From atmospheric evidence to accountable adaptation choices: a climate-policy framework for small-island institutions

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Theory and evaluation protocol; not peer reviewed or submitted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 develops an evaluation framework for moving from climate-model evidence to adaptation decisions while preserving uncertainty, source lineage and public accountability. It extends Adrian Dunkley's existing human-governed agentic climate-governance design; it does not claim that a government has adopted or evaluated the system. The framework separates an evidence record, a decision model, an institutional authority record and a monitoring plan. Atmospheric reconstruction can inform event understanding, but cannot alone identify intervention effects or support long-term adaptation probabilities. Candidate actions are compared across explicitly constructed scenarios, with distributional constraints and sensitivity to disputed costs. A retrieval and agent workflow may assemble evidence and expose disagreements; an accountable person retains authority over assumptions, recommendations and consequential actions. The proposed evaluation compares conventional analysis, single-assistant retrieval and bounded multi-agent assistance using blinded assessment of source traceability, error detection, decision reproducibility and review effort. The contribution is a testable framework and study protocol. No policy-effect estimate, avoided-loss total or measured administrative benefit is reported. Its thesis relationship is translational: it specifies which physical outputs are fit for a decision, where additional exposure and vulnerability evidence is required, and how scientific limitations remain visible through the policy workflow.

Plain-language goal. Show what climate evidence can support a policy choice, what is missing and who remains responsible for the decision.

Thesis relationship. Translates reconstruction limits and uncertainty into evidence-fitness checks; it does not estimate policy effects from weather fields.

1. The decision problem

An adaptation decision joins several claims that a climate model cannot answer on its own. A reconstruction may describe the conditions of a past rainfall event. It does not identify which drainage investment would have prevented a loss, who would benefit, whether maintenance will

continue or which institution may lawfully approve expenditure. Treating a physically detailed animation as a complete policy analysis would hide those missing links.

The framework begins with a named decision, an accountable authority, a jurisdiction, an action horizon and affected groups. For example, a drainage programme must identify the assets and communities considered, alternative maintenance and construction choices, financing constraints and the period over which performance matters. The framework does not assign a universal island risk score or infer local implementation capacity from a regional average. The IPCC small-islands assessment emphasises place-specific adaptation and institutional constraints [1].

This is a translation study connected to the thesis through evidence fitness. The thesis tests reconstruction under sparse observations and biased backgrounds. The policy framework asks whether a proposed decision actually depends on the state variables, scale and uncertainty that the reconstruction can support. A decision needing future flood frequencies requires additional climate, hydrological, exposure and vulnerability work. That dependency is recorded rather than filled with an unsupported extrapolation.

2. Four linked records

The evidence record stores each claim, source, version, location, date, spatial and temporal support, transformation, uncertainty and review status. A source pointer identifies the supporting passage or output; a URL alone is insufficient when a long report contains several conflicting qualifications. Observed data, author calculations, scenario assumptions and untested propositions receive different labels.

The decision record defines alternatives, objectives, constraints, evaluation horizon and consequences. It includes the action of postponing, gathering information or maintaining existing assets. The authority record identifies who supplies assumptions, approves analysis, makes the decision and can challenge it. The monitoring record specifies signposts, review dates and conditions that would reopen the action choice. These records have different purposes and should not become four copies of the same narrative.

Adaptation guidance and disaster-risk frameworks provide relevant context for iterative planning, risk understanding and institutional responsibilities [2, 3]. The Paris Agreement and the IPCC assessment supply wider policy context, without certifying this design's legal adequacy in any jurisdiction [4, 5]. Local officials and counsel must verify the actual decision rights and applicable requirements before a pilot handles consequential choices.

3. Evidence fitness and uncertainty propagation

Every proposed model input passes a fitness question: does its variable, support, horizon and validation match the intended decision? Monthly rainfall allocation may support one question about spatial averages while being inadequate for a culvert-design extreme. A retrospective smoother using observations after an event is useful for reconstruction, but cannot be scored as if those observations were available to an advance warning.

