Beyond In-Distribution Skill: Towards Robust ML Parameterisations for Non-Stationary Climate Systems

Bradley Stanley-Clamp[1,2], Hannah Christensen[3], Ingmar Posner[1]

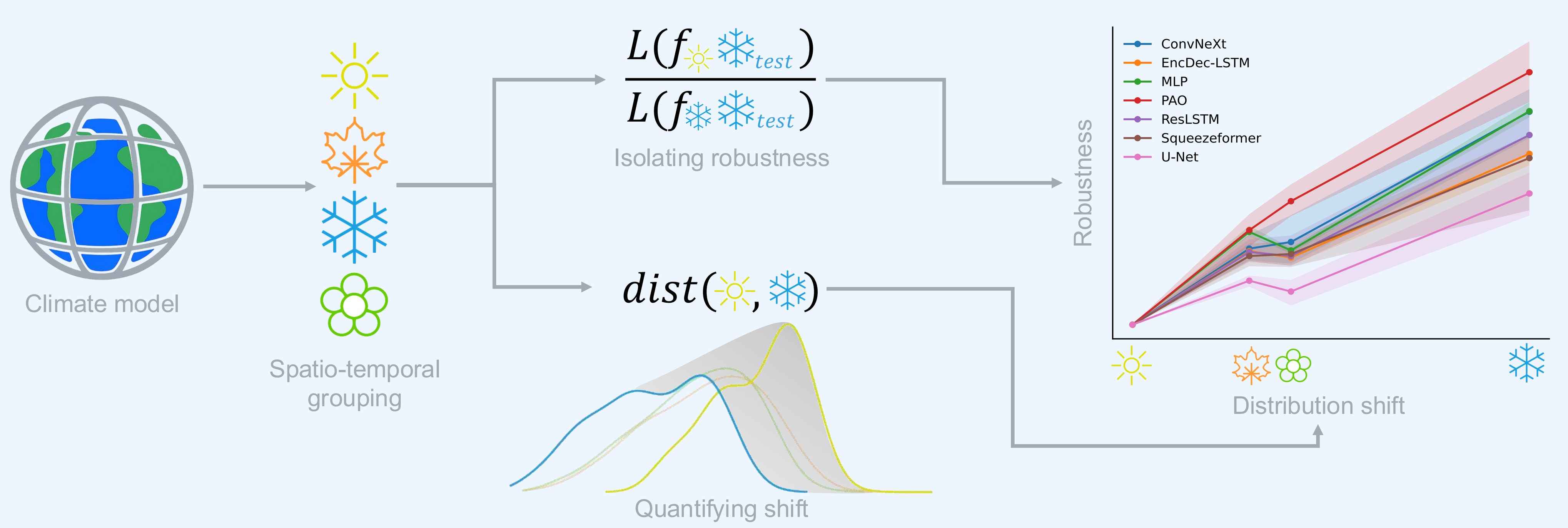
1 Motivation

Current state of the art parameterisations achieve:

- ✓ Skilful prediction
- ✓ Stable coupling
- ? Robust to shift

- Limited physically realistic OOD scenarios
- How does magnitude of shift affect performance?
- Does a single test set performance reflect robustness?

In ML, distribution shift is an edge case; in climate modelling, it's the problem itself.



2 Contributions

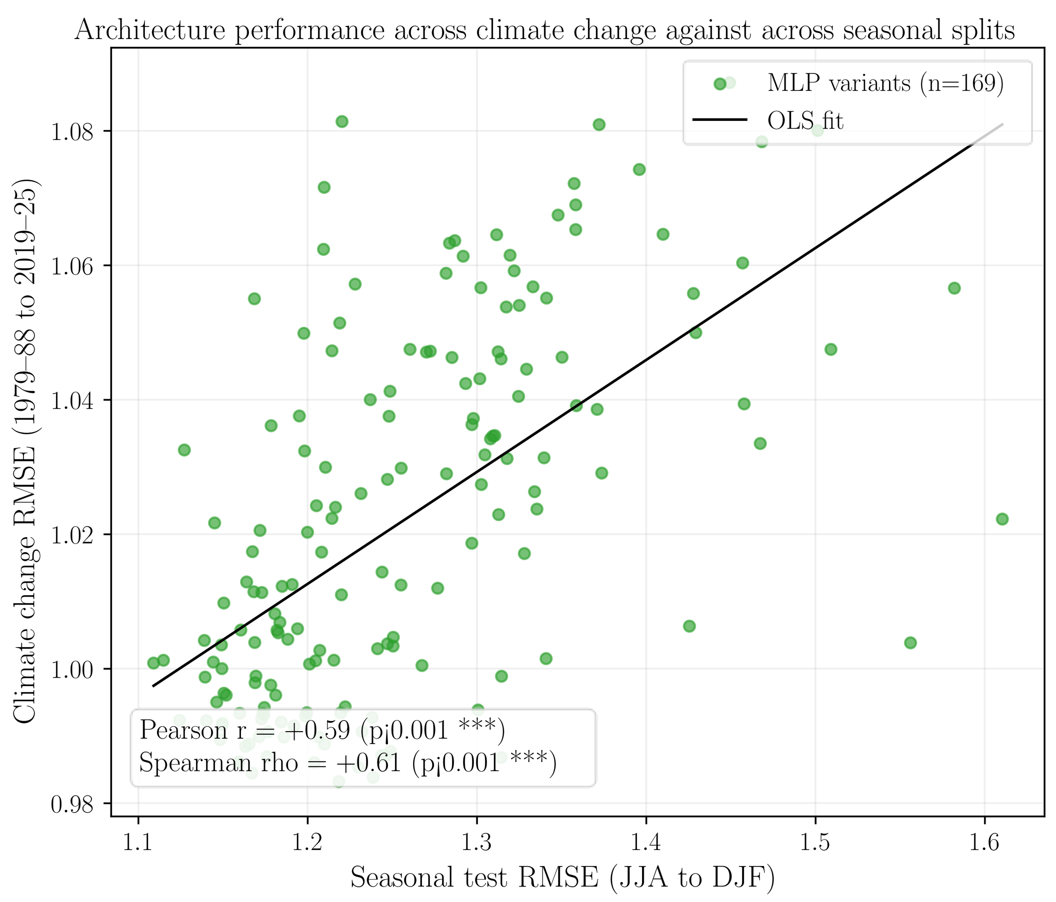
a) Evaluation framework inspired by emergent constraints with zero data collection overhead

