PNNL-39717

Knowledge Graph for End-to-End Traceability of an Integrated Human-Earth System Model

August 2026

Isaias Prieto
Hassan Niazi

Science Undergraduate Laboratory Internship (SULI)

Pacific Northwest National Laboratory (PNNL), Joint Global Change Research Institute (JGCRI)

U.S. DEPARTMENT of ENERGY

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from
the Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062
www.osti.gov
ph: (865) 576-8401
fox: (865) 576-5728
email: reports@osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
or (703) 605-6000
email: info@ntis.gov
Online ordering: http://www.ntis.gov

PNNL-39717

Knowledge Graph for End-to-End Traceability of an Integrated Human-Earth System Model

August 2026

Isaias Prieto
Hassan Niazi

Science Undergraduate Laboratory Internship (SULI)

Pacific Northwest National Laboratory (PNNL), Joint Global Change Research Institute (JGCRI)

Prepared for
the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory

Technical Abstract
Integrated human-Earth system models inform energy-water-land system dynamics and policies, yet their results are difficult to trace through input-data, model structure, scenario configurations, and solved outputs. Because this information is siloed across disconnected artifacts, process-based IAMs have historically lacked a unified, queryable representation. Such lack of traceability prevents researchers from systematically isolating the multi-sector drivers of complex outcomes (such as tracing water-scarcity results back to distant energy-system dynamics) or conducting holistic uncertainty attribution across hundreds of interacting parameters. To address this concern, our work documents the software engineering process of a knowledge graph that unifies these four layers for the Global Change Analysis Model (GCAM-USA_Reference scenario, GCAM v9.1). The graph was built as a relational property graph in DuckDB from the run's own artifacts: the input-preparation dependency map (gcamdata chunk map), the model's XML input files, the run configuration, and the results database (BaseX), successfully mapping the model's declared structure. The resulting graph comprises 204,321 nodes and 1,687,814 edges across 16 node types and 15 edge types, with approximately 16.3 million time-series values stored separately to maintain structural efficiency. To ensure representation fidelity, every edge carries an epistemic-status annotation recording the warrant for the relationship (structural, provenance, dependency, or model-derived), and a machine-readable provenance ledger classifying the origin of every schema element. Evaluation against a fixed five-benchmark suite with locked baselines reports zero structural orphans, zero dangling edge endpoints, and 100% of output-producing technologies traceable to raw input files. Two interactive interfaces present the graph, including a serverless browser application built on DuckDB-Wasm. By establishing the first end-to-end provenance framework for an IAM, this work enables researchers and scientists to systematically audit complex policy scenarios, debug model structures, and trace policy-relevant outputs to their data origins in real time.

General Audience Abstract
Governments and researchers rely on complex computer models to explore how energy systems, agriculture, and the economy might evolve under different natural conditions and policies. The Global Change Analysis Model (GCAM) is one of the most widely used of these tools. However, models of this scale are difficult for information traceability. Because raw data, model structures, run configurations, and final results are scattered across multiple files and formats, tracing how a specific input affects an output becomes challenging. Scientists call this problem epistemic opacity, which is the reality that humans can't track the millions of steps a model takes to reach its conclusions. To address this limitation, we developed a knowledge graph that connects all four of these hidden layers into a single, searchable digital network. Comprising approximately 204,000 entities (such as regions, sectors, technologies, and data files) and 1.7 million relationships, this graph turns a fragmented codebase into a cohesive map. With this tool, questions that previously required timely digging through separate databases can now be answered with a single, nearly instant query. To ensure absolute accuracy, every relationship in the graph is labeled with its origin, backed by a rigorous set of benchmark tests to ensure no data is lost or untraceable. Finally, an interactive browser-based view allows users to explore this network. By making massive simulation models traceable from end-to-end, this work takes a vital step toward making the science behind global change science more transparent and accessible.

1. Introduction

1.1 Context

The Global Change Analysis Model (GCAM) is an open-source, global, market-equilibrium integrated human-Earth systems model that represents the linked behavior of energy, water, land, biogeochemcical, and economic systems (Calvin et al., 2019; Niazi et al., 2024; Kyle et al., 2023), with GCAM-USA extending the model with state-level details for the United States (Binsted et al., 2022; Zhao et al., 2026; Bracken et al., 2026). This project leverages a single solved run, the GCAM-USA_Reference scenario of GCAM (Bond-Lamberty et al., 2026; Zhao et al., 2026). A run rests on the *gcamdata* R package that prepares hundreds of raw input datasets into the XML files the model consumes (Bond-Lamberty et al., 2019; Niazi et al., 2025a; Niazi et al., 2025b), and packages such as *rgcam* extract solved results from the run's BaseX output database (Link & Patel, 2026).

