Towards debiasing climate simulations using unsupervised image-to-image translation networks

James Fulton, Ben Clarke

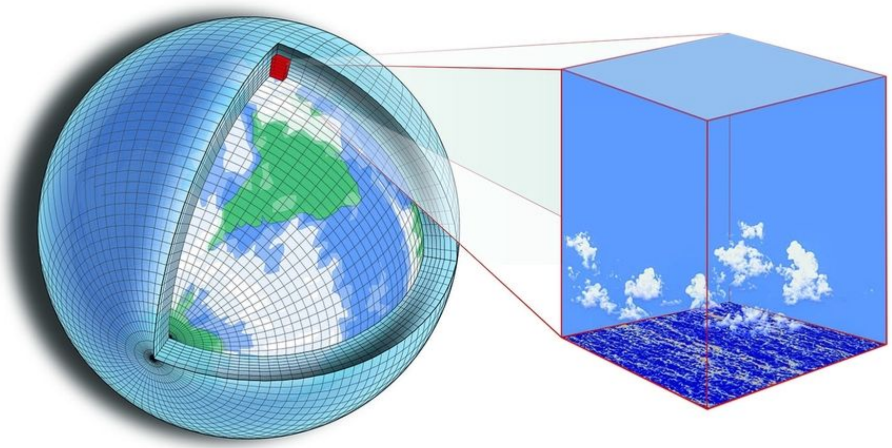
THE UNIVERSITY of EDINBURGH

UNIVERSITY OF OXFORD

NeurIPS 2021 Workshop: Tackling Climate Change with Machine Learning

Motivation

- Climate simulations via GSMs are used widely for research
- They also inform policy, and legal action
- They aren't perfect representations of the world
- We need to make best use of the simulation results via post-processing



https://news.mit.edu/2018/new-climate-modeling-alliance-clima-1212

The status-quo: Quantile mapping

Quantile mapping is defined as:

$$\hat{x}_{obs} = \mathcal{F}_{obs}^{-1}(\mathcal{F}_{GCM}(x_{GCM}; \vec{\theta}); \vec{\theta})$$

where

$$\mathcal{F}_{GCM}(x_{GCM}; \vec{\theta})$$

is the cumulative distribution function calculated from the GCM (simulation) of variable ***x*** at location ***θ***

$$\mathcal{F}_{obs}(x_{obs}; \vec{\theta})$$

is the cumulative distribution function calculated from the observations - and we use its inverse

The status-quo: Quantile mapping

Quantile mapping is defined as:

$$\hat{x}_{obs} = \mathcal{F}_{obs}^{-1}(\mathcal{F}_{GCM}(x_{GCM}; \vec{\theta}); \vec{\theta})$$

Doesn't take into account:

- Spatial correlations
- Cross variable correlations (e.g. temperature vs pressure)

We cannot have corresponding pairs

For this task, we cannot collect prediction pairs like $\{\vec{x}^{\,i}_{GCM}, \vec{x}^{\,i}_{obs}\}$

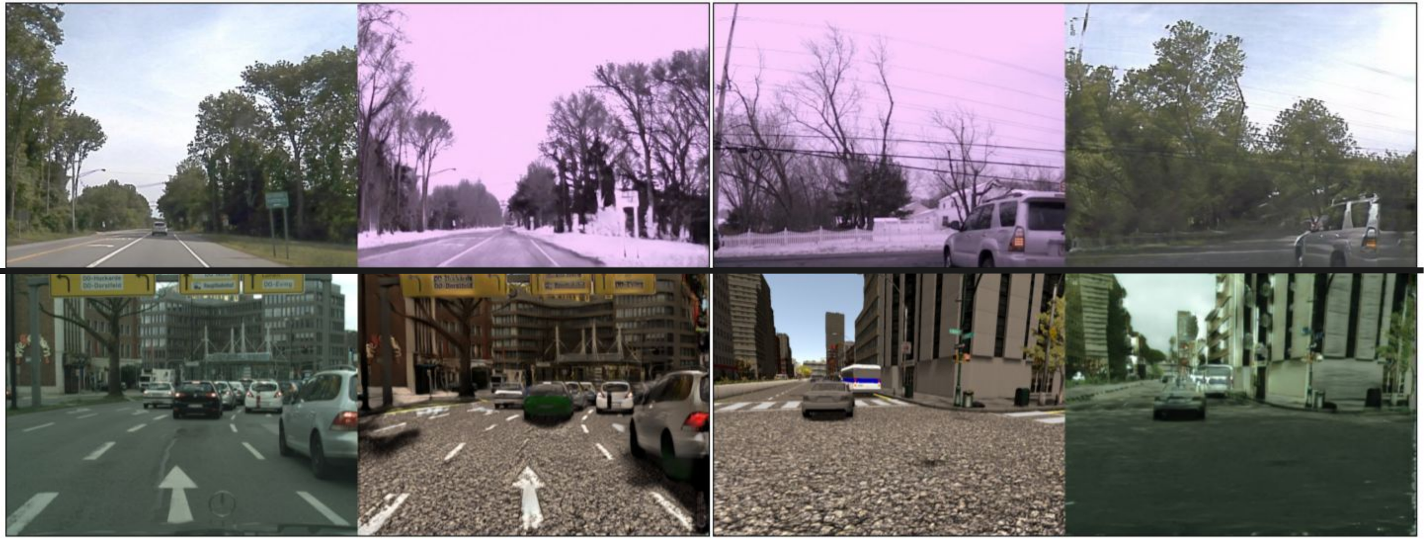
- Even a **perfect simulator** will diverge
- Pairs can only be collected for short time

