Joint-embedding predictive representations for sparse-observation Caribbean atmospheric reconstruction

Adrian Dunkley
Climate Studies Group Mona
The University of the West Indies, Mona
adrian.dunkley@mymona.uwi.edu

Draft v0.1 — 31 October 2026
Theory and evaluation protocol; not peer reviewed or submitted

Abstract

This paper specifies a controlled study of whether joint-embedding predictive representations improve event-window atmospheric reconstruction when observations are sparse and the large-scale background is biased. The proposed representation predicts withheld latent targets from available context, then supplies initial states or prior features to the existing D-LEGOS reconstruction class. The study separates representation learning from physical inference: small latent prediction error is insufficient evidence of recovered wind, temperature or humidity. A decoder or observation operator must support independent evaluation in physical units. Comparisons hold architecture capacity, observation access and tuning budget fixed, and distinguish spatial interpolation within an event from transfer to an unseen event. The primary outcomes are withheld-observation errors and moisture-budget diagnostics, with uncertainty assessed by event-level resampling. Negative controls include temporally permuted context, constant embeddings and an equally sized masked-reconstruction encoder. A staged observing-system simulation experiment precedes any real-event comparison. No JEPA model has been trained for this paper; the contribution is the mathematical specification, falsification criteria and reproducible evaluation design. The central question is whether representation pretraining reduces the additional error caused by sparse observations without hiding errors inherited from the background.

Plain-language goal. Test whether learning to predict hidden representations helps reconstruct an event when observations are scarce.

Thesis relationship. Adds a learned prior to the thesis reconstruction objective; it does not replace physical observation operators.

1. Research question and thesis contribution

The thesis asks how exchange between local learned atmospheric fields interacts with observation scarcity and background error. JEPA supplies a candidate prior for that inverse problem. It does not replace the observation model, prescribe the missing source terms or demonstrate that a detailed image is a physically identified state. The proposed addition is an encoder that offers each

local field a structured starting representation. The primary estimand is the paired difference in withheld-observation error between that initialisation and a matched alternative, evaluated over the same event windows.

I-JEPA predicts representations of withheld image regions from context, while V-JEPA 2 extends predictive representation learning to video and action-conditioned applications [1, 2]. Those demonstrations motivate a research question, rather than establish a climate result. Atmospheric variables carry units, different vertical supports and correlated analysis errors. Masking a weather field therefore requires a scientific information contract that is absent from a generic image-completion task.

S2S-JEPA already applies latent predictive learning to subseasonal-to-seasonal forecasting [3]. The present task is retrospective event reconstruction. An open Weather-JEPA research specification also poses a related matched-budget atmospheric representation question [4]. Its stated status is pre-implementation. The present contribution is the use of such a representation within the thesis event-window inverse problem; no claim of the first atmospheric JEPA application is made.

The proposed paper is a thesis extension on learned priors. It shares the parent thesis observation operators and coupling variants. A separate publication is justified only if the representation comparison produces an identifiable contribution beyond renaming the existing foundation-prior study. If the same experiment answers both questions, this material should become a section of that paper.

2. What a representation can and cannot identify

Let x denote the resolved event-window state, and let $y_k = H_k(x) + \epsilon_k$ denote an observation with operator H_k and error ϵ_k. A context encoder produces $z_C = f_\theta(C)$, and a target encoder produces $z_T = f_{\bar{\theta}}(T)$. The predictor g_ϕ receives the context representation and explicit target coordinates. The coordinates specify position, pressure and valid time; they do not reveal a withheld target value.

$$L_{\mathrm{JEPA}} = \frac{1}{|\mathcal{T}|} \sum_{T \in \mathcal{T}} \left\| g_\phi(f_\theta(C), \xi_T) - \mathrm{sg}(f_{\bar{\theta}}(T)) \right\|_2^2 .$$

Here sg stops gradients into the target branch. Latent coordinates are dimensionless learned features. Their squared distance is not a temperature error, an energy residual or a calibrated probability. A constant representation can make a predictive objective uninformative; representation-collapse diagnostics and a controlled regularisation alternative are required [5]. A masked autoencoder offers a reconstruction baseline with a different target, rather than a guaranteed inferior method [6].