1.2 Problem Statement

Complex simulation models are epistemically demanding: the steps by which they produce results are not accessible to unaided human inspection (Humphreys, 2009), and the sensitivity-analysis community has observed of modern mechanistic models that "while models are becoming more and more complex, they are treated more and more like a black-box, even by model developers themselves" (Rezavi et al., 2021, Section 3.2.1). This is a documented concern within the integrated modeling community itself (Duan et el., 2026), and the community has called for systematic evaluation to establish the credibility of large process-based integrated models (Wilson et al., 2021). The community's existing transparency infrastructure only addressed this partially;the flagship scenario databases and explorers serve harmonized model outputs, not model internals (Cointe, 2024; Kim et al., 2026; Niazi et al. 2024). While GCAM's own provenance tooling in *gcamdata* traces input-data preparation but stops at the produced XML files (Bond-Lamberty et al., 2019). No existing tool connects data provenance, structure, scenario configuration, and solved outputs.

1.3 Purpose and Research questions

The purpose of this project was to design and implement a software product that bridges the architectural gap of lack of systematic traceability of all influencing factors and parameterizations leading to an outcome of interest for a single model run. The resulting system is a knowledge graph that unifies the run's data provenance, declared model structure, scenario configuration, and solved outputs within a single connected, queryable store, accompanied by interactive web interfaces for client-side exploration. This framing adopts Pfenninger's (2024) thesis that releasing open code and data is insufficient on its own, meaning that cognitive understandability must be operationalized as an explicit design goal. Within the vocabulary of model interpretability, the graph operates as a structural transparency instrument by making the model's constituent components, parameter lineages, logical dependencies, and declared relations fully inspectable (Lipton, 2018). This work extends Lipton's taxonomy from its original machine-learning context to process-based simulation models. Ultimately, the graph represents the model's declared topology exclusively; it infers no unstated relationships from raw data, assumes no hidden variables, asserts no statistical mechanisms, and makes no empirical claims beyond the internal logic of the model. Three questions guided this work. Can the four layers be unified so that cross-layer traceability questions resolve in single queries? With what measured fidelity does the graph represent the run? What metadata and confidence levels should be attached to each recorded relationship?

2. Background

2.1 Representing declared structure

Every edge in the graph carries an annotation named *causal_status* in the schema, necessitating a clear definition of what this label can and cannot represent. The causal-inference tradition separates two tasks: inferring structure from data (the discovery program of Spirtes et al., 2000) and using a structure already given (Pearl, 2009). The knowledge graph developed in this work performs no statistical inference; rather, it transcribes GCAM's declared dependency structure. When a model's underlying mechanisms are fully known, relations in the model can in principle be read directly from those mechanisms rather than inferred from data (Bareinboim et al., 2022). However, this guarantee attaches strictly to the model's runnable equations, never to the graph topology alone. Accordingly, *causal_status* is interpreted throughout this report strictly as an *epistemic_status* label rather than the source-warrant for each relationship.

2.2 Transparency and opacity

Interpretability fragments into distinct conceptual dimensions, separating structural transparency from post-hoc explanation (Lipton, 2018). Within this taxonomy, the graph serves as a structural transparency instrument. However, philosophy-of-simulation literature establishes firm bounds regarding what such an instrument can accomplish. Publishing a model's underlying wiring does not dissolve the inherent computational opacity of its solved states (Humphreys, 2009). Furthermore, in complex integrated models, declared modularity often overstates effective modularity because individual components become deeply entangled at the solution point (Lenhard & Winsberg, 2010). In addition, human cognitive explanation is inherently contrastive, contextual, social, and selective (Miller, 2019), presenting an explanatory standard that a raw structural query cannot fulfill on its own.

Two constructive theoretical framings emerge from these limitations. First, the graph functions as a digital counterpart to formal documentation standards. While frameworks such as TRACE prescribe disclosing the entire modeling process to public scrutiny (Grimm et al., 2014), this graph operationalizes the disclosure as an interactive, machine-queryable instance. Second, the epistemic-status labels and the accompanying self-audit ledger function as a documented trust scaffold within the framework of computational reliabilism (Duran & Formanek, 2018). In opaque computational environments, this paradigm grounds system trust in documented reliability indicators, verified provenance chains, audit trails, and architectural flags rather than requiring step-by-step human inspection of every mathematical operation.