- A **biased simulator** may never revisit of initial states where data was collected
- Cause misalign between train and test data

Unsupervised image-to-image translation networks

summer ⟷ winter

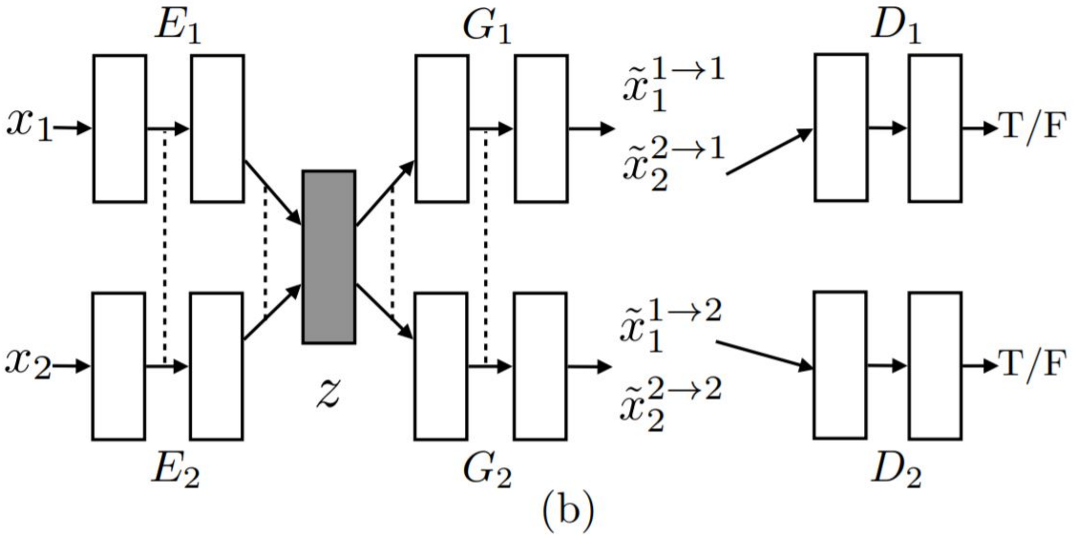
video game ⟷ photo



Liu, Ming-Yu, Thomas Breuel, and Jan Kautz. "Unsupervised image-to-image translation networks." *Advances in neural information processing systems*. 2017.

Zhu, Jun-Yan, et al. "Unpaired image-to-image translation using cycle-consistent adversarial networks." *Proceedings of the IEEE international conference on computer vision*. 2017.

