Agent-based local atmospheric models with a shared physical exchange ledger

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Theory and evaluation protocol; not peer reviewed or submitted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 formulates an agent-based interpretation of D-LEGOS in which each atmospheric subdomain maintains a numerical state, processes permitted observations and exchanges physically defined quantities with neighbouring subdomains. The agents are scientific numerical modules; language-model conversation is not a transport law. A shared oriented interface ledger guarantees cancellation of internal transfers in the summed discrete budget. That identity alone does not establish accurate fluxes, positive humidity, stable integration or an identifiable inverse solution. The proposed study compares uncoupled fields, state-matched interfaces, penalised one-sided fluxes and one shared signed transfer under matched observations, capacity and compute. It introduces explicit tests of update order, stale messages and background bias, followed by a sparse-observation reconstruction comparison in a controlled observing-system simulation experiment. A separate supervisory workflow may prepare manifests and inspect failures, but cannot alter the governing equations or publish results without review. The present contribution is a mathematical model class and an evaluation protocol. No agentic atmospheric reconstruction has been trained for this paper, and the interactive ledger is a constructed teaching model. The paper is a candidate extension of the thesis coupling chapter; it should be merged into that chapter if it does not yield a separate scheduling or communication contribution.

Plain-language goal. Test whether local numerical agents can coordinate reconstruction through one physically defined exchange record.

Thesis relationship. Extends the coupling variants with reproducible update schedules, delayed-message tests and communication costs.

1. What the agents represent

An atmospheric agent is a computational module associated with a physical region and a stated time interval. Its state contains resolved variables and uncertainty, not preferences or a conversational persona. Its actions are numerical updates permitted by the discretisation and inference objective. Neighbours exchange typed messages with units, face geometry, valid time, version and provenance. An accepted message cannot silently change from a mass transfer into an arbitrary learned feature.

This definition borrows the explicit entity, process and scheduling description encouraged by the ODD simulation protocol [1]. Graph networks and learned physical simulators provide relevant precedents for structured interaction [2, 3]. The present question concerns which interface and scheduling restrictions improve a sparsely observed atmospheric reconstruction. Calling a pre-existing domain decomposition an agent system would not, by itself, constitute novelty.

The thesis already compares physically specified coupling variants. This paper earns an independent contribution only if it studies effects that a single synchronously evaluated graph would not answer: delayed messages, unequal local convergence, communication cost, adaptive refinement or reproducible asynchronous execution. The physical meaning of every exchange remains the same in both implementations.

2. Numerical state and information contract

Let region i have horizontal footprint Ω_i and water-vapour mass M_i in kilograms. A local hydrostatic field estimates horizontal wind, temperature and specific humidity in pressure coordinates over the reconstruction window. Terrain intersections are masked, and the observation operator integrates or samples according to the instrument's support. ERA5 supplies a proposed background, not an independently known atmospheric truth [4].

The primary task is retrospective smoothing. All permitted observations inside the event window may inform a reconstructed earlier state, except those reserved for verification. A forecast experiment would restrict observations to those available at issue time, including product latency, and would require a separate evaluation. CHIRPS-based rainfall comparisons in the existing archive do not establish event-level humidity or wind reconstruction [5].

Each message contains the source and destination region, shared-face identifier, orientation, represented area, interval start and end, quantity, units, uncertainty and state-version identifiers. Reject a message if these fields do not match the receiving agent's current discretisation. A duplicated message must be idempotent: accepting it twice cannot move mass twice. This requirement is numerical bookkeeping rather than evidence that the estimated physical transfer is correct.

3. Shared ledger and cancellation proposition

Choose one orientation for each internal interface f. Let B_{if} equal $+1$ or -1 for its two incident regions and zero elsewhere. Let F_f^n be the integrated transfer rate during step n, in kilograms per second, with the orientation fixed by that convention. Let S_i^n represent specified external sources and sinks in the same units. The discrete budget is:

$$M_i^{n+1} = M_i^n - \Delta t \sum_f B_{if} F_f^n + \Delta t S_i^n.$$

For every internal interface, $\sum_i B_{if} = 0$. Summing the update over regions therefore removes the internal-transfer term. The domain mass change equals the recorded external boundary exchange plus sources and sinks, up to floating-point error. This is a direct accounting proposition, not a discovery of a new conservation law. Finite-volume methods supply its established numerical context [6].