2.3 Knowledge graphs of models

Property graphs represent established, first-class structures within the knowledge-graph ecosystem, where annotating graph statements with confidence scores, provenance records, structural contexts, and epistemic metadata is standard practice (Hogan et al., 2021). Consequently, the primary scientific novelty of this work lies in applying a rigorous source-warrant taxonomy specifically to a simulation model's declared-structure graph, rather than in introducing the underlying technique of edge annotation. The closest external effort in multi-domain modeling is The World Avatar, a dynamic knowledge-graph framework in which autonomous agents wrap computational models, process queries, execute simulations, and interact across a shared graph network (Akroyd et al., 2021). The distinction between these approaches is that while The World Avatar frames the model as an active agent interacting with the graph, this architecture represents the model's declared topology, internal parameters, execution lineage, and solved numerical outputs directly as the graph's content. In biological modeling, the MIRIAM standard provides a relevant cross-domain precedent by transforming a mechanistic model's internal components into semantically annotated, machine-queryable, fully inspectable, and programmatically accessible objects (Le Novere et al., 2005). Conversely, energy-domain semantic resources such as the Open Energy Ontology operate at a complementary abstraction level, providing shared vocabularies, domain concepts, class taxonomies, and entity definitions across studies rather than generating an instance graph of an individual run's internal operations (Booshehri et al., 2021)

2.4 Research gap

The integrated human-Earth system modeling, multi-sector dynamics, and integrated assessment modeling (IAM) communities traditionally explain the sensitivity of models' outcomes through several uncertainty characterization paradigms. These include perturbation ensembles (Lamontagne et al., 2018; Wessel et al., 2024; Niazi et al., 2024; Kim et al., 2026; Zhao et al., 2026; Kyle et al., 2023), machine-learning attribution across scenario spaces (Duan et al., 2026), deep-learning emulation architectures (Holmes et al., 2026), and statistical surrogate models (Niazi et al., 2021). Furthermore, the community routinely publishes research findings through output-oriented scenario databases (Cointe, 2024), while documenting model topology in static journal articles and online documentation decoupled from any specific execution deck (Calvin et al., 2019; Binsted et al., 2022). Existing software tools within the Global Change Analysis Model (GCAM) ecosystem, including gcamdata, ModelInterface, interactive dashboards, specialized analysis scripts, and the EPA's GLIMPSE platform (U.S. Environmental Protection Agency), query and display outputs. While GLIMPSE additionally configures scenario parameters and initiates runs, none of these tools exposes input provenance, parameter lineages, structural relationships, or solved output topologies as a unified queryable graph. Note that other more detailed models exist in the GCAM modeling systems, such as superwell, GLORY, Tethys (Niazi et al., 2025c; Zhao et al., 2024; Bracken et al., 2026), but they are not scoped into this exercise, which only focuses on core GCAM.

A documented literature and repository search conducted for this project revealed a distinct research gap. No prior knowledge graph, property graph, relational graph, or formal ontology exists for GCAM or for any process-based IAM, as a solved-run instance that links data provenance, declared structure, scenario configuration, and solved outputs within a single database. Existing systems differ fundamentally in kind. Specifically, gcamdata tracks input provenance exclusively (Bond-Lamberty et al., 2019), The World Avatar binds models as external agents rather than inspectable content (Akroyd et al., 2021), output scenario infrastructure serves final results only (Cointe, 2024), and traditional documentation describes model logic without computational connectivity. This absence claim is presented strictly as the outcome of a systemic search rather than a proven universal negative.

3. Methodology

3.1 Research design

The project follows a methodological framework of tool development paired with benchmark evaluation. The objective is to build a high-fidelity representation of a model directly from a model version's own run artifacts, subsequently measuring the structural and content fidelity of that representation against fixed, locked baselines. Two fundamental design principles guided this construction. First, the graph is built as a strictly read-only representation of the model run and avoids any form of data imputation. To maintain representation integrity, any values that cannot be explicitly resolved are loaded as NULL rather than relying on heuristic estimates. Furthermore, every graph key is version-namespaced, using the default prefix version to ensure that the resulting representation is permanently pinned to the specific software release version that was actually graphed.

3.2 Data sources

Five distinct data sources, all generated by or packaged with the solved GCAM-USA_Reference run, feed the knowledge graph. The first is the *gcamdata* package shipped inside the GCAM install. Its chunk map (*GCAM_DATA_MAP.rda*) records how each data-processing chunk produced its outputs from precursors; its *roxygen* man pages and R source files supply per-chunk documentation (description, details, and defining source file), extracted into a chunk-information table. The second is the XML input deck, defining the region, sector, subsector, and technology hierarchy with goods and declared technology inputs. The third is the run configuration (*configuration_usa.xml*). The fourth is the BaseX results database, extracted through *rgcam* (Link & Patel, 2026) across nine query streams covering flows, outputs, production, costs, prices, emissions, and share-weights; the input-flow backbone alone is roughly 3.5 million rows. The fifth is the solver logs, parsed into an over-period equilibrium status table stored alongside the graph.

