ClimateBench: A benchmark for data-driven climate projections

Duncan Watson-Parris, University of Oxford
Y. Rao, D. Olivié, P. Nowack, G Camps-Valls, P. Stier, S. Bouabid, M. Dewey, E. Fons, J. Gonzalez, P. Harder, K. Jeggle, J. Lenhardt, P. Manshausen, M. Novitasari, L. Ricard, C. Roesch

ECMWF ML Workshop, Virtual

29th March 2022

PHILOSOPHICAL TRANSACTIONS A

royalsocietypublishing.org/journal/rsta

Opinion piece

Check for updates

Machine learning for weather and climate are worlds apart

D. Watson-Parris

Atmospheric, Oceanic and Planetary Physics, Department of Physics, University of Oxford, UK

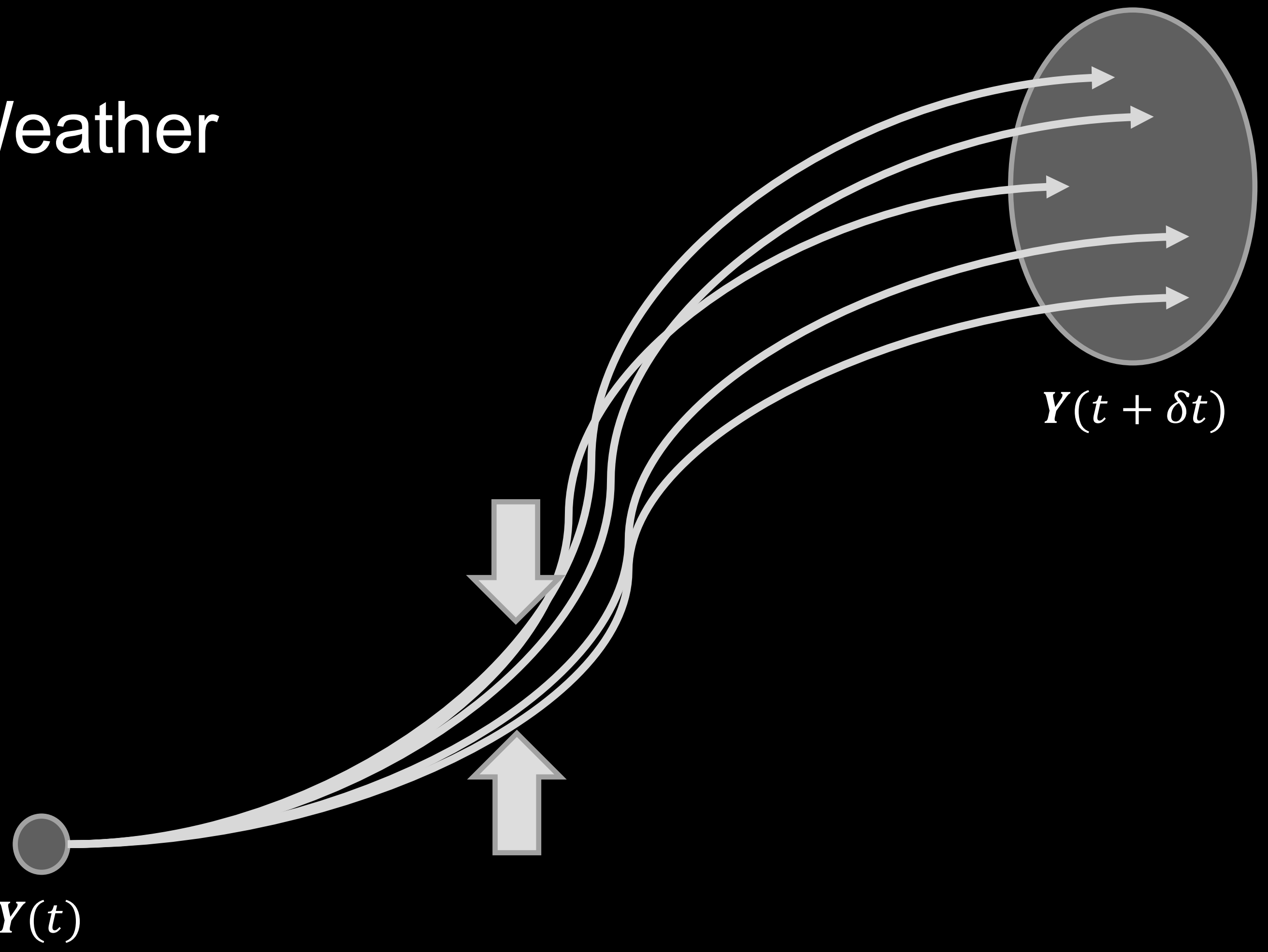
DW-P, 0000-0002-5312-4950

Modern weather and climate models share a common heritage and often even components; however, they are used in different ways to answer fundamentally different questions. As such, attempts to emulate them using machine learning should reflect this. While the use of machine learning to emulate weather forecast models is a relatively new endeavour, there is a rich history of climate model emulation. This is primarily because while weather modelling is an initial condition problem, which intimately depends on the current state of the atmosphere, climate modelling is predominantly a boundary condition problem.

