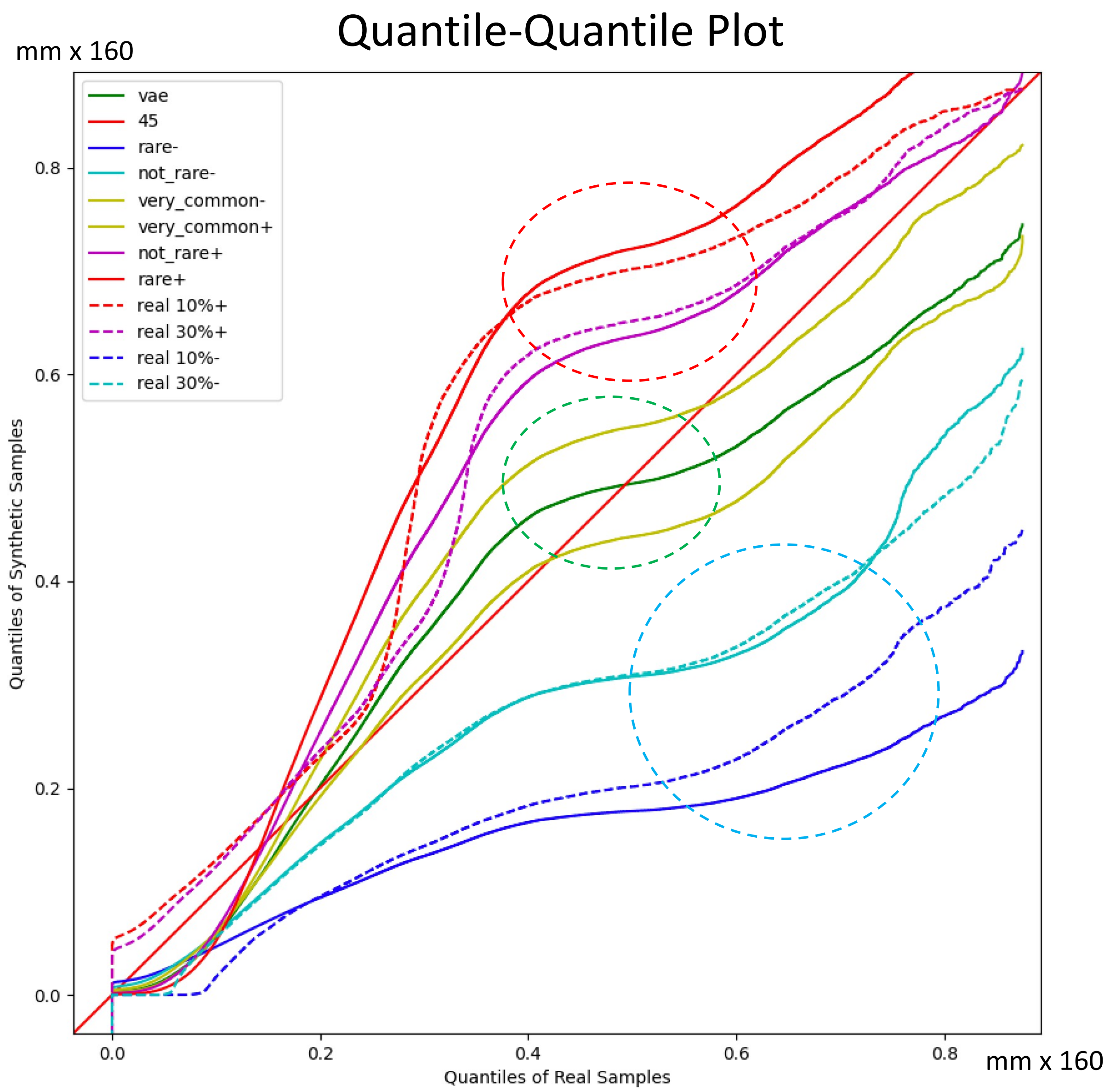


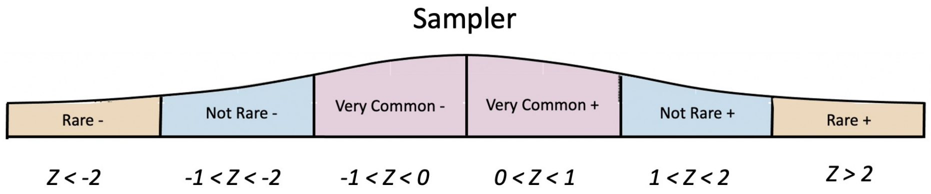
The 10% and 30% samples with more acumulated precipitation in the test set have considerably higher quantiles compared to overall test data (2010-2020)

The 10% and 30% samples with less acumulated precipitation in the test set have considerably lower quantiles compared to overall test data (2010-2020)

QQ-plots can be used for evaluating extreme events compared to the regular distribution

Use Case: Palghar Monsoons - Results



Synthetic samples for higher extremes are coherent with real test data for the 10% and 30% samples with the highest total monthly precipitation

Synthetic common scenarios are the ones closer to the overal test precipitation data distribution

Synthetic samples for lower extremes are coherent with real test data for the 10% and 30% samples with the lowest total monthly precipitation

Final Remarks

This paper explored the efficient use of variational autoencoders as a tool for controlling the synthesis of weather fields considering more extreme scenarios.

An essential aspect of weather generators is controlling the synthesis for different weather scenarios in light of climate change.

We reported that controlling the sampling from the known latent distribution is effectively related to synthesizing samples with more extreme scenarios in the precipitation dataset experimented in our tests.

As further research, we expect to explore models that enable multiple distributions for finer control of synthesis and to tackle data with multiple weather system distributions.

Thank You!