3.3 Pipeline

A three-stage pipeline, driven by a Python package (*gcamkg*) with Makefile targets, moves data from the GCAM installation to the graph. Stage one extracts each slice into flat, gzip-compressed CSVs using eight R and Python scripts, producing 18 build inputs plus three unloaded metadata-view files. Stage two runs an ordered, idempotent SQL build in DuckDB, constructing raw mirrors and then node, scenario, edge, fact, and provenance tables, 56 tables in total, with per-table row-count checks. Stage three runs the read-only benchmarks, renders visualizations, and exports a versioned Parquet set for the browser interface.

3.4 Graph model

The graph itself is a relational property graph under five schema rules. A topology/value split stores deduplicated structural pairs in edge tables while the full time series (about 16.3 million rows) lives in four fact tables keyed to the same identifiers. Region-sage traversal gives every edge touching a global good its own region column, preventing false cross-region supply-chain paths. An inclusive edge policy keeps pass-through, secondary, and cost flows as first-class edges with metadata flags, differing filtering to query time. An epistemic-status annotation labels every edge with one of four active warrants (structural, provenance, dependency, or model_derived), with a fifth value, hypothesized, reserved and deliberately empty. A materialized lineage closure (39,002 rows) precomputes reachability from produced XML files back to raw source CSVs. Alongside the graph, a self-audit ledger classifies every schema element into one of three origin classes (directly sourced, inferred or generalized, or manually designed) with exact source locations and caution flags; it documents its own compromises, recording for instance that DEFINES edges labeled structural have a warrant closer to provenance. The ledger ships as machine-readable JSON and Parquet files.

3.5 Fidelity measures

Fidelity is measured by benchmarks B1 through B5 with locked baselines re-run after every change. The term benchmark is utilized here in the reference-standard sense established for model evaluation (Oreskes et al., 1994). Expected results are recorded in the benchmark code, a pass requires an exact match, and intentional changes require re-locking with justification. B1 tests input-to-output traceability, B2 relationship correctness, B3 one-query assembly, B4 a comparative diagnostic of technology competition, and B5 completeness.

4. Results

4.1 The built graph

The completed graph holds 204,321 nodes and 1,687,814 edges across 16 node types and 15 edge types, all contained within a single DuckDB database comprising approximately 56 tables. The structural layer contains 105 regions, 17,412 sectors, 55,400 subsectors, 107,998 technologies, 1,518 goods, and 17,874 markers. These counts may shift slightly between builds as data extraction files are deployed, representing the current reference figures for this specific software build. The provenance layer's chunk nodes now carry the parsed gcamdata documentation; the solved-outputs layer is backed by approximately 16.3 million fact rows. Cross-layer questions that previously required separate tools (the gcamdata map lineage, the XML deck for structure, the configuration file for the scenario, rgcam queries for results) resolve in single SQL queries against one store.

4.2 Benchmark results

All five benchmarks successfully pass their locked baselines. Benchmark B1 traces the CO_2 emissions of a Texas industrial coal-cogeneration technology back through its defining XML files to the data pipeline steps and raw input CSVs that produced them. Benchmark B2 confirms that modeled input flows possess matching declared structural inputs, while ensuring that every edge type preserves its contracted endpoints and status labels. Benchmark B3 assembles a technology's declared inputs, flowing inputs, outputs, structural placement up to the region level, and its five defining XML files in a single query. Obtaining this information previously required consulting the XML source deck, the data pipeline map, the configuration files, and the primary results database separately. Benchmark B4 returns, within a single query, the model-declared determinants of electricity-technology competition in China for the year 2050. In this run, solar PV leads generation at 25.1 EJ, exhibiting the lowest declared unit cost among competing options, optimal market access, and a share-weight of 1.0. Conversely, combined-cycle has generated 0.33 EJ at roughly 1.5 times the declared cost of solar with a share-weight of 0.24. This query acts as a comparative diagnostic to surface the model's own internal declarations without asserting an external causal claim. Benchmark B5 reports zero technology-level defining XML files (107,998 of 107,998), and 100% of output-producing technologies (80,227 of 80,227) traceable to raw source data through the lineage closure.