Watson-Parris 2021

ML for Weather and Climate are worlds apart

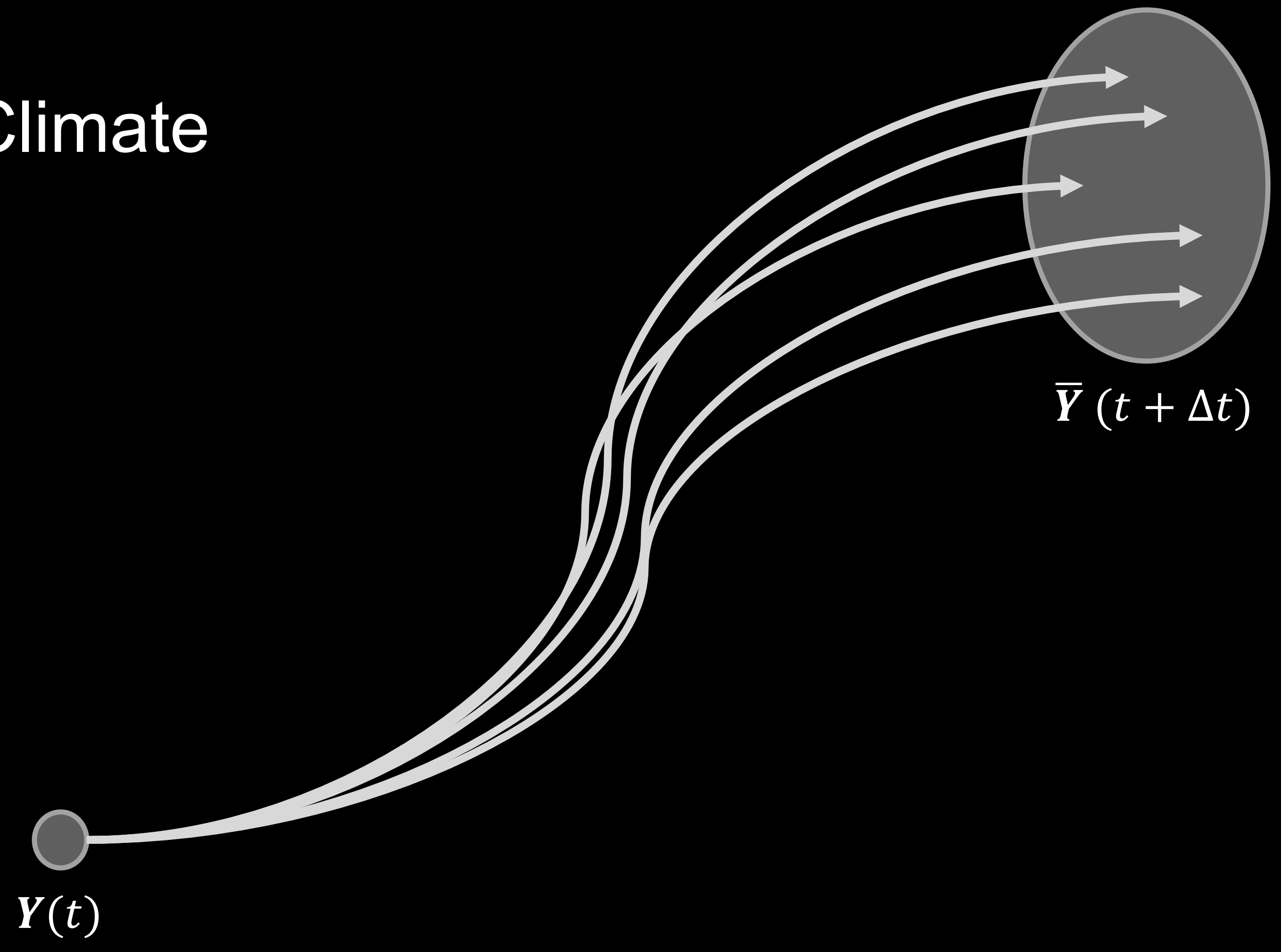
Weather



Watson-Parris 2021

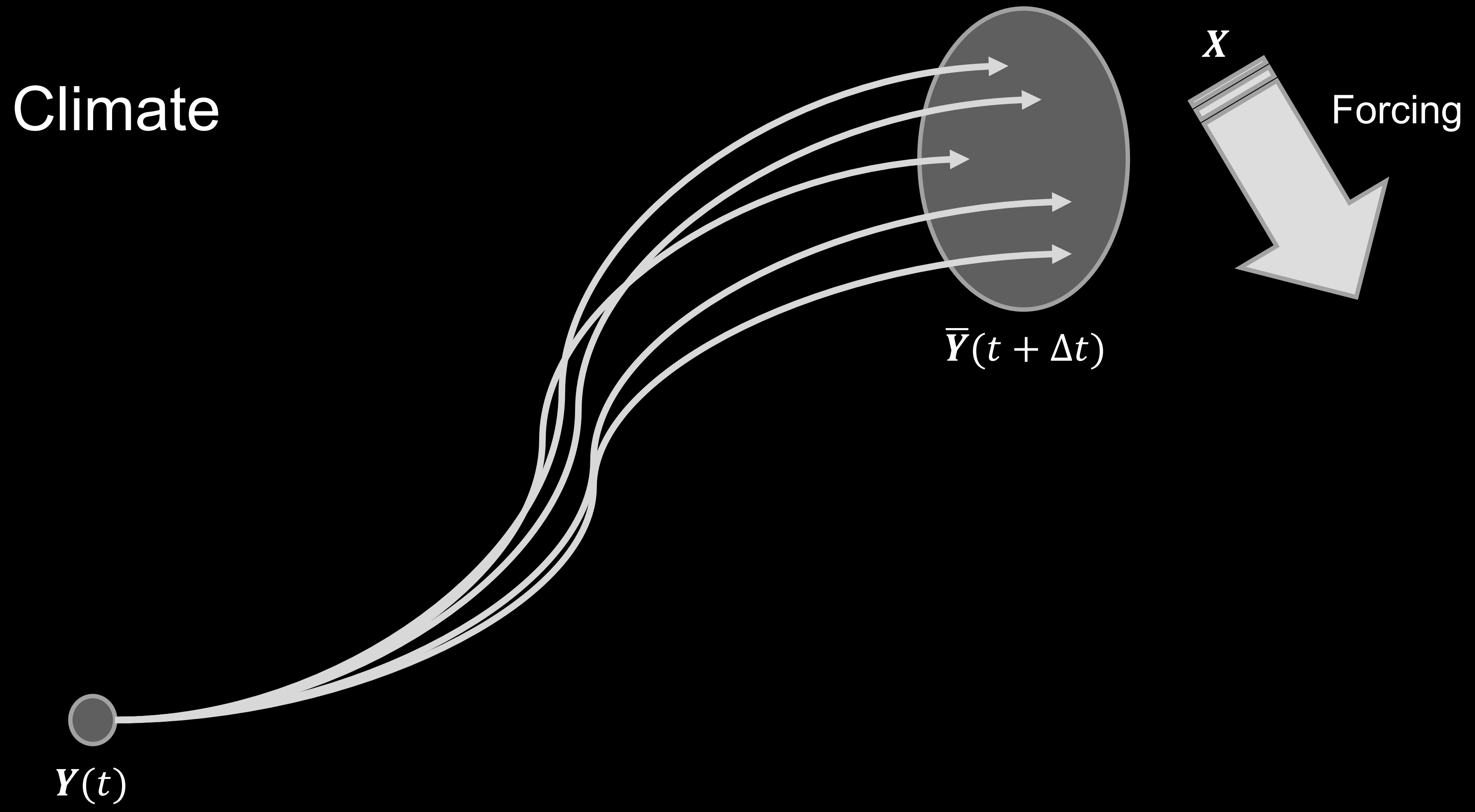
ML for Weather and Climate are worlds apart

Climate



Watson-Parris 2021

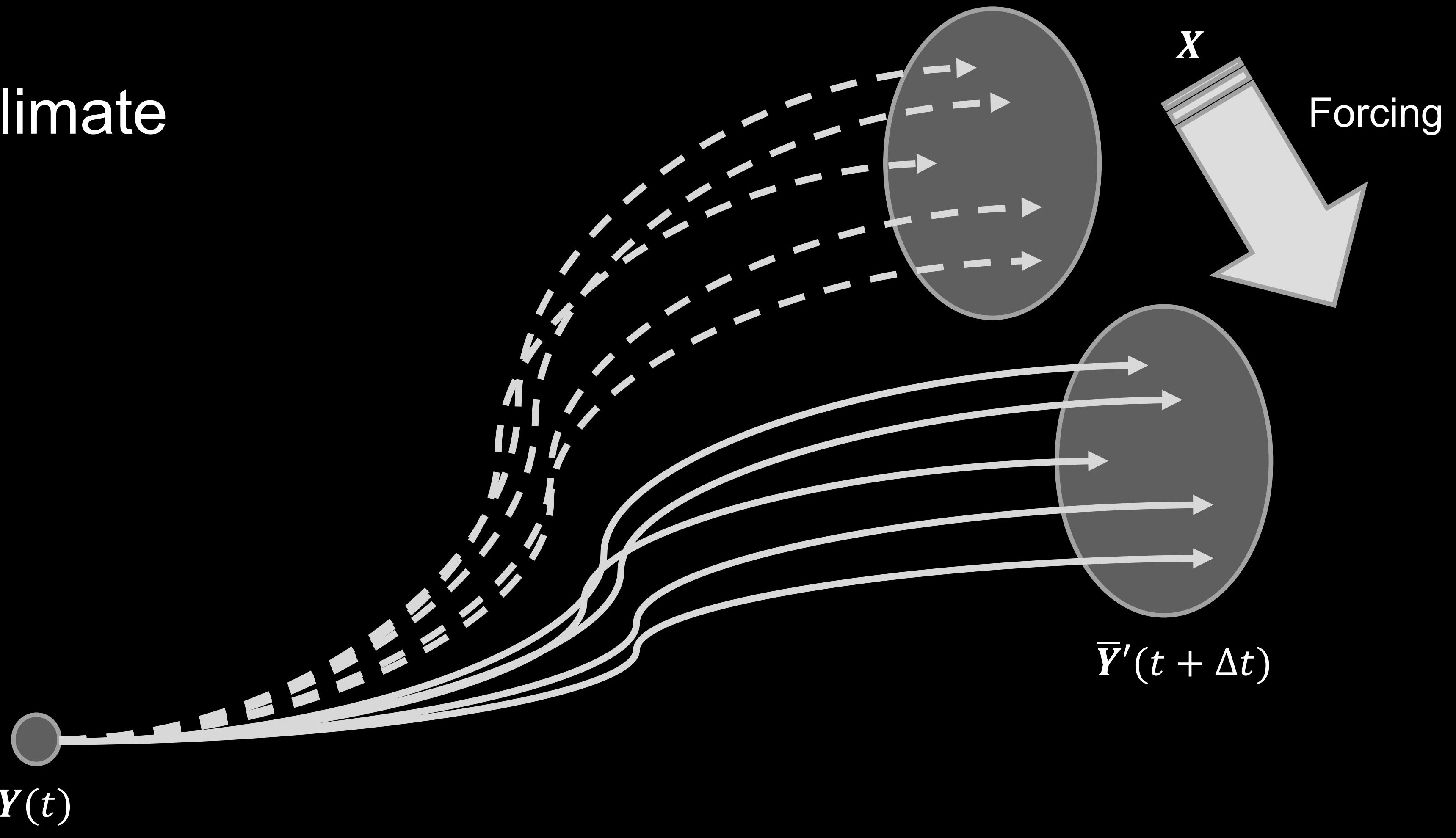
ML for Weather and Climate are worlds apart