An elementary non-identifiability argument sets the limit. If two distinct states have identical available observations and identical context representations, a deterministic predictor must give them the same prediction. It cannot recover which state occurred without further information. This argument does not depend on network size. Evaluating independently withheld physical observations is therefore essential, even when the latent objective becomes very small.

3. Data, support and information access

The first dataset is an observing-system simulation experiment with known truth, documented boundary forcing and an observation network sampled from that truth. A different model or

controlled perturbation must generate the background so the experiment does not become a trivial identical-twin reconstruction. The synthetic observation noise, covariance and missingness rules are frozen before comparisons.

ERA5 is a proposed real-event background [7]. CHIRPS provides a possible rainfall-product comparison at its own accumulation and spatial support [8]. Neither is automatically an independent target if the same product or assimilated instruments shaped the encoder inputs. A source-dependence table will mark the shared inputs. The existing monthly precipitation archive establishes useful prework, but does not establish availability of the required pressure-level event tensors, station observations or independent humidity profiles.

Three mask designs test distinct questions. Random point masking is an interpolation diagnostic. Contiguous geographic masks test missing regions. Entire withheld events test transfer. A fourth sensitivity removes an instrument class, such as rainfall retrievals, to check whether apparent improvement depends on a particular observing system. Stratification by season and terrain is fixed using training data. Report both the fraction of withheld values and their physical support; a percentage alone hides whether an entire mountain region is missing.

4. Connection to the D-LEGOS inverse problem

Each local field maps longitude, latitude, pressure and time to horizontal wind, temperature and specific humidity. The primary thesis formulation remains hydrostatic and restricted to the scales supported by its background. Finer numerical output is not evidence of resolved deep convection. Domain-decomposition precedent motivates local representations, but does not establish their advantage for this atmospheric inverse problem [9].

The representation supplies initial parameters or a finite-variance prior penalty. The physical reconstruction objective then combines observations, background, bounded model discrepancy and shared-interface terms. Weak-constraint assimilation provides the relevant distinction between uncertain dynamics and perfect-model assumptions [10]. The candidate objective is:

$$J(x) = J_{\mathrm{obs}}(x) + J_{\mathrm{bg}}(x) + J_{\mathrm{phys}}(x) + J_{\mathrm{int}}(x) + \lambda_z \|E(x) - z_C\|^2_{\Sigma_z^{-1}}.$$

The map E and covariance Σ_z must be fixed or estimated within the training split. If covariance parameters are learned, include the likelihood normalisation and explicit priors; merely increasing their variances must not erase a bad fit. Compare initialisation alone with a persistent latent penalty to determine which mechanism is responsible for an observed difference.

Physics-informed losses are soft preferences unless a hard equality is built into the parameterisation [11]. Hard downscaling constraints concern specified statistical aggregates, and their precedent should not be described as proof of atmospheric moisture closure [12]. The proposed decoder is assessed against physical observations and independent budget operators, not only against the same residuals used in training.

5. Comparisons that isolate the representation

The experiment contains six matched arms: the background alone; a seasonally and terrain-conditioned statistical reconstruction; the D-LEGOS field with random initialisation; the same

field initialised by a masked autoencoder; the field initialised by JEPA; and JEPA with the latent-prior term retained during inference. A full-domain neural field provides a capacity-matched non-decomposed comparison where feasible. Hyperparameter search, optimisation stopping rules and total compute must be declared for each arm.

Global learned forecasting systems show that useful representations and physical structure can coexist, but their forecasting tasks do not validate this event-window smoother [13, 14]. WeatherBench 2 supplies benchmarking context for separating variables, scales and verification choices [15]. The real-event test here remains a reconstruction study because in-window observations after the reconstructed time may be used. A future-only forecast arm would require a different information cutoff and a separate analysis.