b) Characterised robustness amid varying predictability across test sets

c) Composing simple experts can improve robustness to radiation

3 Seasonality as a proxy for climate change

- **Emergent constraints:** A model that cannot reproduce key observed climate behaviours is unlikely to reliably predict future responses
- **Hypothesis:** An architecture that generalises across seasons will generalise to future climates



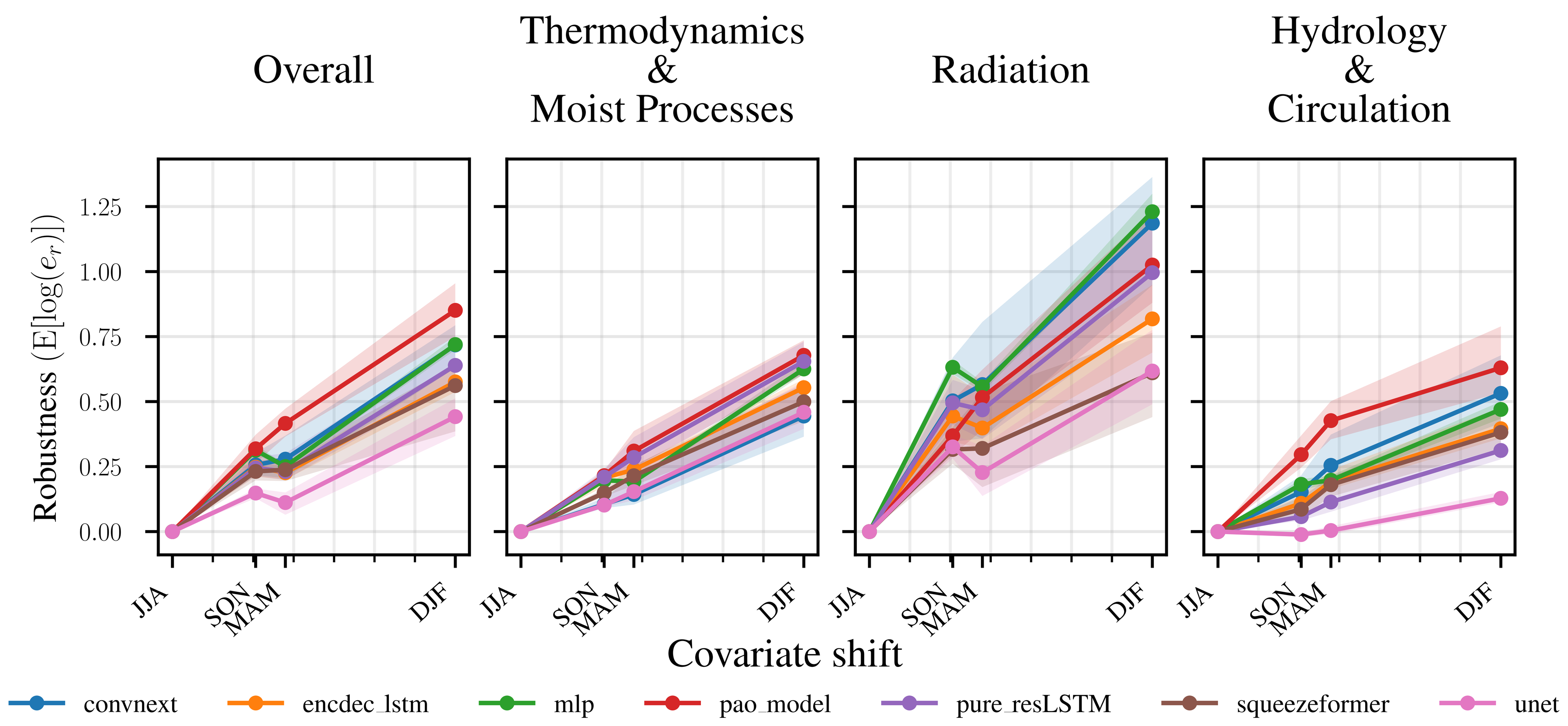
ERA5 Monthly:
Inputs:
- Near surface air temperature,
- 500hPa geopotential height,
- Precipitable water