The proposition assumes that both regions use the same stored transfer, time interval and orientation. Independently predicted outward fluxes do not satisfy the identity unless their sum vanishes. A penalty can encourage that agreement, while an explicit shared variable enforces the discrete cancellation. Conservative PINNs and parallel domain-decomposed PINNs are relevant precedents to inspect when specifying the baseline implementation [7, 8].

Cancellation does not ensure $M_i^{n+1} \geq 0$. An explicit update can remove more water than a region contains. A donor-based restriction or another positivity-preserving numerical method must be specified and tested separately. It also does not ensure stability, correct transport direction or useful rainfall inference. The manuscript reports these properties in separate columns rather than describing the ledger as a universal physical guarantee.

4. Four coupling variants and a scheduling extension

Variant C0 fits each local field independently. C1 penalises differences in state at a shared interface. C2 gives each side its own outward transfer and penalises their inconsistency. C3 stores one oriented transfer used by both incident regions. These are the existing thesis coupling choices, restated so the agent implementation can be checked against the same mathematical objects.

The new scheduling comparison uses a synchronous reference, a deterministic staggered update and a bounded-delay exchange. For a bounded-delay run, record the age of every received state and the interval for which a transfer was negotiated. If the implementation uses an old state to compute a new transfer, both agents still book that exact transfer in the same ledger interval. If that cannot be ensured, the run is classified as an inconsistent communication variant rather than a conserving implementation.

A rollback within a numerical solver is different from replacing an inconvenient scientific result. Any local rejection must restore both incident ledger entries atomically, and the log records the attempted step. Unequal local time steps require a common exchange interval or a conservative subcycling construction. Reusing one instantaneous flux across mismatched intervals would violate the stated proposition even if the interface variable has a reassuring name.

5. Local inference and unresolved processes

Local states are fitted against observations, background and bounded physical discrepancy. The discrepancy terms represent unresolved tendencies and imperfect equations. They are prescribed or restricted in the first comparison; allowing arbitrary source fields would let a poor reconstruction fit its own residuals. Weak-constraint 4D-Var provides the relevant model-error context [9].

The local objective can be written schematically as:

$$J = \sum_i \left(J_{\mathrm{obs},i} + J_{\mathrm{bg},i} + J_{\mathrm{dyn},i}\right) + J_{\mathrm{interface}}.$$

Observation-error covariance must account for correlated instruments. A local agent cannot treat every retrieved pixel as independent merely because each arrives in a different message. Physics-informed objectives motivate this joint fitting class, but their optimisation and approximation limits require explicit controls [10, 11]. A low residual can coexist with a wrong state when the inverse problem is weakly identified.

Hard statistical constraints on downscaled rainfall remain a separate object [12]. Exact preservation of a supplied coarse precipitation average does not establish water-vapour budget closure across a three-dimensional atmosphere. The proposed ledger is evaluated with physical units and specified external terms, rather than borrowing the accuracy claims of the existing monthly rainfall experiment.

6. Experimental stages and baselines

Stage one uses a deterministic transport or exchange problem with known truth and a mesh-refinement comparison. It tests cancellation, non-negativity, orientation, duplicate-message handling, update-order sensitivity and numerical convergence. Introduce deliberate stale messages and one-sided bookings so the checks can demonstrate their ability to detect failure. Passing this stage establishes software properties for that constructed model.

Stage two uses an observing-system simulation experiment with an imperfect background, controlled observation scarcity and known state. Freeze the event windows, observation masks, error covariance, model capacity and compute budgets before fitting. Compare C0–C3, a full-domain neural field, a statistical reconstruction and a conventional assimilation reference where implementation is feasible. Report any unmatched cost directly.

Stage three is conditional on acquiring the required real observations and quality-controlled pressure-level inputs. It withholds complete events and instrument classes. WeatherBench-style evaluation provides useful guidance on explicit variables and verification scales, while this task remains a smoother [13]. GraphCast is a forecasting precedent, not a directly interchangeable task or a source of local performance claims [14].

The agent-versus-graph comparison must match the physical update and available information. If an asynchronous method simply uses a newer observation that the synchronous baseline was denied, the difference answers an information-access question. It cannot be attributed to agency. A strong sequential implementation also controls whether additional orchestration is useful at the proposed domain size.