Ablations remove coordinates, physical penalties, shared fluxes or pretraining in turn. A temporally shuffled context tests whether the encoder uses event evolution. A constant embedding tests whether the downstream field ignores the representation. A frozen random encoder controls for additional features. If these controls perform as well as the trained encoder, the result would weaken the proposed mechanism even if all neural variants beat an inadequate background.

6. Outcomes, uncertainty and stopping rules

The primary score is an event-weighted, variable-specific error on observations excluded from both pretraining fine-tuning and reconstruction. Wind, temperature and humidity retain their own units; a combined standardised score is secondary and uses training-only scales. Report location, level and event errors alongside the aggregate. A representation that improves temperature while damaging humidity should not be hidden behind one weighted number.

Moisture diagnostics include the domain budget residual, shared-face mismatch, negative-humidity frequency and sensitivity to background bias. These measure different properties. A zero internal accounting residual does not show that transport is correct. Independent discretisations and withheld collocation points reduce the chance of reporting the training loss as external verification. Known PINN failure modes motivate explicit adverse cases, rather than assuming a larger loss weight repairs the inference [16].

If the method returns ensembles, assess calibration and a proper probabilistic score, retaining the distinction between probabilistic forecasts and conditional reconstruction distributions [17]. The resampling unit is the event, with additional spatial blocking where warranted. The number of events and minimum practically relevant difference must follow an observing-system pilot and be frozen before the final held-out comparison; neither is invented here.

Stop the scientific comparison if target leakage is found, if withheld instruments share an undocumented preprocessing path with inputs, or if preprocessing was fitted on the test events. A negative comparison is reportable. The method does not earn a claim of benefit merely by converging or producing smoother maps.

7. Reproducibility and available prework

Existing prework includes the D-LEGOS research-design chapter, the paper series on coupled fields and observation support, archived monthly Jamaican rainfall comparisons and a foundation-prior temperature study. Their outputs must remain separate from this experiment. No archived score

should be relabelled as a JEPA result. The current web companion is a constructed masking instrument: it explains context and targets without claiming a trained encoder.

The required run package will contain acquisition manifests, dataset versions, event lists, spatial masks, preprocessing statistics, model configurations, seeds, frozen checkpoints, unit-level predictions where permitted, observation operators and exact commands. Hash the training and evaluation manifests before fitting. Failed runs remain in the log with their stopping reason. A fresh evaluation command must fail clearly if an input is missing, rather than generate a plausible replacement.

8. Interpretation and limitations

A useful result would show that a specified representation reduces a specified withheld physical error under the matched information budget. It would not show that JEPA has learned the full climate system, recovered unobserved convection or made long-term climate projections. Reconstruction skill within the ERA5-resolved class cannot be extrapolated automatically to kilometre-scale processes or another island.

The main alternative explanation is that pretraining transfers the background's regularities while concealing its bias. Background perturbations, source-class holdouts and adverse controls therefore have equal standing with the headline comparison. Another possibility is that pretraining is simply an expensive initialiser that offers no advantage over a smaller seasonal model. Compute and data requirements belong in the result, even if the physical score improves.

9. Current conclusion

The contribution at this checkpoint is a falsifiable representation-learning extension to the D-LEGOS reconstruction programme. Its equations define the separation between latent prediction and physical identification; its protocol identifies the data and comparisons needed to test usefulness. Empirical improvement remains unreported until those runs exist.

Data, authorship and assistance

Adrian Dunkley is the sole named author. Climate affiliation: Climate Studies Group Mona, Department of Physics, The University of the West Indies, Mona. Contact: `adrian.dunkley@mymona.uwi.edu`. The draft extends the author's thesis direction and existing research programme. AI assisted literature organisation, drafting and the interactive companion; the author must review claims, affiliations and declarations before external submission. No new participant data, climate-model training or independent reproduction is claimed. Some proposed references have only abstract or bibliographic checks, as identified in the source record; full-text verification remains a submission requirement. Funding, conflicts and final institutional declarations require author confirmation.

References

[1] Assran et al. (2023). Self-Supervised Learning from Images with a Joint-Embedding Predictive Architecture. `https://arxiv.org/abs/2301.08243`.