Watson-Parris 2021

ML for Weather and Climate are worlds apart

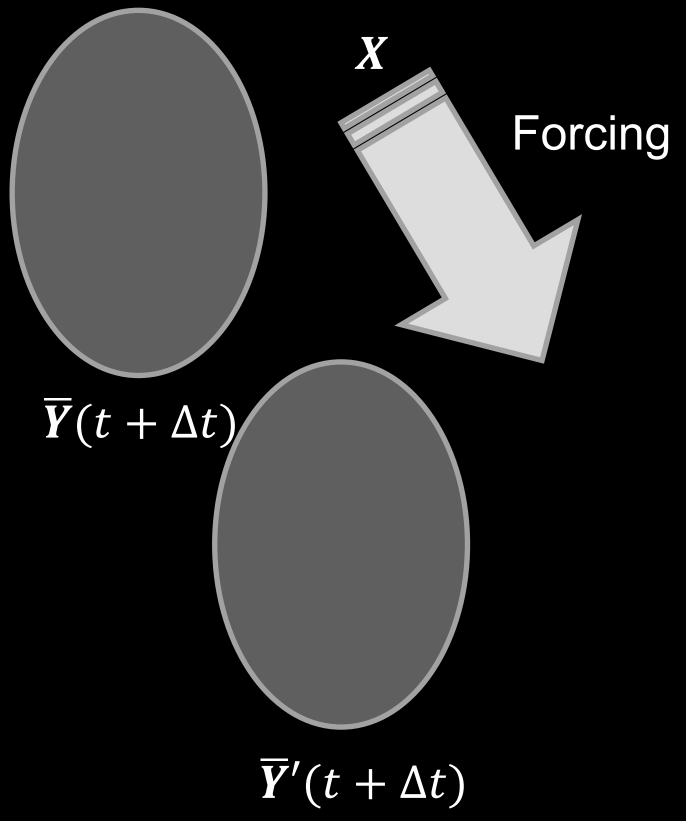
Climate



Watson-Parris 2021

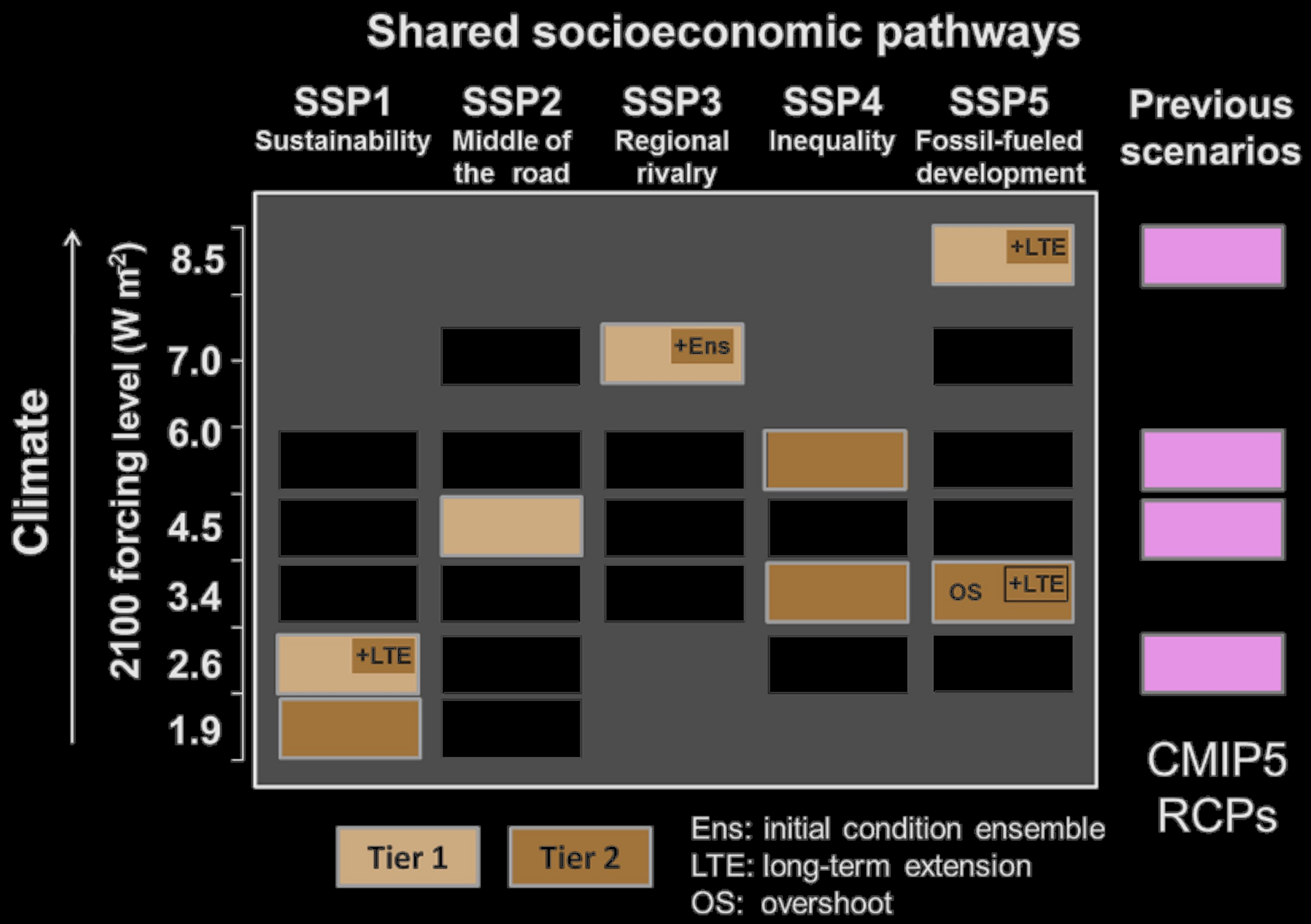
ML for Weather and Climate are worlds apart