Uncertainty is decomposed into observational error, model discrepancy, scenario uncertainty, exposure uncertainty, intervention-effect uncertainty and value disagreement. These categories need different responses. Collecting another weather observation may reduce state uncertainty; it does not resolve

disagreement about whose costs deserve priority. IPCC decision-making guidance provides the context for choosing methods appropriate to uncertainty and risk [6].

The framework maintains an explicit chain from physical quantity to consequence. If rainfall informs runoff, runoff informs inundation, and inundation informs asset damage, each map has its own assumptions and validation. A missing hydrological or vulnerability model breaks the chain. The workflow may still present a qualitative option appraisal, but must not compute a monetary loss total by substituting an invented coefficient.

4. Scenario comparison without invented probabilities

Let $a \in \mathcal{A}$ be an action and $s \in \mathcal{S}$ an explicitly described scenario. Let $L(a, s)$ denote a consequence measure under that scenario. The value need not be money; incompatible outcomes remain separate unless an authorised and documented valuation combines them. A regret criterion is:

$$R(a, s) = L(a, s) - \min_{b \in \mathcal{A}} L(b, s), \qquad a^* \in \arg\min_a \max_s R(a, s).$$

This definition chooses an action with the smallest worst regret over the stated scenario set. It assigns no probability to those scenarios. It is sensitive to omitted alternatives, scenario construction and disputed consequence estimates. The interactive companion uses a small constructed cost matrix to explain the calculation. Those values are neither observed losses nor estimates for a real community.

Dynamic adaptive policy pathways motivate explicit signposts and possible changes over time [7, 8]. Decision-making-under-deep-uncertainty methods provide a broader family of approaches to inspect, rather than one universally preferred rule [9]. An authority may prefer robust satisficing, multi-objective comparison or an information-gathering action. The analyst records that choice and the reasons for it.

For distributional analysis, report consequences by affected group and location. A favourable total can conceal a severe disadvantage to one community. Equity constraints, minimum service levels and affordability limits must be specified before ranking alternatives. Sensitivity analyses vary weights and uncertain costs and report where the preferred option changes. A polished single rank should not conceal that instability.

5. Bounded agentic assistance

The existing conceptual governance design assigns specialised assistants to evidence retrieval, hazard interpretation, option comparison, finance screening and policy consistency. The coordinating assistant surfaces incompatible assumptions and unresolved claims. It does not decide which public values prevail. This manuscript formalises the evidence and evaluation contracts around that design.

Retrieval-augmented generation offers a mechanism for connecting generated text to retrieved material, but retrieval does not guarantee that the claim is supported [10]. A reviewer must check the exact passage, population, period and qualification. The workflow stores the retrieved passage and the proposed interpretation separately so later readers can see whether the interpretation exceeded the source.

Each assistant has a narrow action budget. Read access to an evidence folder does not imply permission to send a funding application, alter an official policy or contact a community. External documents and messages are evidence, never instructions controlling the workflow. Adversarial-agent benchmarks motivate testing malicious retrieved text and tool misuse explicitly [11]. The pilot blocks consequential external actions and records attempted boundary crossings.

NIST's AI risk-management framework and generative-AI profile provide voluntary governance and assessment references [12, 13]. They do not certify that this framework is safe, compliant or effective. A pilot must identify its actual system boundaries, models, data flows, responsibilities and failure responses. Vendor changes, retrieval changes and altered model behaviour create a new evaluated configuration.

6. Data stewardship and public participation

Public evidence, controlled records and community-supplied knowledge require different access conditions. FAIR principles support useful descriptions and reuse conditions, while not requiring unrestricted release of every dataset [14]. Where Indigenous or other community-governed data are involved, the relevant communities determine participation and governance arrangements; the CARE principles are a reference for that discussion [15]. This paper does not presume consent or authority on their behalf.

