CASANZ 2026 · Advanced Monitoring, Remote Sensing and AI Systems

Agentic optimisation of WRF

An agent loop for autonomous, reasoning-driven tuning of numerical weather prediction models

Danny Craggs[1] · Peter D'Abreton[2] · Andrew Twidle[2]

[1] Zephyr Environmental · [2] Ideagen Envirosuite

Notes

Affiliation

My co-authors and I carried out this work at Ideagen Envirosuite in early 2026.

AI assistance

The work presented today was prepared with the assistance of AI.

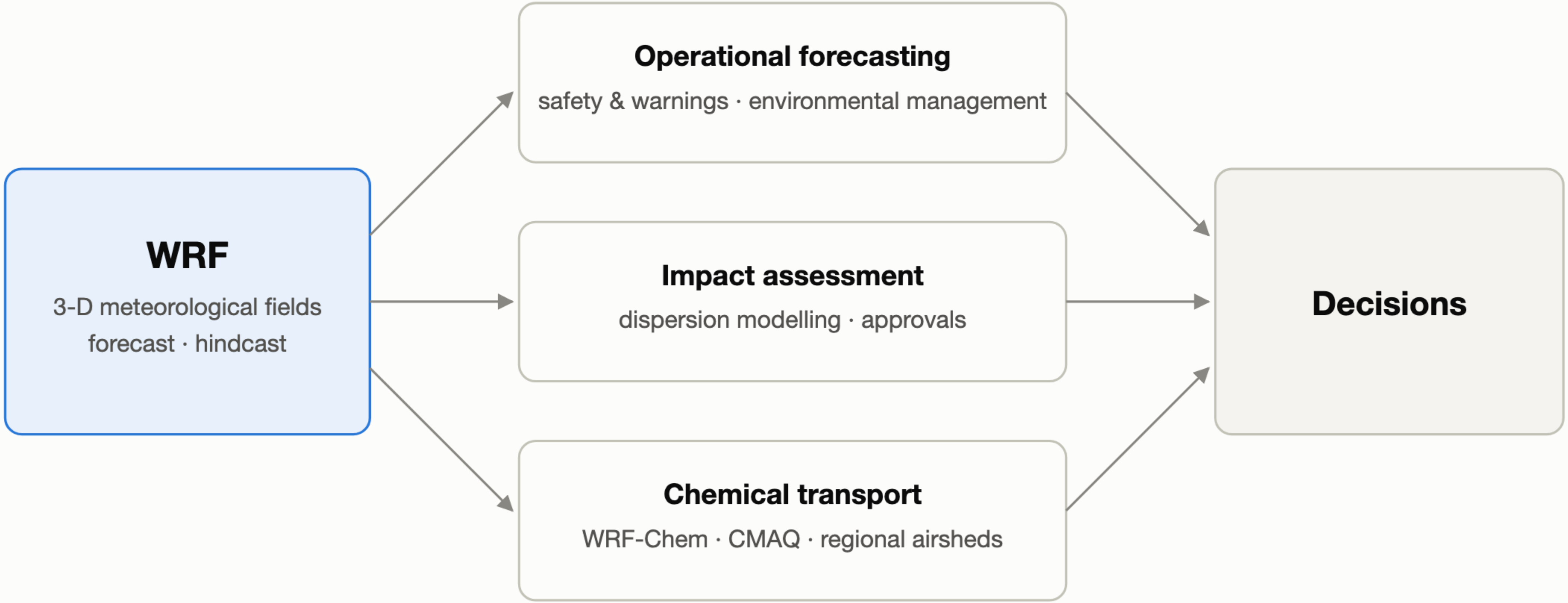
0

The problem

The Weather Research and Forecasting Model (WRF) is powerful and useful but has high technical and scientific barriers to entry.

0 — The problem

Use cases for WRF



0 — The problem

Problem statement

Tuning a WRF configuration across a vast configuration space is difficult.

Automated methods exist (e.g. grid search, Bayesian calibration) but are computationally constrained to searching across a narrow set of levers.

Manual iteration is a reliable method of configuration tuning and is computationally efficient but is constrained by the labour of experienced specialists.

1

The opportunity

Artificial Intelligence, both predictive machine learning and generative AI, is driving advances in scientific research and application.