Climate

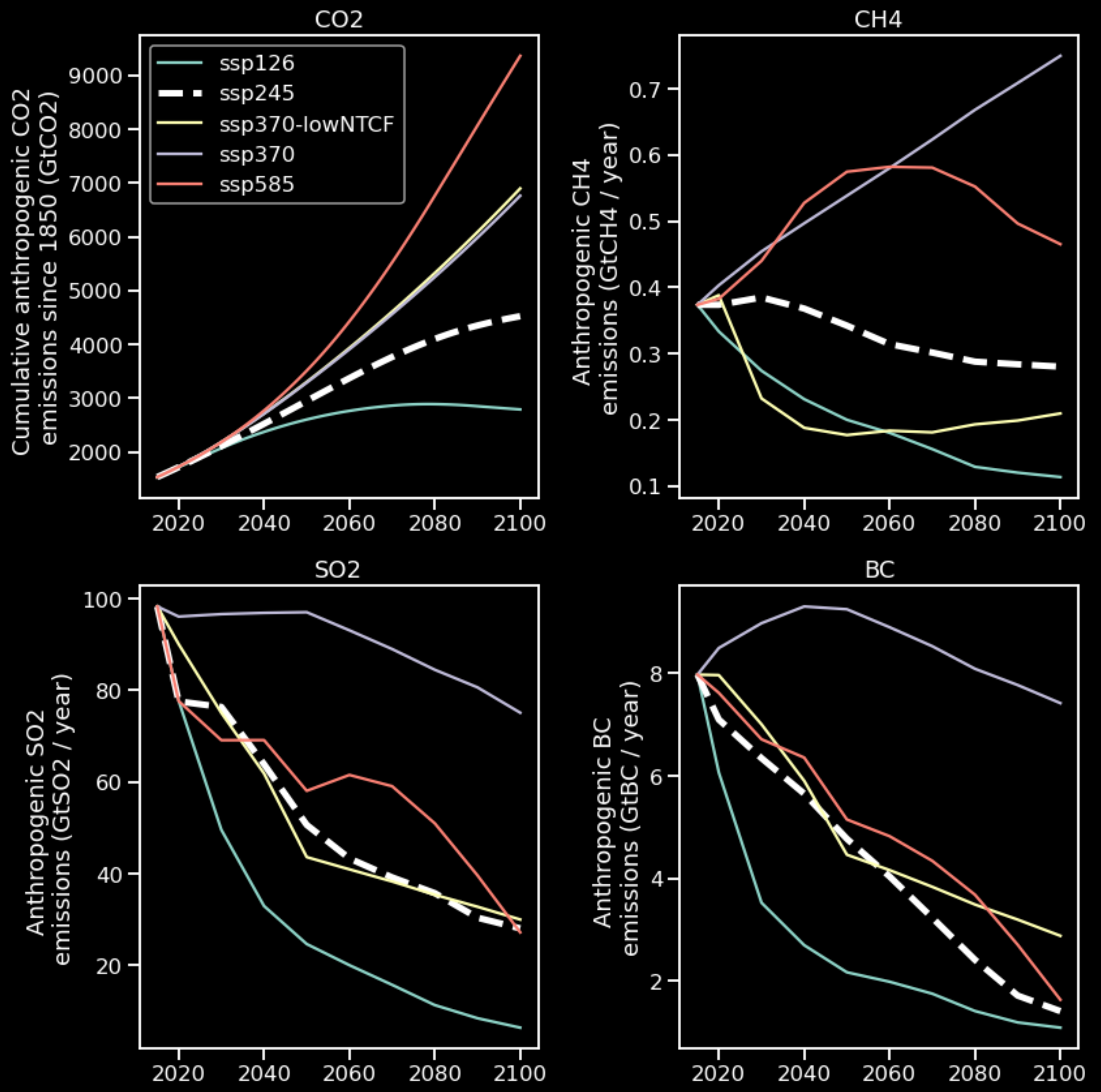


Watson-Parris 2021

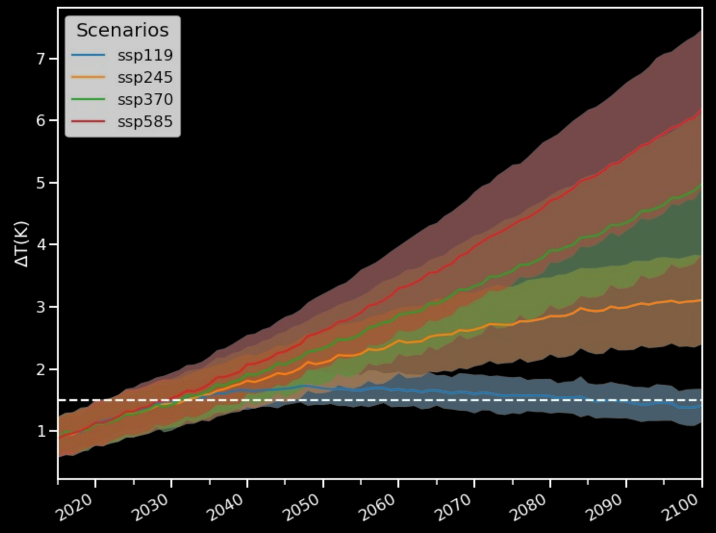
Climate projections: ScenarioMIP



Exploring scenario uncertainty: (Simple) Emulators

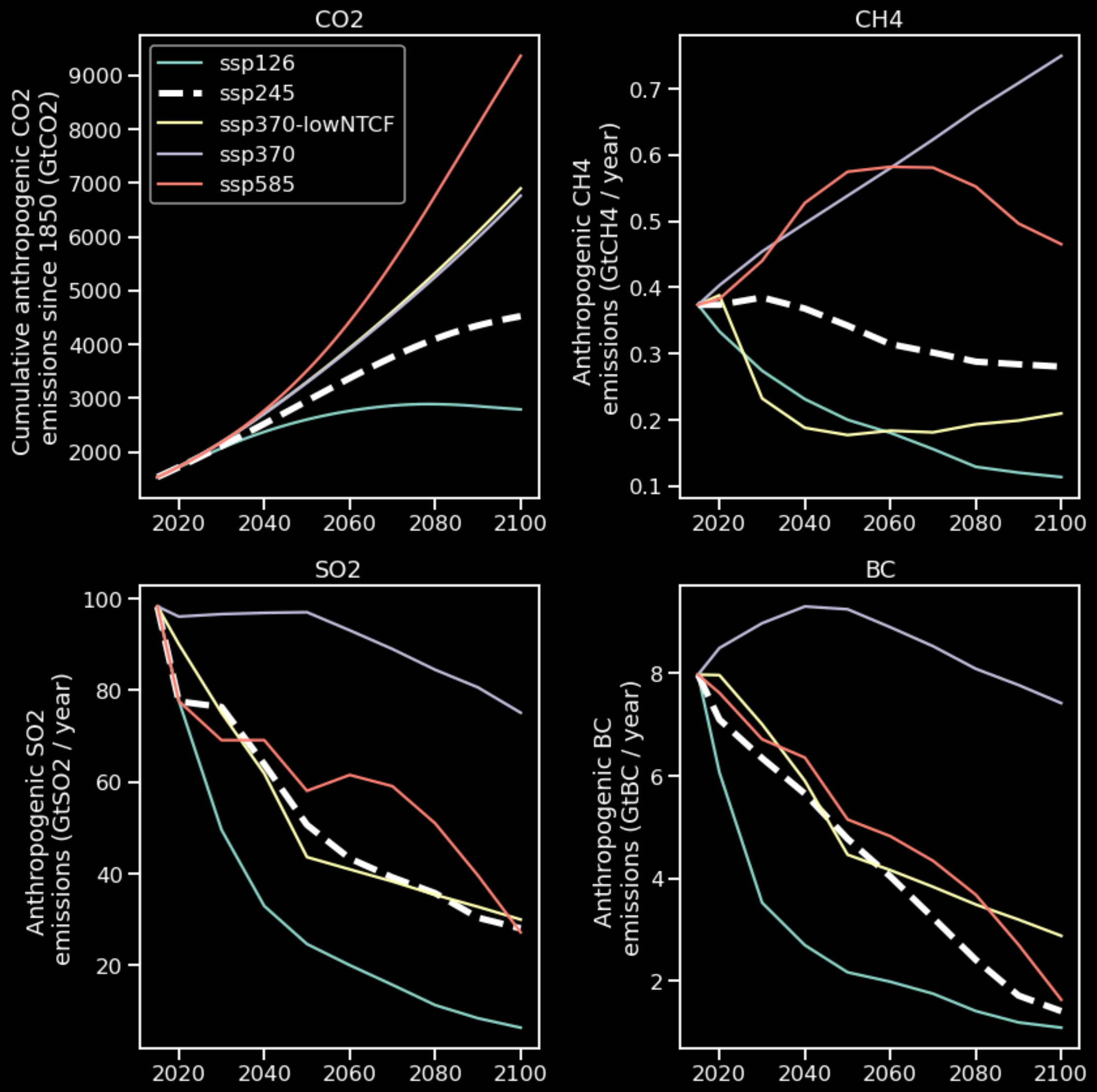


Emulate

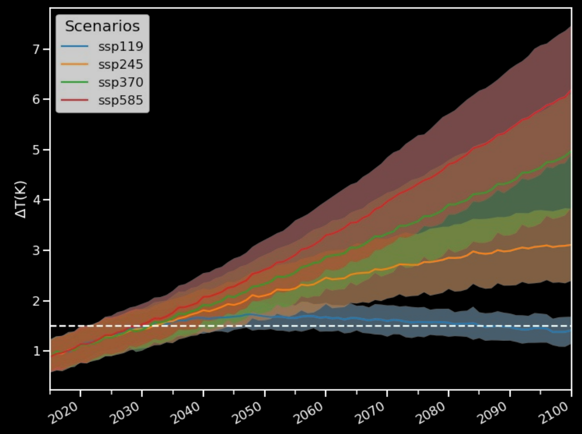


E.g., FAIR, MAGGIC, WASP etc.