The UNIT network



Components of loss

- L1 loss on image reconstruction
- L1 cycle consistency loss
- GAN loss on translation

Data sources and extent

Climate simulation data source

- HadGEM3
- C20C+ archive
- Historical recreation scenario

“Observations” source

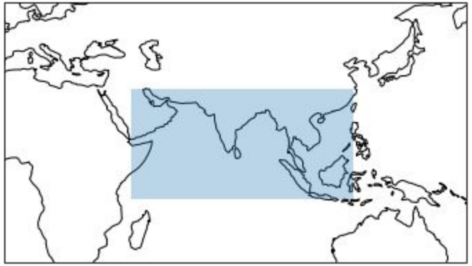
- ERA5

Time overlap

- Jan 1979 - Dec 2013
- Approx. 12 000 days

Spatial extent

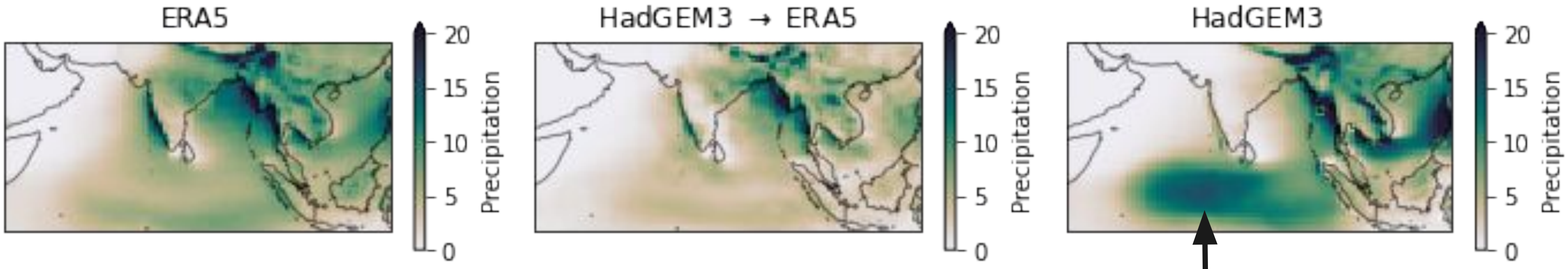
- South Asian monsoon region
- Has large known biases in GCMs
- Hard case to get correct
- 8°S - 30°N
- 44°E - 121°E



Physical variables

- Daily min, mean and max temperature
- Daily total precipitation
- Z500 geopotential height