4.3 Epistemic-status annotation, the ledger, interfaces, and products and deliverables

Every edge carries its warrant label; the reserved hypothesized value remains empty, and the ledger classifies every schema element with caution flags against its own choices, a documented trust scaffold rather than a validation claim. Two web interfaces expose the graph. The primary one is a static, serverless React and TypeScript application that loads the Parquet export into DuckDB-Wasm and runs every query client-side, offering search, focused subgraph exploration, upstream and downstream dependency chains, an alternatives view, an edge inspector with status badges, a meta-graph view, and an integrated ledger display. A second view renders the complete graph in raw WebGL within one self-contained HTML files. Deliverables comprise the build repository with its extraction scripts, SQL model, pipeline package, and benchmarks the built database and Parquet export, the ledger, and two interfaces.

5. Discussion, Limitations, and Future Work

The graph closes a specific gap between the community's input-side provenance tooling (Bond-Lamberty et al., 2019) and its output-side scenario infrastructure (Cointe, 2024): the run itself now has a connected, queryable representation. The novelty lies at the integration level rather than the component level. Coarse-grained lineage (Cui & Widom, 2003), property-graph modeling (Hogan et al., 2021), and established extraction tooling are deliberately reused; what is new is persisting the lineage, closing it transitively, and joining it across the execution boundary to declared structure, configuration, and solved outputs, with a source-warrant taxonomy applied to a simulation model's declared-structure graph. Against The World Avatar the distinction is content versus agent (Akroyd et al., 2021). Against sensitivity analysis the relationship is complementary: the graph answers reachability at zero simulation cost and thereby scopes the perturbation experiments that alone establish magnitude (Saltelli et al., 2008; Weiser, 1984; Lamontagne et al., 2018; Kyle et al., 2023; Wessel et al., 2024; Niazi et al., 2024; Kim et al., 2026; Zhao et al., 2026).

The scope conditions are firm. The graph captures one model, one scenario, and one run exactly the isolated operating point the sensitivity analysis warns against generalizing from (Saltelli et al., 2008). It records declared structure without interventional validation, and declared modularity may overstate effective modularity in a model whose components entangle at the solution point. The opacity of the equilibrium solve itself serves publication of the wiring (Humphreys, 2009), and passing every benchmark says nothing about GCAM's correspondence to the real world (Oreskes et al., 1994). Future work follows directly: perturbation validation, testing whether declared paths predict which outputs move when an input changes; mapping the schema onto the W3C PROV data model for interoperability (Moreau & Missier, 2013); extension to multiple runs and scenarios to enable structural comparison; extension to other models (Niazi et al., 2025c; Zhao et al., 2024; Bracken et al., 2026) and sectors (Misener and Niazi, 2025; Zhang et al., 2024; Wolfram et al., 2026) and service as grounding substrate for ensemble and attribution studies (Duan et al., 2026).

6. Conclusion

Complex simulation models risk becoming opaque or computationally inaccessible, even to their own primary developers (Razavi et al., 2021). Within the integrated human-Earth system and multisector dynamics modeling community, researchers have widely documented both the interpretability critique and the partial reach of current transparency infrastructure. Existing platforms successfully serve output scenario data and trace input datasets, but the execution phase bridging inputs to outputs has lacked a connected, queryable representation. This gap highlights a broader methodological challenge: providing open-source code, raw data files, input scripts, and output databases alone does not guarantee system understandability, cognitive accessibility, structural clarity, or meaningful peer scrutiny (Pfenninger, 2024).

To address this architectural deficit, this study engineered a unified graph representation for a complete GCAM-USA v9.1 Reference run. The resulting knowledge graph unifies 204,321 nodes and 1,687,814 edges across four interconnected functional domains: raw data provenance, declared model topology, scenario parameters, and solved economic outputs. This infrastructure incorporates explicit epistemic-status labels on every edge, a machine-readable schema audit ledger, a locked-baseline benchmark suite, and a serverless browser interface for client-side exploration. Validation testing confirms zero structural orphans, zero dangling endpoints, 100% technology coverage, and complete lineage closure, while explicitly tracking unit-verification debt rather than concealing data anomalies.

Crucially, this graph model represents the declared structure of the system while maintaining strict epistemic boundaries. The framework infers no unstated relationships, asserts no external causal claims, avoids statistical extrapolations, and speaks strictly to the internal reality of the model world (Oreskes et al., 1994). Within these bounded scope conditions, this research demonstrates that end-to-end, single-query traceability across an integrated multi-sector dynamics model run is fully achievable, using standard, open-source, reproducible tooling. Ultimately, this work establishes a concrete, operational instrument for model understandability, providing a transparent evidence structure for researchers, scientists, policy analysts, model developers, and external security auditors.