Output:
- $\log(precip)$

- Only true if climate change signal is strong
- Seasonal shift is larger than climate change, and dependent on region

4 ML parameterisations are not robust

1. Train on on one spatio-temporal group and test on others
2. Isolate robustness from predictability challenges
3. Quantify covariate shift with energy distance



- Degradation is related to covariate shift
- Degradation occurs across groups but varies based on dynamics
- Robustness seems to be architecture agnostic

We need a new approach

5 Composing for generalisation

Predicting surface radiation fields

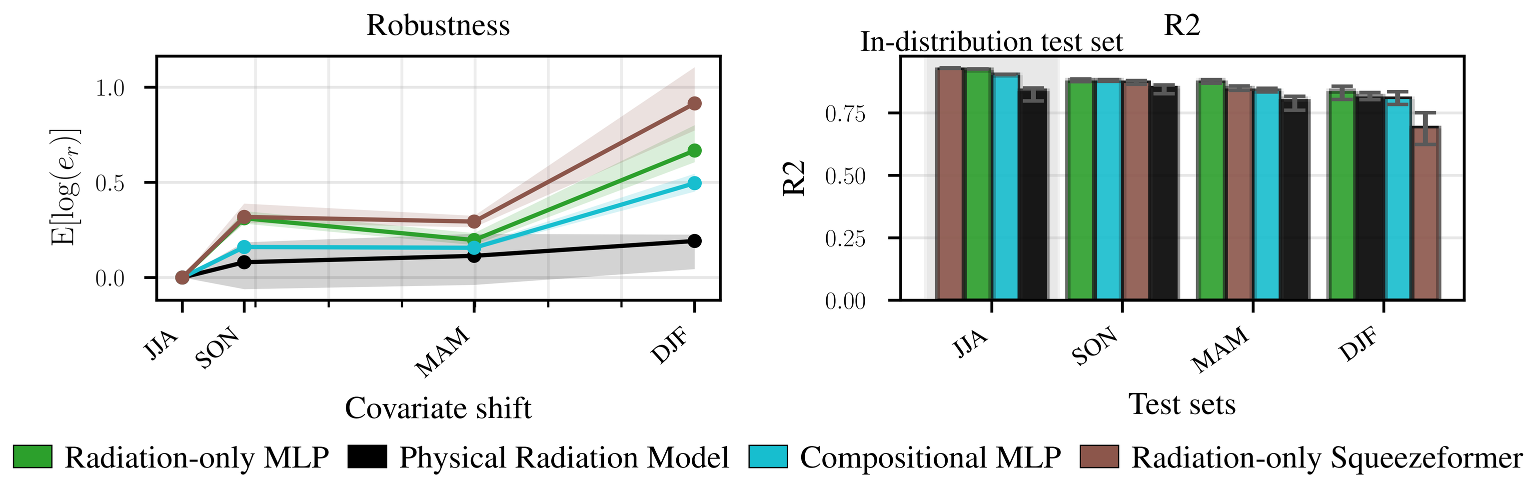
Compositional models: predictions are constructed by composing behaviours that are season agnostic

$$R_{net} = (1 - w_{cloud}) \cdot R_{clear} + w_{cloud} \cdot R_{cloud}$$

Physical experts

Scenario weighting

Experts for clear/cloudy transmittance, effective albedo and emissivity



- Physical priors beat ML under shift: a 38-parameter model outperforms large neural networks out-of-distribution
- Skill and robustness trade off: higher in-distribution accuracy degrades faster under seasonal shift
- Compositional structure helps: regime-specific experts improve OOD robustness without sacrificing ML-level skill

How do can we learn season agnostic behaviours?

[1]Applied AI Lab, University of Oxford, Oxford, United Kingdom.
[2]Intelligent Earth UKRI AI Centre for Doctoral Training in AI for the Environment, University of Oxford, Oxford, United Kingdom.
[3]Atmospheric, Oceanic and Planetary Physics, University of Oxford, Oxford, United Kingdom.

Paper

Sharing is encouraged

Abstract

bradleysc@robots.ox.ac.uk