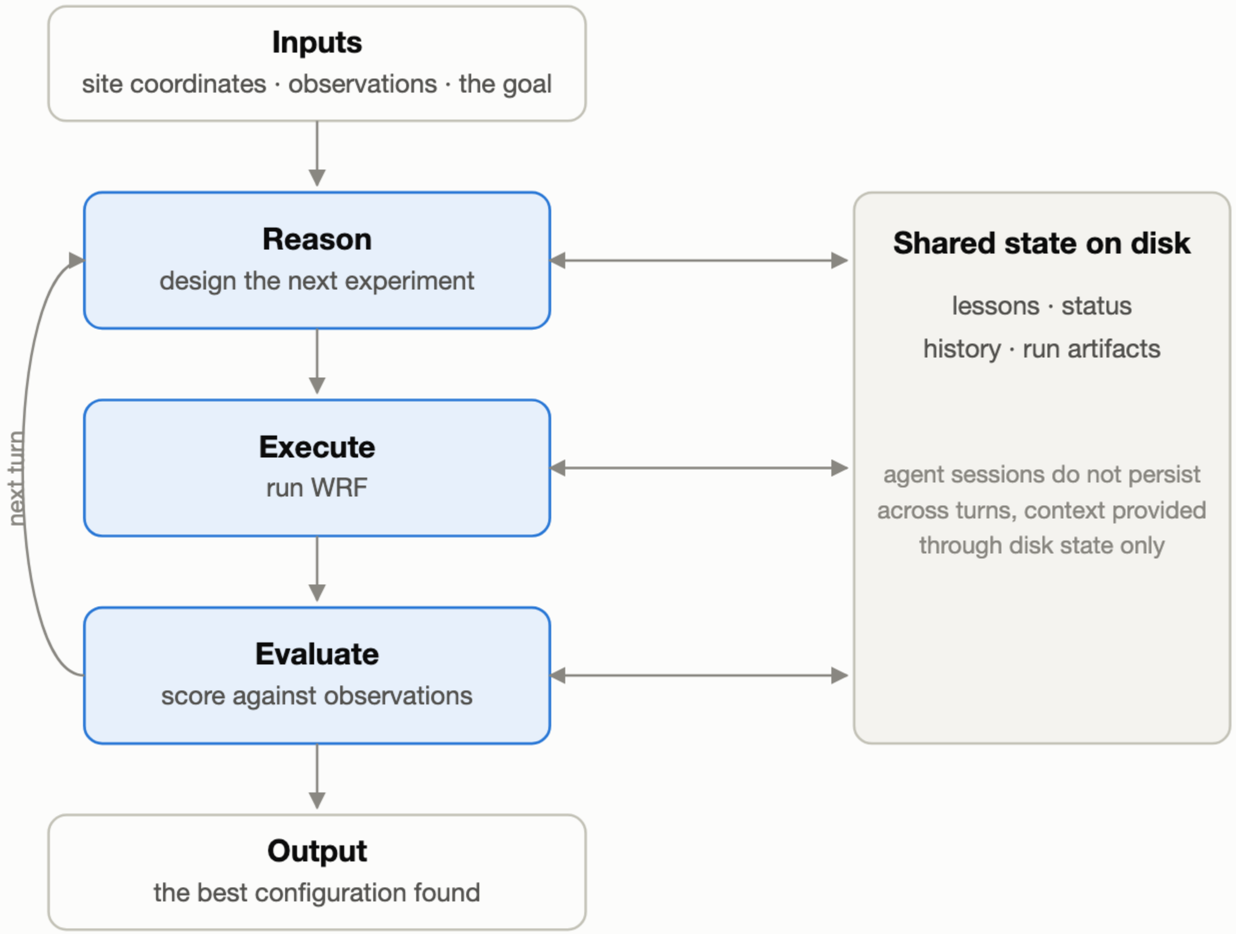
AI in meteorology

PREDICTIVE ML	**Computationally efficient substitute for dynamical downscaling**	**Pure ML forecasting (no physics)** ~1000× energy saving vs physics models ECMWF AIFS
GENERATIVE AI	**Diffusion models for downscaling**	**AI scientist?** an AI agent pursuing an objective