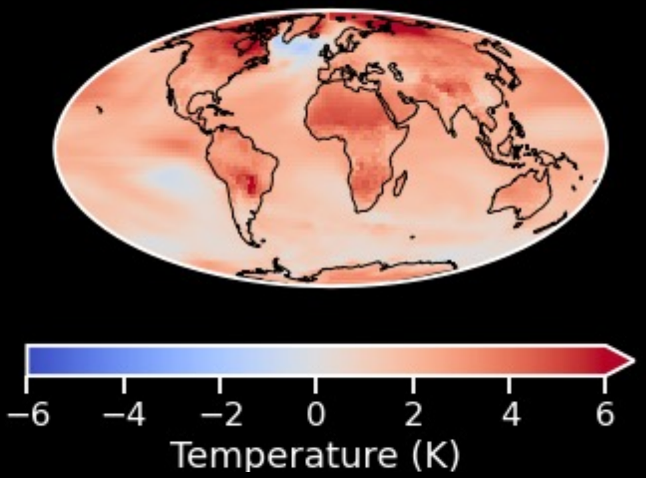
Exploring scenario uncertainty: Pattern Scaling



Emulate



E.g., MESMER



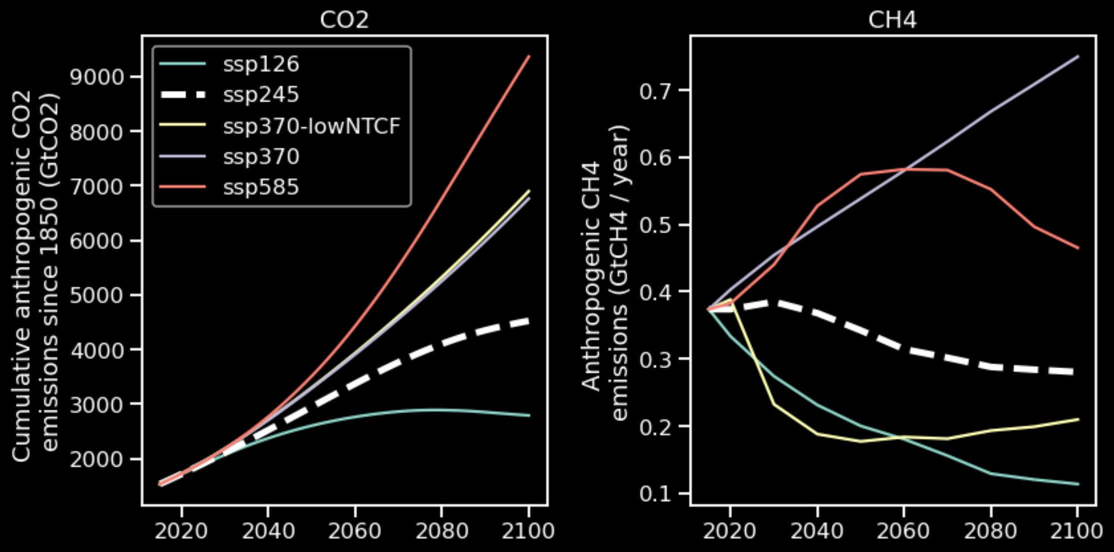
Climate
WeatherBench

Leaderboard

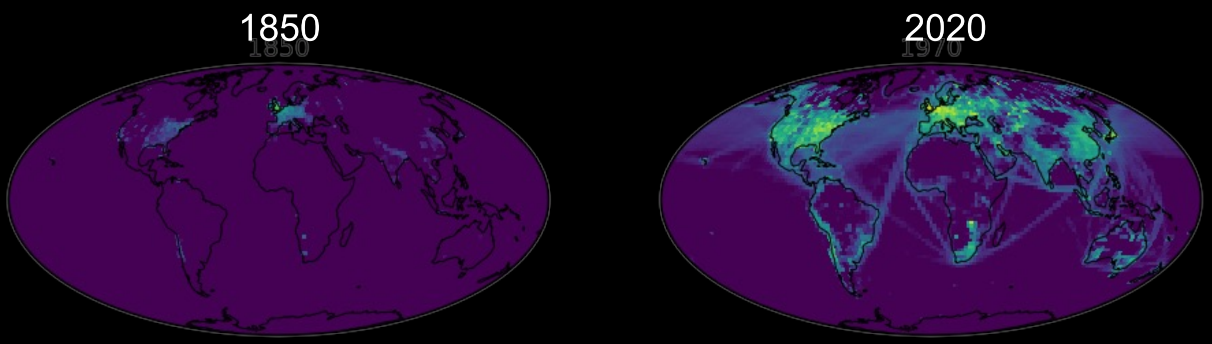
Model	Z500 RMSE (3 / 5 days) [m^2/s^2]	T850 RMSE (3 / 5 days) [K]	Notes	Reference
Operational IFS	154 / 334	1.36 / 2.03	ECWMF physical model (10 km)	Rasp et al. 2020
Rasp and Thuerey 2020 (direct/continuous)	**268 / 499**	**1.65 / 2.41**	Resnet with CMIP pretraining (5.625 deg)	Rasp and Thuerey 2020
IFS T63	268 / 463	1.85 / 2.52	Lower resolution physical model (approx. 1.9 deg)	Rasp et al. 2020
Weyn et al. 2020 (iterative)	**373 / 611**	**1.98 / 2.87**	UNet with cube-sphere mapping (2 deg)	Weyn et al. 2020
IFS T42	489 / 743	3.09 / 3.83	Lower resolution physical model (approx. 2.8 deg)	Rasp et al. 2020
Weekly climatology	816	3.50	Climatology for each calendar week	Rasp et al. 2020

Watson-Parris et al. *ESSOAr 2022*

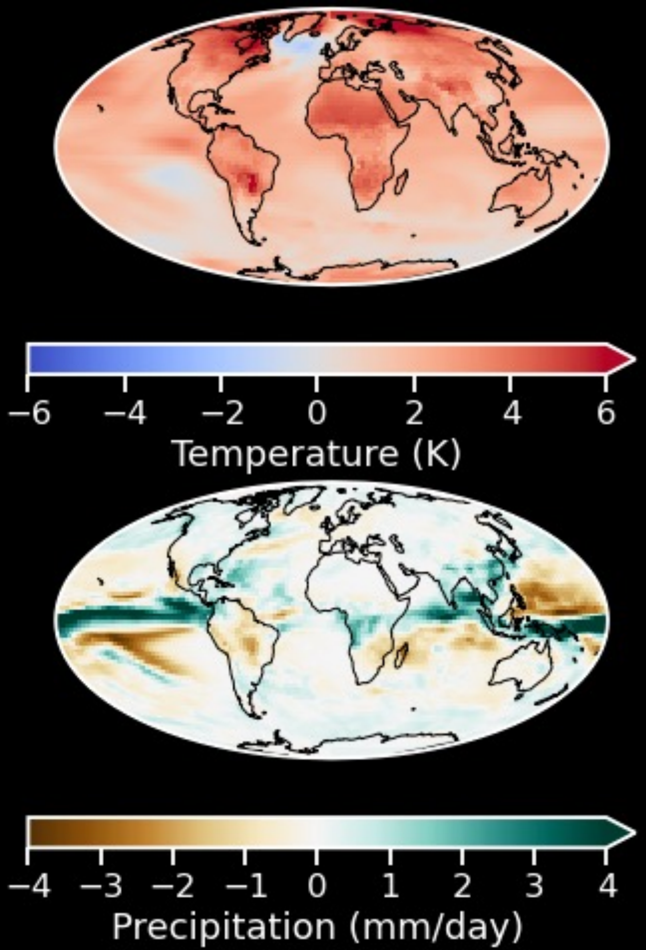
Exploring scenario uncertainty: ClimateBench