Acknowledgements

I express my sincere gratitude to the Science Undergraduate Laboratory Internship (SULI) program, the U.S. Department of Energy, Pacific Northwest National Laboratory (PNNL), and the Joint Global Change Research Institute (JGCRI)for providing the opportunity to conduct this research at the intersection of integrated human-Earth system modeling, computational provenance, graph databases, and software architecture. I am particularly indebted to my mentor, Hassan Niazi, for his technical guidance, domain expertise, architectural insights, and steadfast support throughout the internship. I also extend my appreciation to my home academic institution, the University of Guam, for supporting my educational trajectory. Finally, I thank the broader PNNL colleagues, the GCAM development team, and the JGCRI modeling community for their valuable discussions, code reviews, domain feedback, and collaborative contributions. H.N. was supported through DOE Office of Science BER's SFA Integrated Multisector Multiscale Modeling (IM3) project.

References

Akroyd, J., Mosbach, S., Bhave, A., & Kraft, M. (2021). Universal digital twin: A dynamic knowledge graph. *Data-Centric Engineering, 2*, Article e14. https://doi.org/10.1017/dce.2021.10

Bareinboim, E., Correa, J. D., Ibeling, D., & Icard, T. (2022). On Pearl's hierarchy and the foundations of causal inference. In H. Geffner, R. Dechter, & J. Y. Halpern (Eds.), *Probabilistic and causal inference: The works of Judea Pearl* (pp. 507-556). Association for Computing Machinery.

Binsted, M., Iyer, G., Patel, P., Graham, N. T., Ou, Y., Khan, Z., Kholod, N., Narayan, K., Hejazi, M., Kim, S., Calvin, K., & Wise, M. (2022). GCAM-USA v5.3_water_dispatch: Integrated modeling of subnational US energy, water, and land systems within a global framework. *Geoscientific Model Development, 15*(6), 2533-2559. https://doi.org/10.5194/gmd-15-2533-2022

Bond-Lamberty, B., Dorheim, K., Cui, R., Horowitz, R., Snyder, A., Calvin, K., Feng, L., Hoesly, R., Horing, J., Kyle, G. P., Link, R., Patel, P., Roney, C., Staniszewski, A., Turner, S., Chen, M., Feijoo, F., Hartin, C., Hejazi, M., … Clarke, L. (2019). gcamdata: An R package for preparation, synthesis, and tracking of input data for the GCAM integrated human-earth systems model. *Journal of Open Research Software, 7*(1), Article 6. https://doi.org/10.5334/jors.232

Bond-Lamberty, B., Patel, P., Lurz, J., Kyle, P., Calvin, K., Smith, S., Snyder, A., Dorheim, K. R., Binsted, M., Link, R., Kim, S., Narayan, K., Graham, N., Staniszewski, A., Lochner, E., Feng, L., Roney, C., Lynch, C., Horing, J., … Iyer, G. (2026). *JGCRI/gcam-core: GCAM 9.1* (Version gcam-v9.1) Computer software. Zenodo. https://doi.org/10.5281/zenodo.20510944

Bracken, C., Niazi, H., Thurber, T., Thompson, I., Tamaddun, K., Eldardiry, H., Mongird, K., Voisin, N., Sun, N., & Rice, J. (2026). *High-resolution monthly sectoral water demands for the U.S. over 1980-2100*. https://doi.org/10.31223/X56B7C

Calvin, K., Patel, P., Clarke, L., Asrar, G., Bond-Lamberty, B., Cui, R. Y., Di Vittorio, A., Dorheim, K., Edmonds, J., Hartin, C., Hejazi, M., Horowitz, R., Iyer, G., Kyle, P., Kim, S., Link, R., McJeon, H., Smith, S. J., Snyder, A., … Wise, M. (2019). GCAM v5.1: Representing the linkages between energy, water, land, climate, and economic systems. *Geoscientific Model Development, 12*(2), 677-698. https://doi.org/10.5194/gmd-12-677-2019

Cointe, B. (2024). The AR6 Scenario Explorer and the history of IPCC Scenarios Databases: Evolutions and challenges for transparency, pluralism and policy-relevance. *npj Climate Action, 3*, Article 3. https://doi.org/10.1038/s44168-023-00075-0

Cui, Y., & Widom, J. (2003). Lineage tracing for general data warehouse transformations. *The VLDB Journal, 12*(1), 41-58. https://doi.org/10.1007/s00778-002-0083-8

Dash, D., & Druzdzel, M. J. (2008). A note on the correctness of the causal ordering algorithm. *Artificial Intelligence, 172*(15), 1800-1808. https://doi.org/10.1016/j.artint.2008.06.005