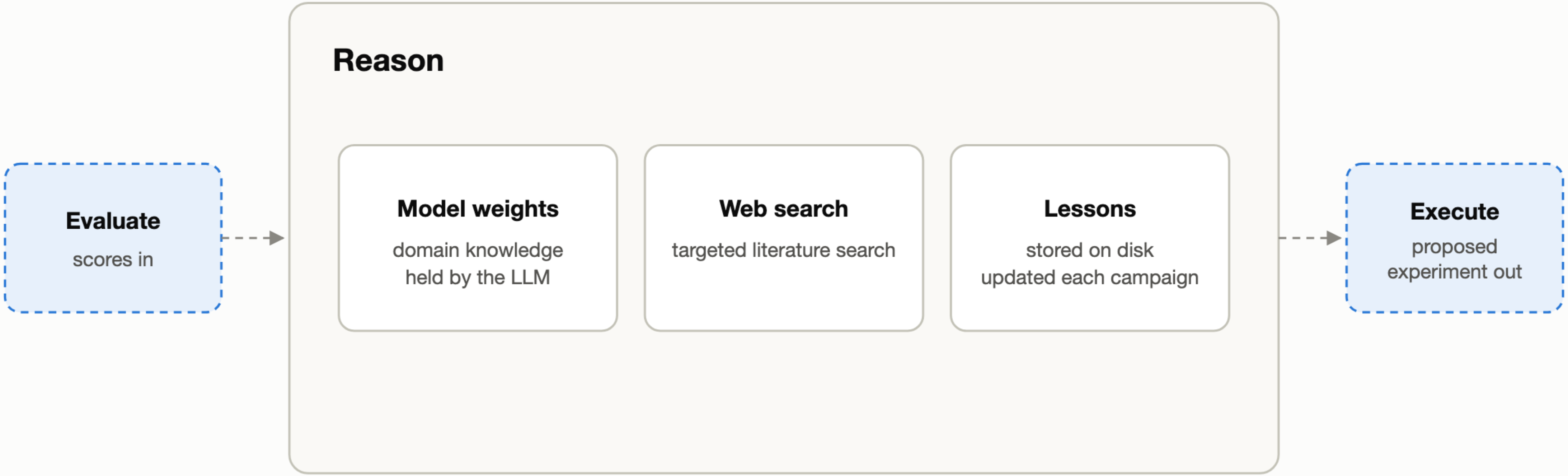
Can an agent loop autonomously produce a strong WRF configuration for a site, with no human involvement?

2

The loop

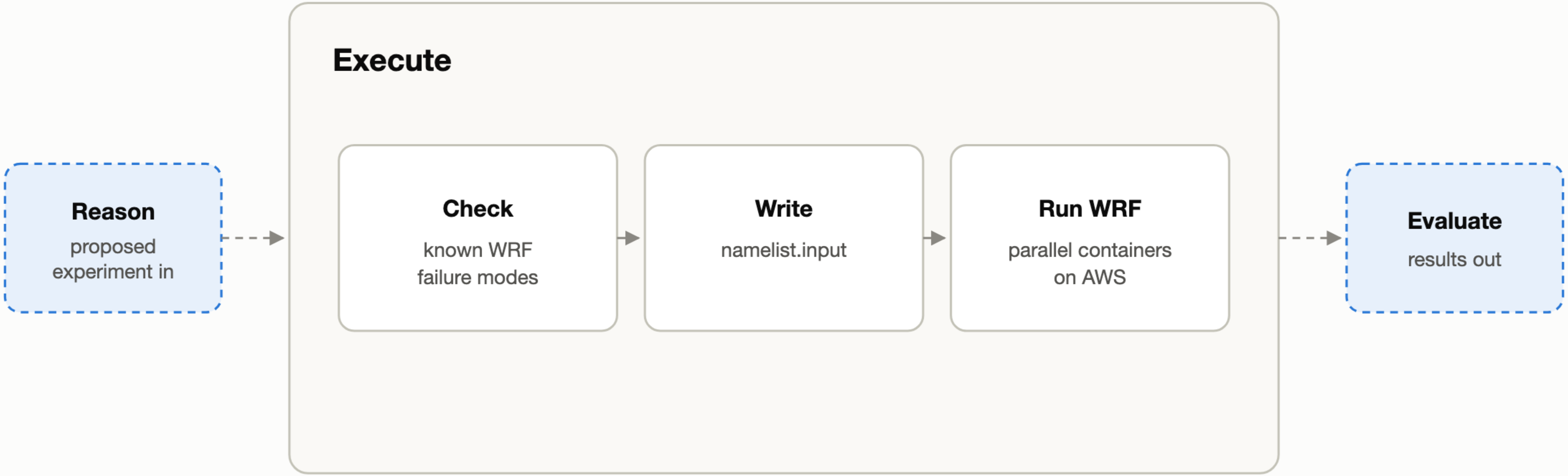

Inputs
site coordinates · observations · the goal
Reason
design the next experiment
Execute
run WRF
Evaluate
score against observations
next turn
Output
the best configuration found
Shared state on disk
lessons · status
history · run artifacts
agent sessions do not persist across turns, context provided through disk state only