Spatially resolved emissions of SO2 and BC

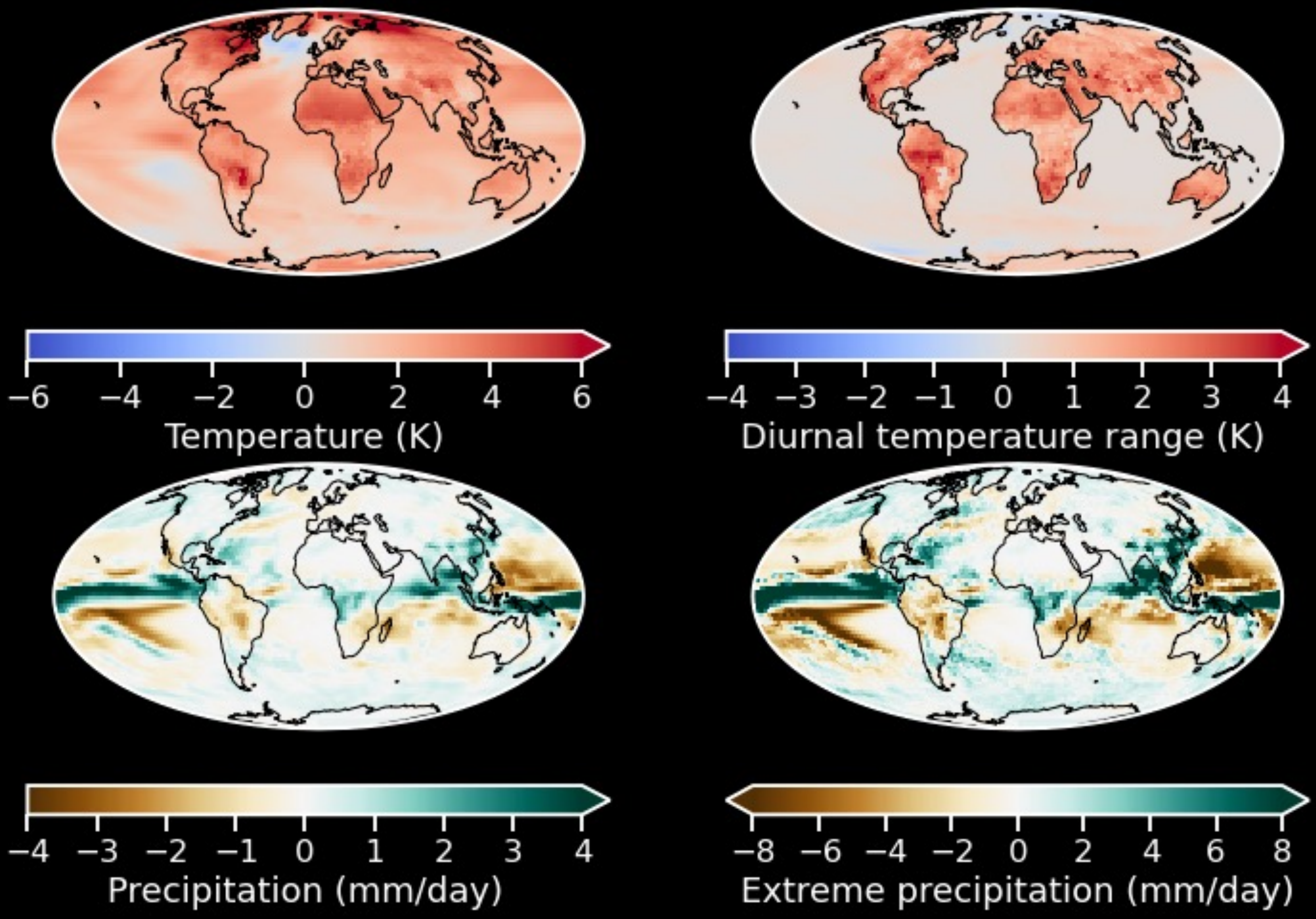


Emulate



Evaluation

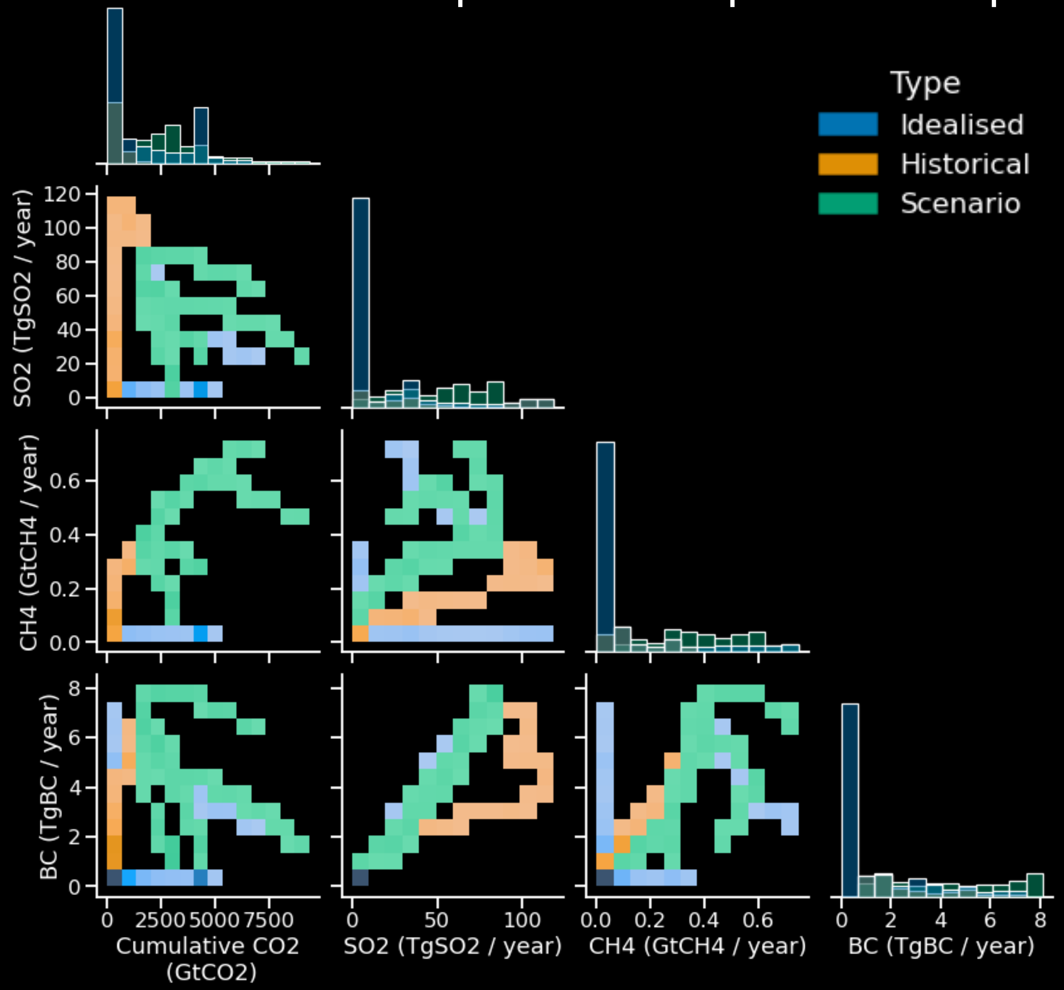
Given SSP245, estimate:



Ranked by RMSE for 2050 - 2100

ClimateBench training dataset

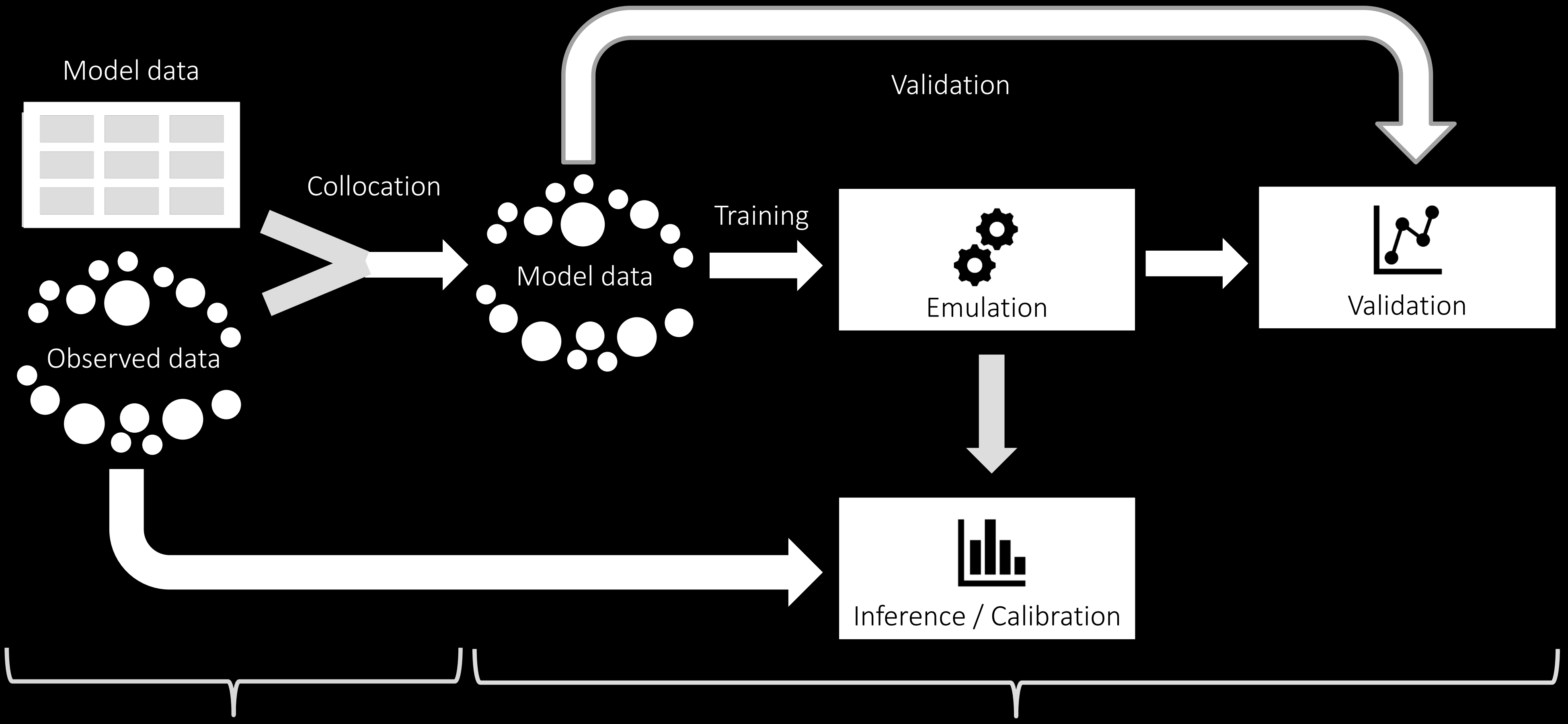
Pre-processed input4MIPS inputs and NorESM2 outputs for multiple experiments