7. Metrics and adverse tests

Primary scientific outcomes are withheld wind, temperature and humidity errors, plus rainfall-related diagnostics at their true support. Budget outcomes include domain drift, external-boundary closure and shared-face mismatch. Computational outcomes include wall time, peak memory, communication volume, rejected updates and age of consumed messages. Count hardware and orchestration overhead, rather than reporting only local network evaluation time.

For probabilistic outputs, use a proper score and assess calibration at the variable and event level [15]. Resampling must respect event and spatial dependence. A scheduling policy that produces rare catastrophic failures cannot be represented adequately by mean error alone. Report the worst observed event and all rejected runs with their input conditions.

Adverse cases include reversed interface signs, duplicated transfers, missing external sources, excessive time steps, delayed state updates, background bias and removal of a strategically important observing site. An accounting test should fail when a deliberately inconsistent transfer is injected. A reconstruction test should remain capable of finding that the statistically simpler baseline is better. The expected outcome is not prescribed.

8. A bounded role for language-model agents

A language-model assistant may prepare a data manifest, propose a configuration change, explain a failed run or compare retained outputs. It does not estimate a physical flux by conversation, bypass dataset licences or turn a planned experiment into a completed result. Any configuration change produces a new version with an explicit reason and requires approval before the held-out evaluation is reopened.

The scientific simulator and the research assistant therefore have different action spaces. Numerical agents operate within the solver. Research assistants operate around the experiment, with file and execution permissions controlled separately. A useful future workflow study could compare this assistance against a conventional script, but that would require its own outcome measures and cannot serve as validation of atmospheric dynamics.

9. Reproducibility, overlap and current conclusion

Prework includes the thesis coupling specification, shared-ledger numerical demonstrations, archived model comparisons and the existing world-model interface. These establish a starting implementation context. They do not establish that an agent-based atmosphere has reconstructed a Caribbean storm. The web instrument uses three constructed chambers and two shared transfers; its unit total is an accounting demonstration.

The release package will preserve the region graph, face geometry, unit conventions, update schedule, message log, source and configuration hashes, checkpoints, held-out predictions and one command for reproducing the primary comparison. The ODD account must be detailed enough to reimplement the schedule without relying on a product name. Experimental evidence is unreported at this checkpoint.

The model class is useful only if explicit local interaction improves an identified scientific or computational outcome. Its strongest current property is the stated cancellation identity. The proposed experiments determine whether the additional scheduling machinery contributes enough to justify a separate paper beyond the parent thesis coupling study.

Data, authorship and assistance

Adrian Dunkley is the sole named author. Affiliation: Climate Studies Group Mona, Department of Physics, The University of the West Indies, Mona. Contact: `adrian.dunkley@mymona.uwi.edu`. This is a theory and protocol draft, not an operational forecasting system. AI assisted drafting, source organisation and the interactive companion. The author must verify the numerical specification, bibliography, institutional relationship, funding and declarations before external submission. Proposed references with incomplete full-text checks are marked in the source record. No participant study, new trained atmospheric model or government deployment is asserted.

References

[1] Grimm et al. (2020). The ODD Protocol for Describing Agent-Based and Other Simulation Models: A Second Update to Improve Clarity, Replication, and Structural Realism. https://jasss.soc.surrey.ac.uk/23/2/7.html.

[2] Battaglia et al. (2018). Relational inductive biases, deep learning, and graph networks. https://arxiv.org/abs/1806.01261.

[3] Sanchez-Gonzalez et al. (2020). Learning to Simulate Complex Physics with Graph Networks. https://arxiv.org/abs/2002.09405.

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

[5] 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.

[6] LeVeque (2002). Finite Volume Methods for Hyperbolic Problems. https://doi.org/10.1017/CBO9780511791253.

[7] Jagtap, Kharazmi and Karniadakis (2020). Conservative physics-informed neural networks on discrete domains for conservation laws: Applications to forward and inverse problems. https://doi.org/10.1016/j.cma.2020.113028.

[8] 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.

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

[10] 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.

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

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

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

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

[15] 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.

odd Publisher abstract and introduction inspected; explicit simulation reporting.

relational Primary abstract inspected; graph-network structural motivation.

gns Primary abstract inspected; particle message-passing simulator precedent.

era5 Publisher article inspected; dataset identity and reanalysis status.

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

leveque Book identity checked; primary book contents remain a pre-submission reading requirement.

cpinn Bibliographic record checked; 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.

failure Primary abstract inspected; failure-mode motivation.

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

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

graphcast Existing register identity; current publisher full text inaccessible.

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