How an agent arrives at a proposed configuration

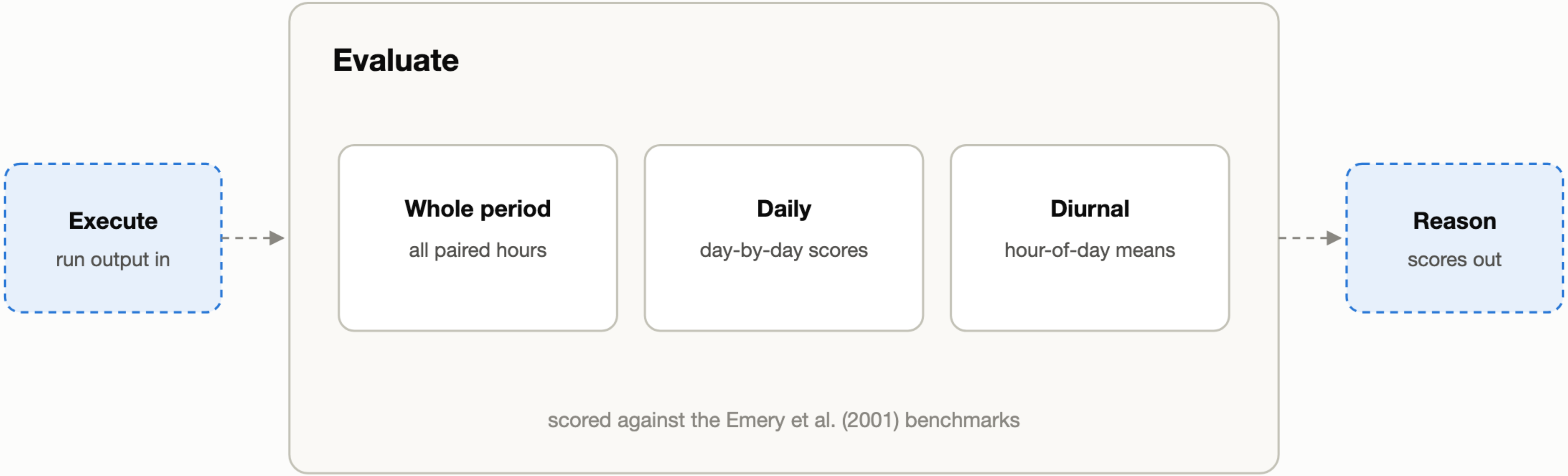


2 — The loop · Execute

How an experiment is executed

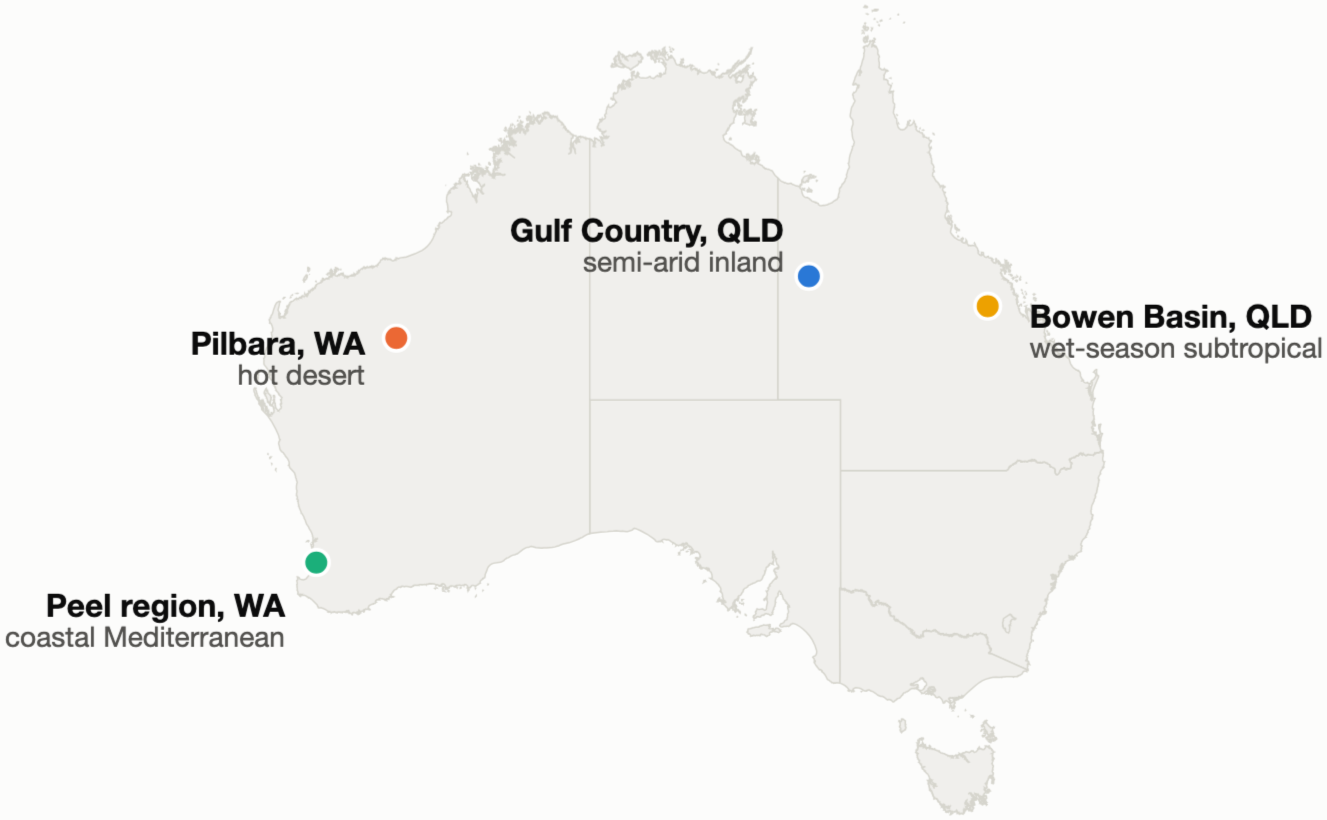


How an experiment is evaluated



3

Campaign design



Design specification

INPUTS	GFS analysis forcing (NCEP, f000) · MODIS 20-class land use (WPS geog) · 1 month of surface observations (2 m T · 10 m WS · 10 m WD)
AGENT & COMPUTE	AWS EC2, 32 physical cores · up to 4 parallel WRF runs at 8 cores each · Claude Opus 4.7 & 4.8
VARIABLES	all WRF namelist.input settings available · domain & nesting changes disincentivised by the wall-clock budget
STOPPING CONDITION	a) a configuration passes all 8 Emery benchmarks versus observations, or b) the 168 h wall-clock budget runs out