- Selection of future 'Socio-economic pathways' designed to span range of likely future scenarios
- Two idealised simulations modelling the effect of abrupt and gradual increases in CO2
- Core historical experiment plus two detection and attribution simulations of the historical period with either aerosol or GHG held constant
- Together they do a good job of spanning the input space – although aerosol species are correlated

Seland et al. 2020

ESEm: Earth System Emulator



https://github.com/duncanwp/ESEm

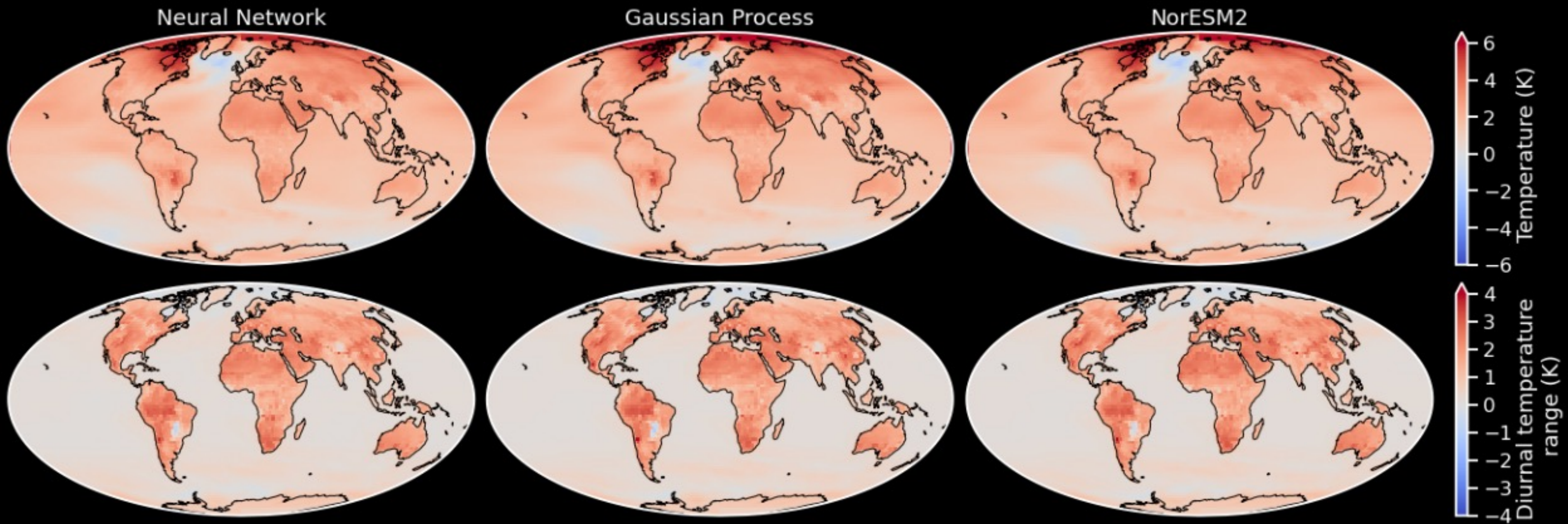
Watson-Parris et al. *GMD 2021*

Hackathon Baseline Models

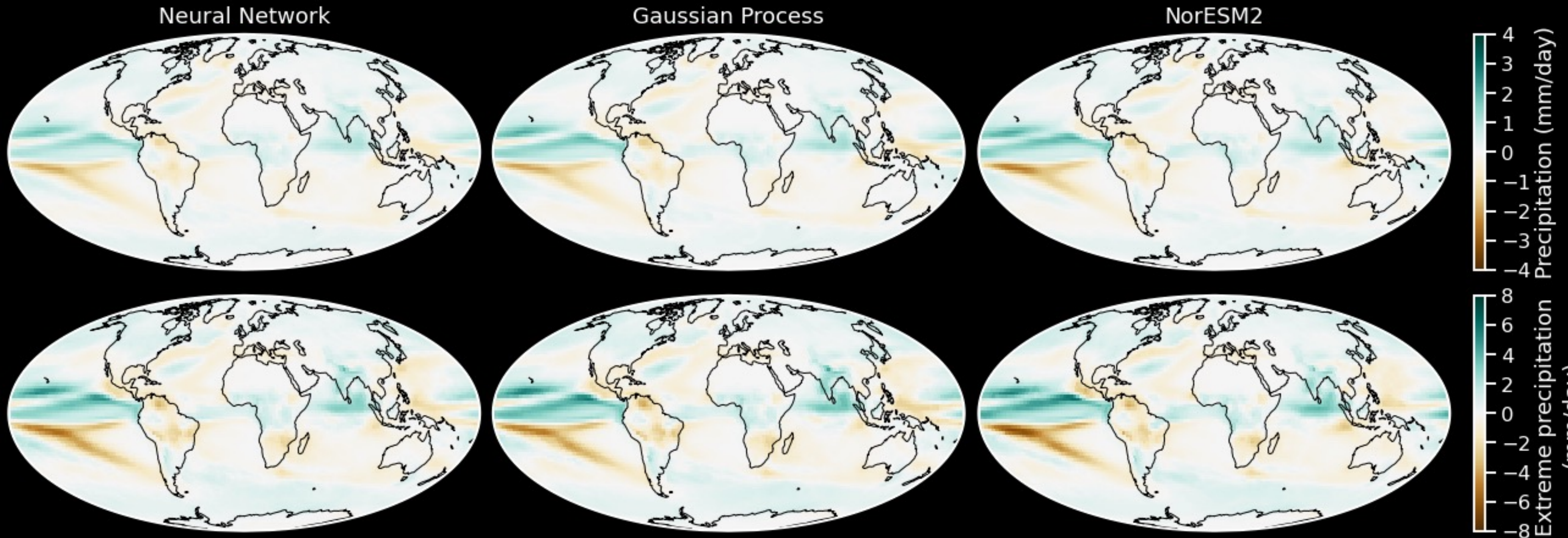
- 8 teams participated in the 3rd NOAA AI workshop hackathon
- Competed to score the best RMSE averaged across all variables
- Three best models so far: Random Forest (RF) and Gaussian process (GP) trained using each year independently. LSTM accounts for temporal evolution

Variable	GP	RF	LSTM	(UKESM)
Mean temperature	0.39 K	0.49 K	0.39 K	1.8 K
Temperature range	0.17 K	0.16 K	0.17 K	1.2 K
Mean precipitation	0.53 mm/day	0.57 mm/day	0.55 mm /day	0.8 mm/day
90th percentile precipitation	1.5 mm/day	1.7 mm / day	1.5 mm /day	2.4 mm/day