UNIT improves monsoon circulation bias

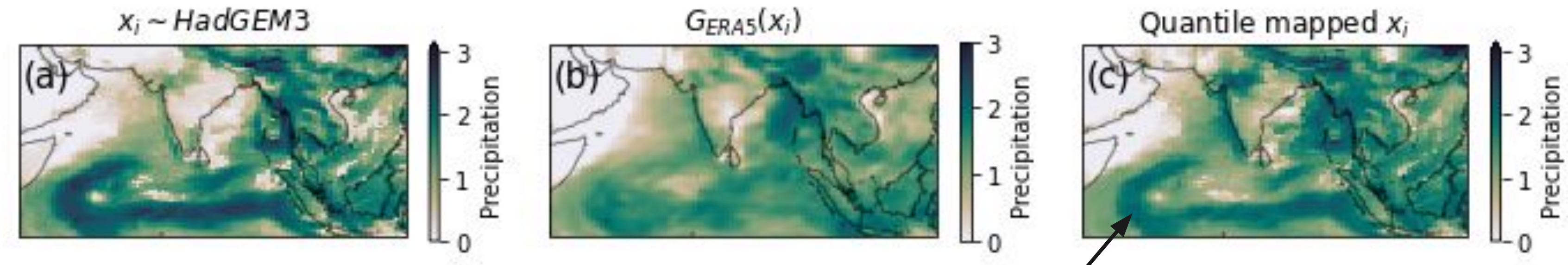


UNIT transform
removes bias

Precipitation bias

Single example

For precipitation only, other 4 variables not plotted



A sample HadGEM3 field with a biased precipitation weather system

Its UNIT translation

Its quantile mapped translation

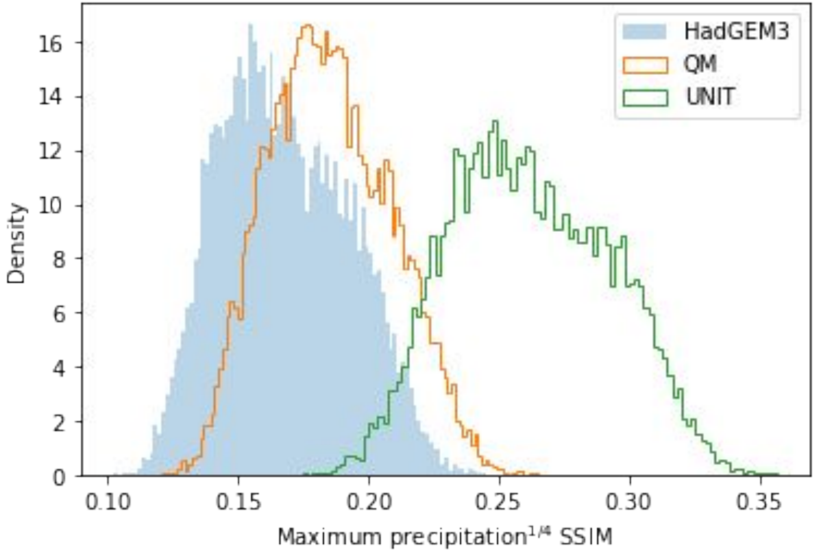
Unrealistic weather system

Spatial matching

To assess spatial plausibility of fields

1. Take each precipitation field from HadGEM3/QM/UNIT
2. Measure structural similarity index measure to all ERA5 precipitation fields
3. Keep score of best match

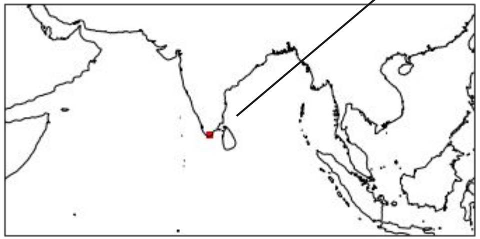
Distribution of best matches

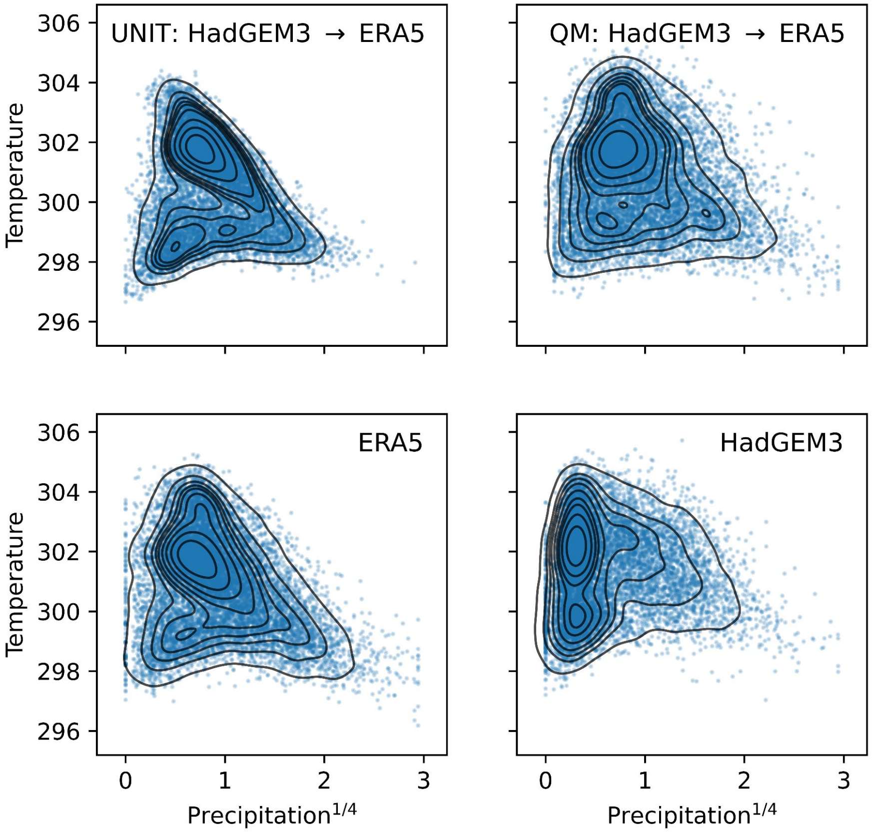


The UNIT translated HadGEM3 data had spatial structures with better matches to the ERA5 data,.

Cross-variable biases

- Joint distribution sampled from single spatial location
- UNIT captures, but over-exaggerates joint distribution structure





Conclusion

UNIT network

- Removes major bias in HadGEM3 precipitation
- Produces precipitation spatial patterns which better match the observations, when compared to baseline
- Corrects the joint distribution of variables at individual spatial locations

Further developments in techniques that would be beneficial

- Translation with distributions with long tails
- Avoid collapse towards modes in joint-distributions - i.e. more diversity

This work was supported by Microsoft AI for Earth grant

james.fulton@ed.ac.uk