4

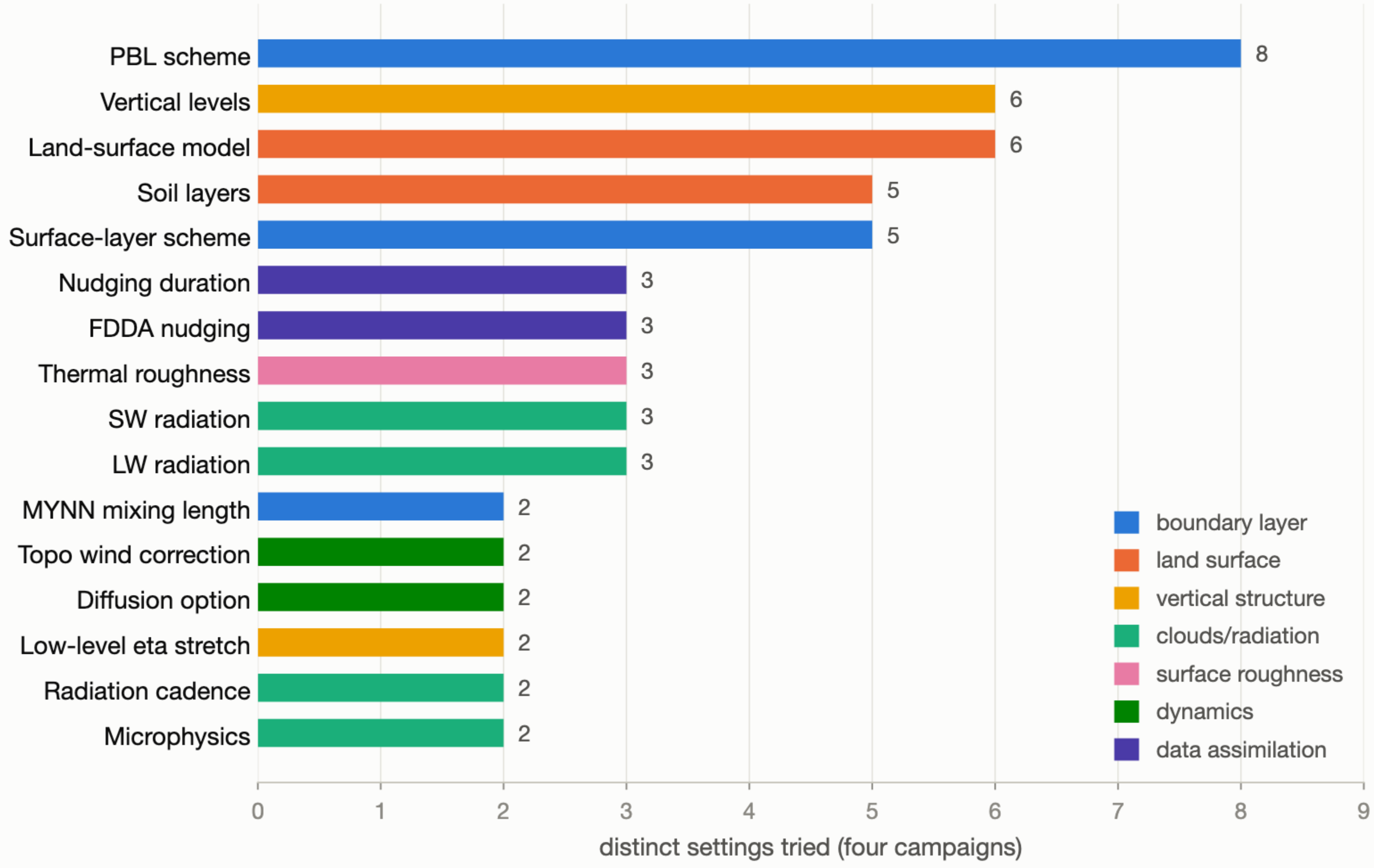
Results

Two of four campaigns passed all eight Emery benchmarks.