Duan, H., Sun, Y., Tang, Y., Iyer, G., & Li, X. (2026). Leveraging machine learning to reveal transparency in integrated assessment model ensembles. *Climatic Change, 179*(5), Article 98. https://doi.org/10.1007/s10584-026-04181-w

Freire, J., Koop, D., Santos, E., & Silva, C. T. (2008). Provenance for computational tasks: A survey. *Computing in Science & Engineering, 10*(3), 11-21. https://doi.org/10.1109/MCSE.2008.79

Grimm, V., Augusiak, J., Focks, A., Frank, B. M., Gabsi, F., Johnston, A. S. A., Liu, C., Martin, B. T., Meli, M., Radchuk, V., Thorbek, P., & Railsback, S. F. (2014). Towards better modelling and decision support: Documenting model development, testing, and analysis using TRACE. *Ecological Modelling, 280*, 129-139. https://doi.org/10.1016/j.ecolmodel.2014.01.018

Hogan, A., Blomqvist, E., Cochez, M., d'Amato, C., de Melo, G., Gutiérrez, C., Kirrane, S., Labra Gayo, J. E., Navigli, R., Neumaier, S., Ngonga Ngomo, A.-C., Polleres, A., Rashid, S. M., Rula, A., Schmelzeisen, L., Sequeda, J., Staab, S., & Zimmermann, A. (2021). Knowledge graphs. *ACM Computing Surveys, 54*(4), Article 71. https://doi.org/10.1145/3447772

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association, 81*(396), 945-960. https://doi.org/10.1080/01621459.1986.10478354

Humphreys, P. (2009). The philosophical novelty of computer simulation methods. *Synthese, 169*(3), 615-626. https://doi.org/10.1007/s11229-008-9435-2

Kim, G. J., Wild, T., Zhang, Y., Binsted, M., Niazi, H., Chowdhury, A. K., Zhao, M., & Lamontagne, J. (2026). *Beyond Water: Global Hydropower Futures Shaped by Interacting Hydrological and Socio-Technical Drivers*. Research Square. https://doi.org/10.21203/rs.3.rs-9464688/v1

Kyle, P., Ollenburger, M., Zhang, X., Niazi, H., Durga, S., & Ou, Y. (2023). Assessing Multi-Dimensional Impacts of Achieving Sustainability Goals by Projecting the Sustainable Agriculture Matrix Into the Future. *Earth's Future*, *11*(10), e2022EF003323. https://doi.org/10.1029/2022EF003323

Lamontagne, J. R., Reed, P. M., Link, R., Calvin, K. V., Clarke, L. E., & Edmonds, J. A. (2018). Large ensemble analytic framework for consequence-driven discovery of climate change scenarios. *Earth's Future, 6*(3), 488-504. https://doi.org/10.1002/2017EF000701

Link, R., & Patel, P. (2026). *rgcam: Tools for importing and working with GCAM results* (Version 1.2.1) [Computer software]. GitHub. https://github.com/JGCRI/rgcam

Lipton, Z. C. (2018). The mythos of model interpretability. *Communications of the ACM, 61*(10), 36-43. https://doi.org/10.1145/3233231

Misener, C. P., & Niazi, H. (2025). *Mapping Nitrogen Pathways in Organic Manure through GCAM* (PNNL--38052). Pacific Northwest National Laboratory (PNNL), Richland, WA (United States). https://doi.org/10.2172/2584635

Moreau, L., & Missier, P. (Eds.). (2013, April 30). *PROV-DM: The PROV data model* (W3C Recommendation REC-prov-dm-20130430). World Wide Web Consortium. https://www.w3.org/TR/2013/REC-prov-dm-20130430/

Niazi, H., Smith, S. J., Weber, M. A., Hoesly, R. M., Pimenta, I. J., & Stelmach, T. J. (2025a). *Core Model Proposal #390: Base Year Update: Initial Updates Preparing for the New Base Year* (PNNL--37780). Pacific Northwest National Laboratory (PNNL), Richland, WA (United States). https://doi.org/10.2172/3000478

Niazi, H., Chadsey, R. P., Zhang, Y., Zagoruichyk, A., Patel, P. L., & Kholod, N. (2025b). *Core Model Proposal 401: Ukraine as an independent region in GCAM* (PNNL-37799). Pacific Northwest National Laboratory (PNNL), Richland, WA (United States). https://doi.org/10.2172/3000091