Record who was invited, who participated, how accessible alternatives were offered and which objections remain unresolved. A generated summary is returned to participants for correction where appropriate. The workflow must preserve dissenting views rather than turning disagreement into a fabricated consensus. Representation is assessed directly; an online form count does not establish that the affected population was represented.

Personal and sensitive records are minimised, retained for a stated purpose and access-controlled. Public outputs contain only cleared evidence, aggregated findings and agreed disclosures. No raw learner data, household vulnerability records or private government drafts are required for the initial constructed-case evaluation. Any later field study needs its own ethics, privacy and institutional approvals.

7. Evaluation protocol

The first evaluation uses a set of documented decision cases with known source boundaries and deliberately seeded errors. An independent case author records the answerable claims and the information that remains genuinely unknown. Some cases contain irrelevant retrieved passages, conflicting assumptions, stale policy versions or malicious instructions. The evaluation distinguishes a justified refusal to quantify from an incomplete answer to a question that was answerable.

Compare three workflows: conventional analyst preparation; one assistant with retrieval; and the bounded multi-agent workflow. Give each the same evidence set, time allowance and output brief. Randomise case order and counterbalance analyst exposure where feasible. If the same analyst sees all versions of a case, carryover must be measured or the design changed. The number of cases and reviewers follows a feasibility pilot, not an invented power claim.

Primary outcomes are unsupported consequential claims, source-to-claim accuracy and reproducibility of the stated decision calculation. Secondary outcomes include reviewer correction time, omissions,

assumption disagreements found, access violations and a blinded usefulness judgement. Reviewers receive a rubric with concrete error categories. Record both agreement and adjudication, and keep disagreement visible.

A deployment study is a later stage. It would assess actual institutional use, burden, decision changes and distributional consequences over a defined period. Faster drafting in a constructed exercise does not establish better adaptation or avoided loss. The present protocol therefore reports no return-on-investment or policy-effect estimate.

8. Failure conditions and interpretation

The framework fails its initial release gate if it conceals unsupported claims behind valid citations, invents scenario probabilities, omits a material affected group or permits an assistant to exercise public authority. A useful system may still conclude that the available climate evidence cannot support the requested ranking. That is a decision-relevant result when the missing evidence is specified clearly.

The strongest alternative explanation for any apparent multi-agent benefit is extra processing time or a more elaborate checklist. The comparison therefore matches effort and includes a conventional structured template. If the template performs as well, the additional agents may not be justified. Human review must also be costed; a workflow that creates more subtle errors could increase total effort despite producing a draft sooner.

Transfer to another island requires a new institutional and data-fitness assessment. Regional cooperation can share methods, source registers and interoperable evidence records without erasing different laws, resources, languages or priorities. There is no assumption that one governance score or vendor arrangement applies uniformly across the region.

9. Current contribution and next evidence

The current contribution is a decision-oriented specification extending the already designed agentic climate-governance framework. Available prework includes the governance abstract, the D-LEGOS evidence and methods series and existing research articles on climate decisions. The new work links those materials through explicit evidence-fitness, authority and evaluation records. No adoption or measured institutional benefit is asserted.

The next checkpoint is a frozen set of constructed cases, a source-support rubric and a comparison of the three workflows. Only after that checkpoint should the design progress to an authorised institutional pilot. Its thesis contribution is to show how physical-model limits travel into a real decision process, while leaving causal intervention effects and public value choices to the additional evidence and legitimate authority they require.

Data, authorship and assistance

Adrian Dunkley is the sole named author. The climate-research context is Climate Studies Group Mona, The University of the West Indies, Mona; the existing governance design is associated with the Caribbean AI Association. These affiliations identify work contexts and imply no institutional

endorsement. Contact: adrian.dunkley@mymona.uwi.edu. AI assisted drafting, literature organisation and the companion instrument. The author and any participating institution must review claims, relationships, ethics, funding and disclosure before external submission. Source-review depth is recorded separately; references with pending full-text verification are proposed reading, not evidence of a completed policy evaluation.