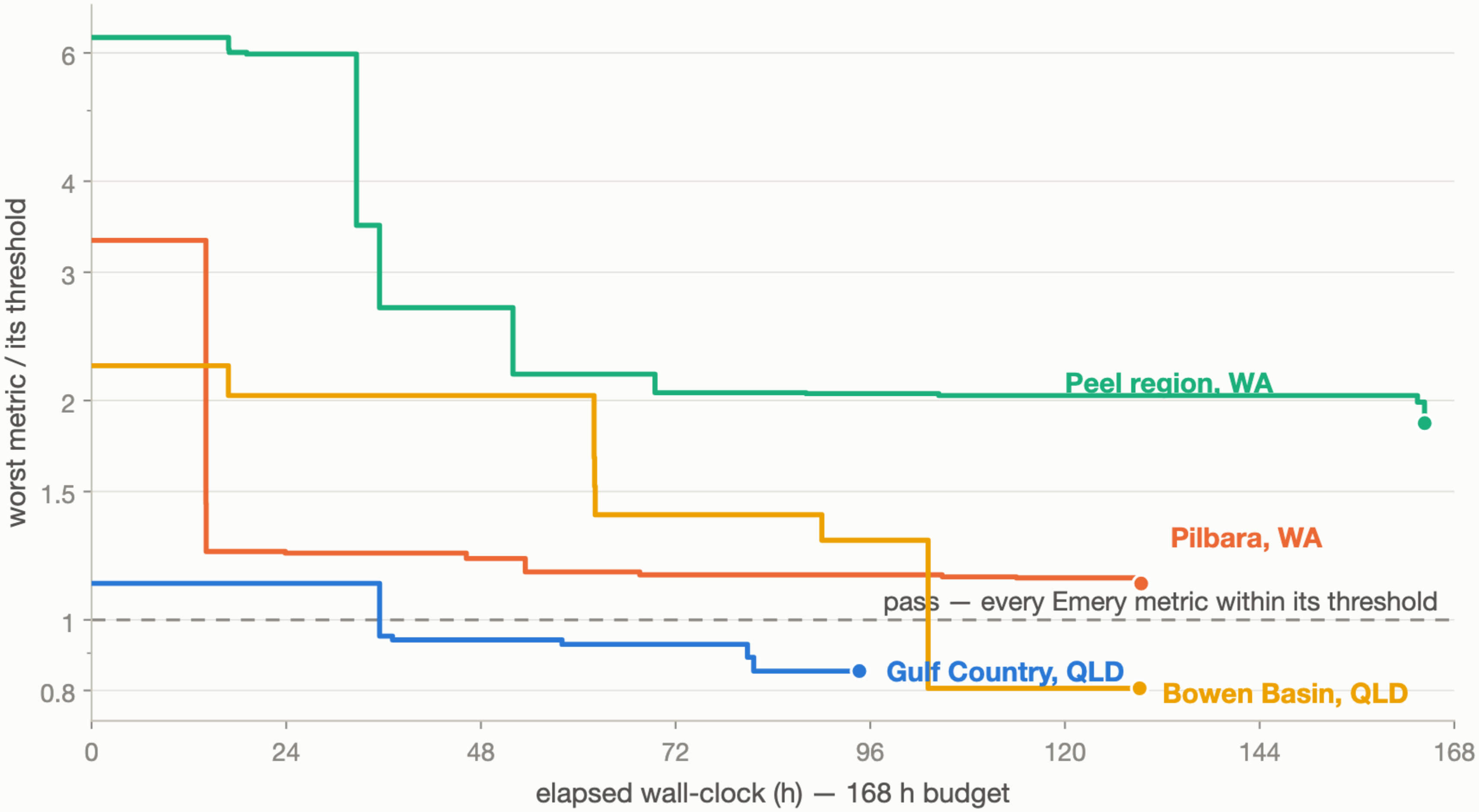
Key performance statistics

	T2m bias °C	T2m GE °C	T2m IOA	WS10 bias m/s	WS10 RMSE m/s	WS10 IOA	WD10 bias °	WD10 GE °	Emery
Emery benchmark	≤ ±0.5	≤ 2.0	≥ 0.8	≤ ±0.5	≤ 2.0	≥ 0.6	≤ ±10	≤ 30	
Gulf Country, QLD	–0.05	1.51	0.97	–0.09	0.91	0.89	1.1	19.5	**8 / 8**
Pilbara, WA	–0.04	1.58	0.98	0.11	0.97	0.93	–6.4	33.6	**7 / 8**
Peel region, WA	–0.01	1.51	0.98	0.93	2.00	0.74	5.8	28.8	**7 / 8**
Bowen Basin, QLD	0.06	1.24	0.95	0.26	1.32	0.83	3.1	22.7	**8 / 8**