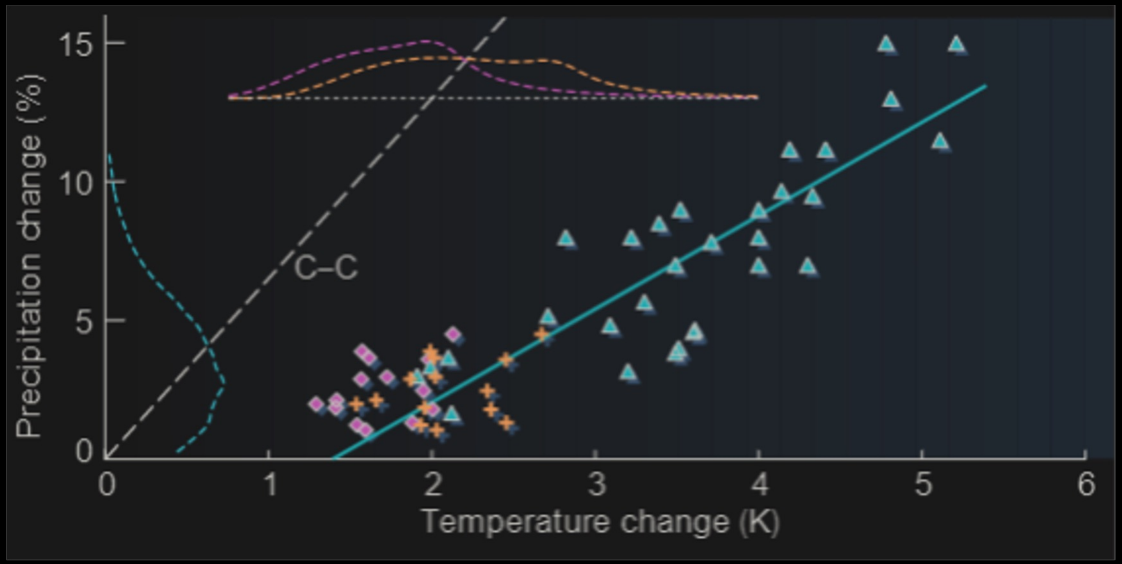
Baseline Models



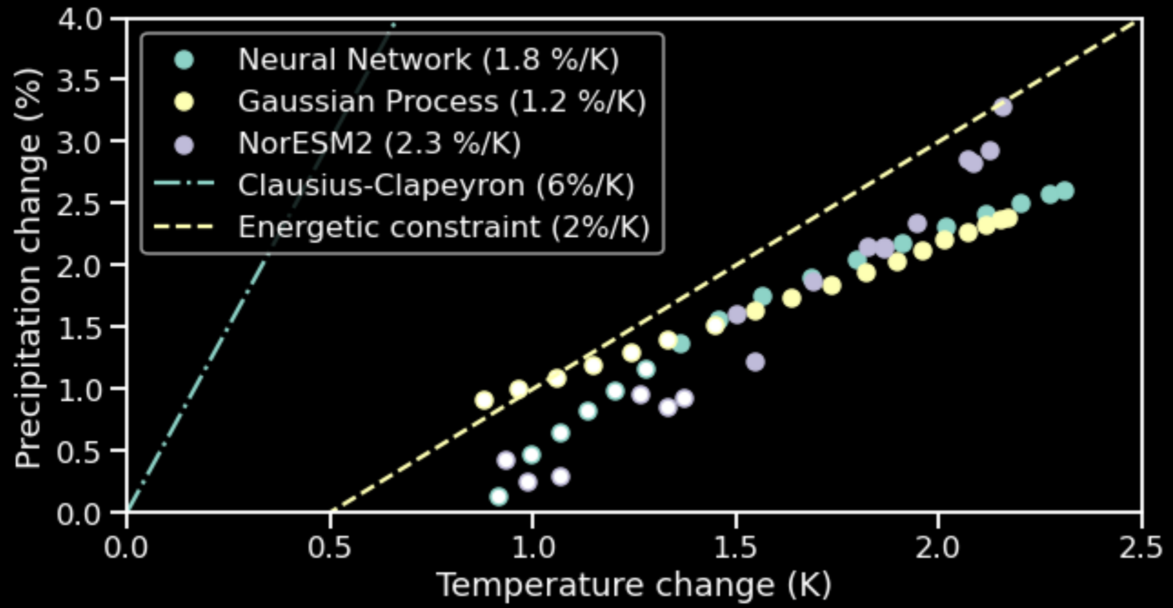
Baseline Models



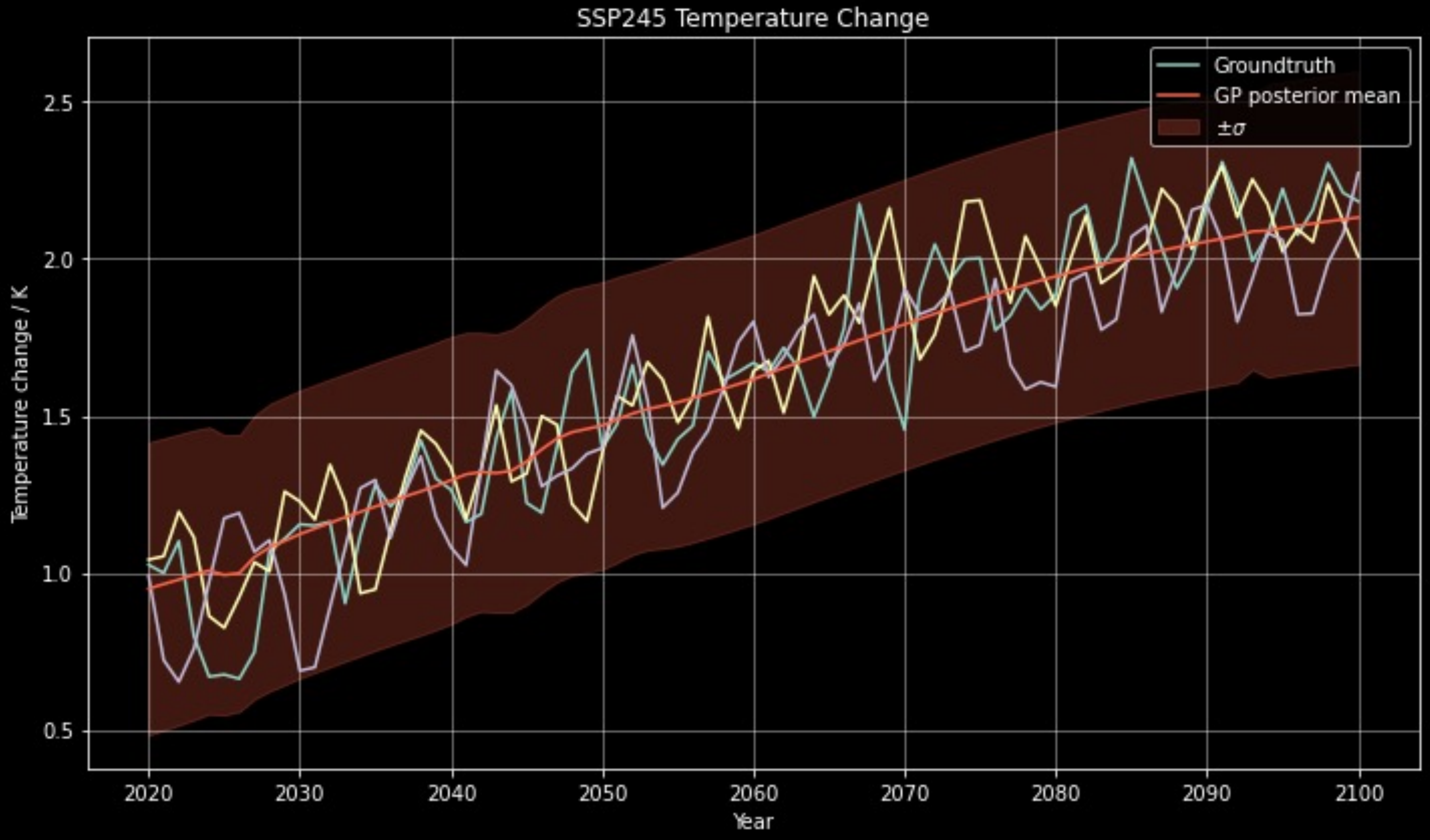
ClimateBench: Energy conservation



Allen & Ingram 2002



GP Model: Sampling Internal Variability



Summary

- An open-source dataset is presented to allow the training and benchmarking of spatially resolved Earth system emulators
- Machine learning techniques are used to create such emulators with excellent accuracy across multiple climate variables
- Each approach has its benefits: GPs allow sampling of variability; CNNs can capture spatial covariances and RFs provide sensitivities
- We have physical constraints on the (global) response, and these could be encoded using recent physically-informed ML approaches
- Build your own model: https://github.com/duncanwp/climatebench !

Spare slides