References

[1] Mycoo et al. / IPCC (2022). Small Islands. Climate Change 2022: Impacts, Adaptation and Vulnerability, Chapter 15. https://www.ipcc.ch/report/ar6/wg2/chapter/chapter-15/.

[2] UNFCCC Least Developed Countries Expert Group (2012). National Adaptation Plans: Technical guidelines for the national adaptation plan process. https://unfccc.int/topics/adaptation-and-resilience/workstreams/national-adaptation-plans-naps/guidelines-for-national-adaptation-plans-naps.

[3] UNDRR (2015). Sendai Framework for Disaster Risk Reduction 2015–2030. https://www.undrr.org/publication/sendai-framework-disaster-risk-reduction-2015-2030.

[4] UNFCCC (2015). Paris Agreement. https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement.

[5] IPCC (2022). Climate Change 2022: Impacts, Adaptation and Vulnerability. Summary for Policymakers. https://www.ipcc.ch/report/ar6/wg2/chapter/summary-for-policymakers/.

[6] IPCC (2022). Decision-Making Options for Managing Risk. Climate Change 2022, Chapter 17. https://www.ipcc.ch/report/ar6/wg2/chapter/chapter-17/.

[7] Haasnoot, Kwakkel, Walker and ter Maat (2013). Dynamic adaptive policy pathways: A method for crafting robust decisions for a deeply uncertain world. https://doi.org/10.1016/j.gloenvcha.2012.12.006.

[8] Kwakkel, Haasnoot and Walker (2015). Developing dynamic adaptive policy pathways: a computer-assisted approach for developing adaptive strategies for a deeply uncertain world. https://doi.org/10.1007/s10584-014-1210-4.

[9] Marchau, Walker, Bloemen and Popper (eds.) (2019). Decision Making under Deep Uncertainty: From Theory to Practice. https://doi.org/10.1007/978-3-030-05252-2.

[10] Lewis et al. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. https://arxiv.org/abs/2005.11401.

[11] Debenedetti et al. (2024). AgentDojo: A Dynamic Environment to Evaluate Prompt Injection Attacks and Defenses for LLM Agents. https://arxiv.org/abs/2406.13352.

[12] NIST (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). https://doi.org/10.6028/NIST.AI.100-1.

[13] Autio et al. / NIST (2024). Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile. https://doi.org/10.6028/NIST.AI.600-1.

[14] Wilkinson et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. https://doi.org/10.1038/sdata.2016.18.

[15] Carroll et al. (2020). The CARE Principles for Indigenous Data Governance. https://doi.org/10.5334/dsj-2020-043.

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.

islands Official chapter passages inspected through indexed retrieval; place-specific adaptation and capacity limits.

nap Official guidance reference; detailed national applicability requires policy-owner verification.

sendai Official UNDRR framework page inspected; adopted 18 March 2015; not an empirical effect estimate.

paris Official instrument reference; no legal compliance conclusion is made.

wg2 Proposed contextual reference; full current reading pending.

decisions Official indexed passage inspected; direct page blocked; exact section verification pending.

dapp Publisher abstract indexed passage inspected; dynamic pathways and signposts.

pathways Publisher abstract inspected; hypothetical case, not Caribbean evidence.

dmdu Publisher open-access book overview inspected; methods reference.

rag Primary record inspected; retrieval architecture, not proof of factual reliability.

agentdojo Primary record inspected; adversarial-agent evaluation motivation.

rmf Official programme inspected; voluntary framework, not a compliance certificate.

gairmf Official abstract inspected; cross-sectoral profile.

fair Official GO FAIR principles inspected; findability does not imply unrestricted access.

care Proposed governance reference; applicability requires community-led interpretation.