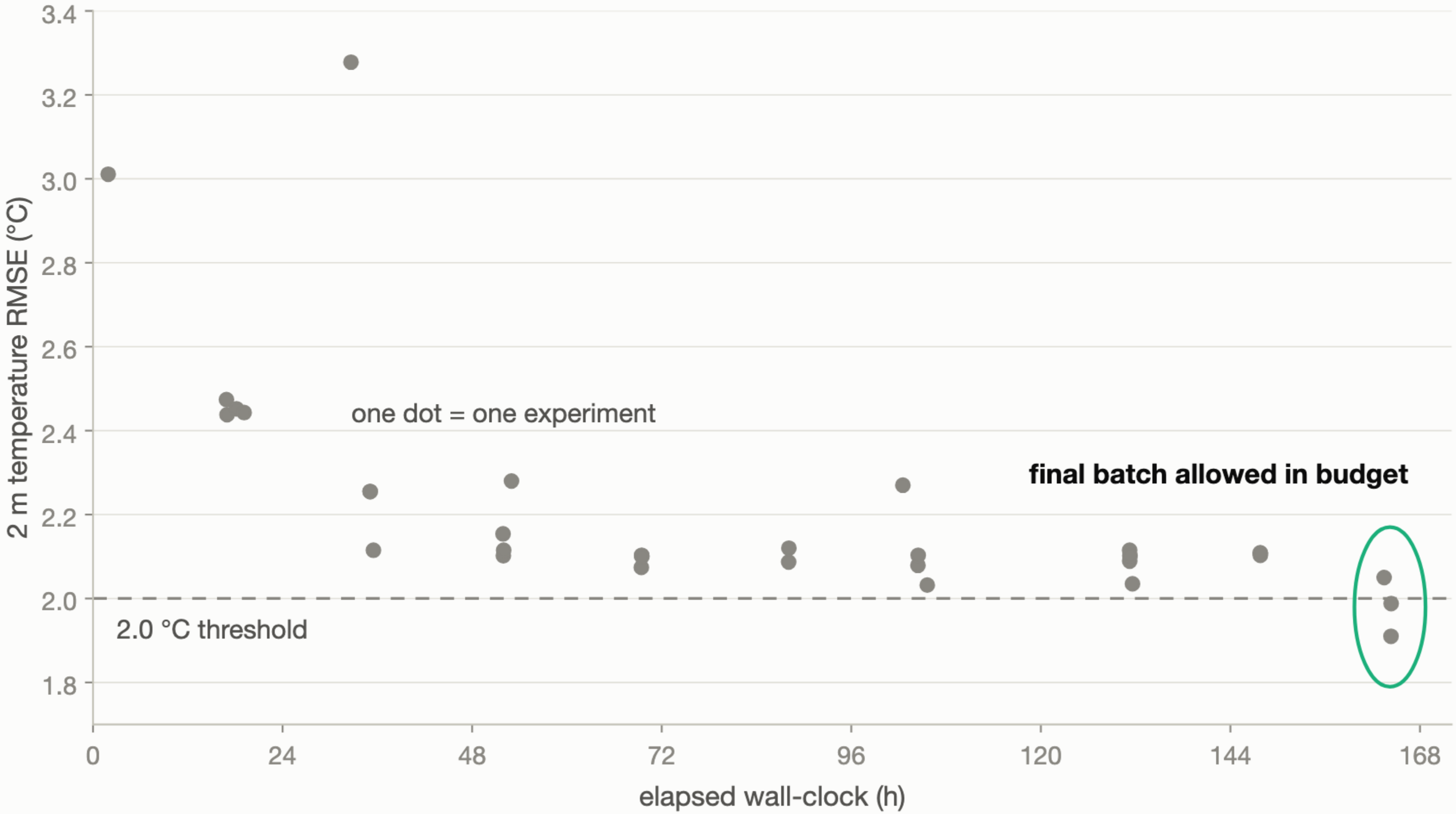
Search effort distribution



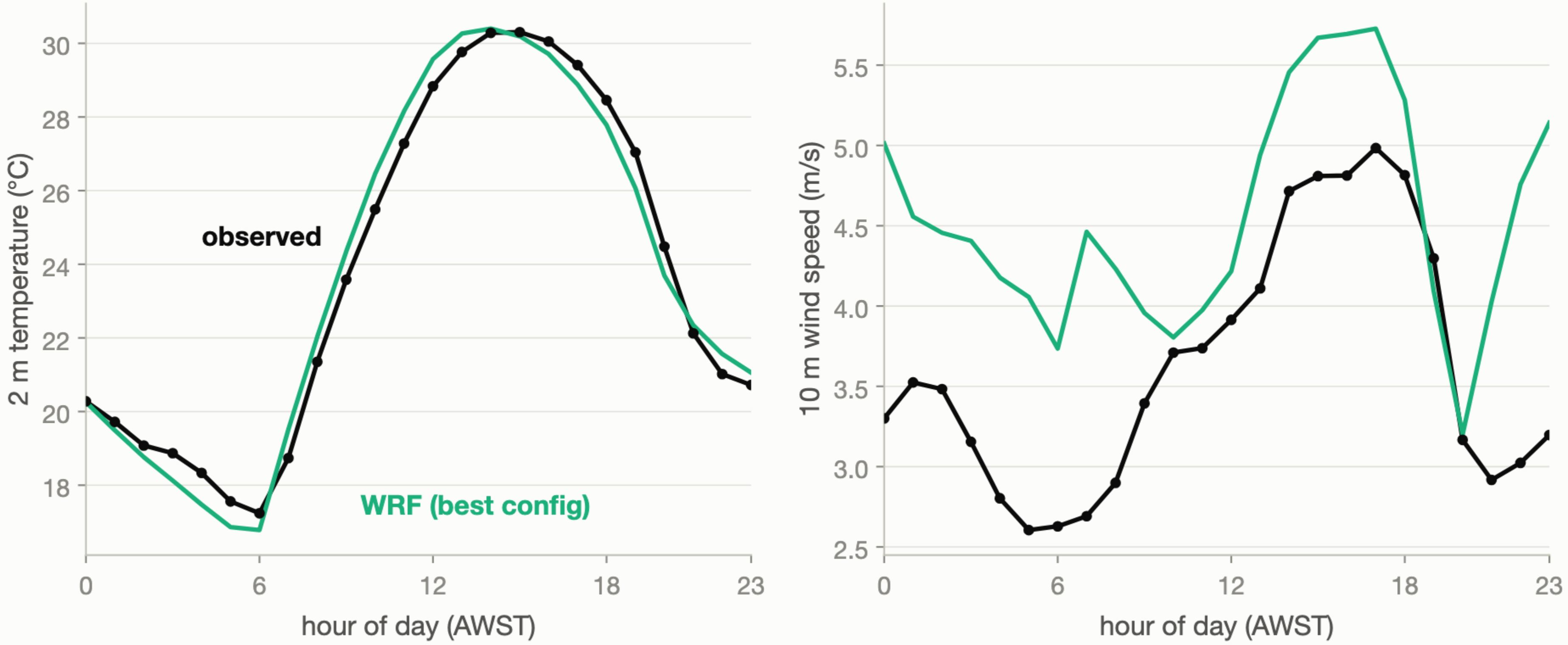
Convergence



Peel campaign convergence



Peel configuration against observations



Peel region, WA · mean diurnal cycle over the scored month

Closing

Findings

1. The agent proposed strong WRF configurations with no human intervention.

2. The agent started cold, with no prior knowledge, no human guidance and a constrained search space. Real-world use would relax these constraints and should reach high-quality configurations at lower cost.

3. The approach should transfer to other meteorological and atmospheric science problems where a deterministic objective can be defined.

Questions

Danny Craggs · danny.craggs@zephyrenviro.com