Niazi, H., Ferencz, S. B., Graham, N. T., Yoon, J., Wild, T. B., Hejazi, M., Watson, D. J., & Vernon, C. R. (2025c). Long-term hydro-economic analysis tool for evaluating global groundwater cost and supply: Superwell v1.1. *Geoscientific Model Development*, *18*(5), 1737–1767. https://doi.org/10.5194/gmd-18-1737-2025

Niazi, H., Wild, T. B., Turner, S. W. D., Graham, N. T., Hejazi, M., Msangi, S., Kim, S., Lamontagne, J. R., & Zhao, M. (2024). Global peak water limit of future groundwater withdrawals. *Nature Sustainability*, *7*(4), 413–422. https://doi.org/10.1038/s41893-024-01306-w

Niazi, M. H. K., Morales Nápoles, O., & Van Wesenbeeck, B. K. (2021). Probabilistic Characterization of the Vegetated Hydrodynamic System Using Non-Parametric Bayesian Networks. *Water*, *13*(4), 398. https://doi.org/10.3390/w13040398

Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994). Verification, validation, and confirmation of numerical models in the earth sciences. *Science, 263*(5147), 641-646. https://doi.org/10.1126/science.263.5147.641

Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511803161

Pfenninger, S. (2024). Open code and data are not enough: Understandability as design goal for energy system models. *Progress in Energy, 6*(3), Article 033002. https://doi.org/10.1088/2516-1083/ad371e

Razavi, S., Jakeman, A., Saltelli, A., Prieur, C., Iooss, B., Borgonovo, E., Plischke, E., Lo Piano, S., Iwanaga, T., Becker, W., Tarantola, S., Guillaume, J. H. A., Jakeman, J., Gupta, H., Melillo, N., Rabitti, G., Chabridon, V., Duan, Q., Sun, X., … Maier, H. R. (2021). The future of sensitivity analysis: An essential discipline for systems modeling and policy support. *Environmental Modelling & Software, 137*, Article 104954. https://doi.org/10.1016/j.envsoft.2020.104954

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., & Tarantola, S. (2008). *Global sensitivity analysis: The primer*. John Wiley & Sons. https://doi.org/10.1002/9780470725184

Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, prediction, and search* (2nd ed.). MIT Press. https://doi.org/10.7551/mitpress/1754.001.0001

Weiser, M. (1984). Program slicing. *IEEE Transactions on Software Engineering, SE-10*(4), 352-357. https://doi.org/10.1109/TSE.1984.5010248

Wessel, J., Iyer, G., Wild, T., Ou, Y., McJeon, H., & Lamontagne, J. (2024). Large ensemble exploration of global energy transitions under national emissions pledges. *Earth's Future, 12*(10), e2024EF004754. https://doi.org/10.1029/2024EF004754

Wilson, C., Guivarch, C., Kriegler, E., van Ruijven, B., van Vuuren, D. P., Krey, V., Schwanitz, V. J., & Thompson, E. L. (2021). Evaluating process-based integrated assessment models of climate change mitigation. *Climatic Change, 166*(1-2), Article 3. https://doi.org/10.1007/s10584-021-03099-9

Wolfram, P., Niazi, H., Kyle, P., & McJeon, H. (2026). Ensuring consistency between biogeochemical planetary boundaries. *Nature Sustainability*, *9*(4), 488–493. https://doi.org/10.1038/s41893-026-01770-6

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford University Press. https://doi.org/10.1093/0195155270.001.0001

Zhang, X., Sabo, R., Rosa, L., Niazi, H., Kyle, P., Byun, J. S., Wang, Y., Yan, X., Gu, B., & Davidson, E. A. (2024). Nitrogen management during decarbonization. *Nature Reviews Earth & Environment*, *5*(10), 717–731. https://doi.org/10.1038/s43017-024-00586-2

Zhao, M., Wild, T., Ahsan, H., Patel, P., Snyder, A., Binsted, M., Vernon, C., Khan, Z., Sun, N., Burleyson, C., Niazi, H., Graham, N., Wolfram, P., & Rice, J. (2026). *U.S. water system responses to future global human-Earth system change*. Research Square. https://doi.org/10.21203/rs.3.rs-9182467/v1

Zhao, M., Wild, T. B., Graham, N. T., Kim, S. H., Binsted, M., Chowdhury, A. F. M. K., Msangi, S., Patel, P. L., Vernon, C. R., Niazi, H., Li, H.-Y., & Abeshu, G. W. (2024). GCAM–GLORY v1.0: Representing global reservoir water storage in a multi-sector human–Earth system model. *Geoscientific Model Development*, *17*(14), 5587–5617. https://doi.org/10.5194/gmd-17-5587-2024

Pacific Northwest National Laboratory

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99354

1-888-375-PNNL (7665)

www.pnnl.gov