[2] Assran et al. (2025). V-JEPA 2: Self-Supervised Video Models Enable Understanding, Prediction and Planning. https://arxiv.org/abs/2506.09985.

[3] Dong and Mengaldo (2026). S2S-JEPA: Predicting the Predictable at Subseasonal-to-Seasonal Timescales. https://arxiv.org/abs/2610.03106.

[4] Weather-JEPA project (2026). Weather-JEPA: atmospheric representation-learning research specification (accessed 31 October 2026). https://kdsmithtexnc.github.io/weather-jepa/.

[5] Bardes, Ponce and LeCun (2022). VICReg: Variance-Invariance-Covariance Regularization for Self-Supervised Learning. https://arxiv.org/abs/2105.04906.

[6] He et al. (2022). Masked Autoencoders Are Scalable Vision Learners. https://arxiv.org/abs/2111.06377.

[7] Hersbach et al. (2020). The ERA5 global reanalysis. https://doi.org/10.1002/qj.3803.

[8] Funk et al. (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. https://doi.org/10.1038/sdata.2015.66.

[9] Moseley, Markham and Nissen-Meyer (2021). Finite Basis Physics-Informed Neural Networks (FBPINNs): a scalable domain decomposition approach for solving differential equations. https://arxiv.org/abs/2107.07871.

[10] Trémolet (2006). Accounting for an imperfect model in 4D-Var. https://doi.org/10.1256/qj.05.224.

[11] Raissi, Perdikaris and Karniadakis (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. https://doi.org/10.1016/j.jcp.2018.10.045.

[12] Harder et al. (2023). Hard-Constrained Deep Learning for Climate Downscaling. https://jmlr.org/papers/v24/23-0158.html.

[13] Kochkov et al. (2024). Neural general circulation models for weather and climate. https://doi.org/10.1038/s41586-024-07744-y.

[14] Lam et al. (2023). Learning skillful medium-range global weather forecasting. https://doi.org/10.1126/science.adi2336.

[15] Rasp et al. (2024). WeatherBench 2: A benchmark for the next generation of data-driven global weather models. https://arxiv.org/abs/2308.15560.

[16] Krishnapriyan et al. (2021). Characterizing possible failure modes in physics-informed neural networks. https://arxiv.org/abs/2109.01050.

[17] Gneiting and Raftery (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. https://doi.org/10.1198/016214506000001437.

Reference review record

The following record distinguishes primary abstract checks from proposed references requiring additional reading. It is not a statement that all full texts were verified.

jepa Primary abstract inspected; image representation learning, not climate validation.

vjepa Primary abstract inspected; video and robot tasks, not atmospheric evidence.

s2sjepa Primary arXiv abstract and metadata inspected through indexed search; preprint dated 2 October 2026; direct full-text retrieval disabled; source-reported forecast results not reproduced or transferred to this reconstruction protocol.

weatherjepa Primary project specification inspected; undated, accessed 31 October 2026; self-described pre-implementation research, not a peer-reviewed result.

vicreg Primary abstract inspected; representation-collapse motivation.

mae Primary abstract inspected; masked reconstruction baseline.

era5 Publisher article inspected; dataset identity and reanalysis status.

chirps Source identity checked against existing archive; publisher full text inaccessible in this pass.

fbpinn Primary abstract inspected; domain-decomposition precedent.

4dvar Publisher abstract inspected; weak-constraint model-error formulation; issue year 2006, online publication 2007.

pinn Existing source register verified; detailed full-text claim check retained as submission work.

harder Publisher abstract inspected; statistical hard constraints distinct from atmospheric moisture closure.

neuralgcm Existing register identity; current publisher full text inaccessible.

graphcast Existing register identity; current publisher full text inaccessible.

weatherbench Primary record inspected; variable/scale-specific benchmarking precedent.

failure Primary abstract inspected; failure-mode motivation.

scoring Bibliographic identity from existing manuscript; publisher inaccessible in this pass.