Dear Reviewers and System Dynamics 2026 Conference Committee,

We sincerely thank the reviewers for their thoughtful and constructive comments. We carefully considered all comments and addressed the majority of them in the revised manuscript. Where page-limit constraints prevented extensive expansion, we made concise revisions to improve clarity, organization, terminology, and readability, and added targeted clarifications where feasible.

Below, we provide a point-by-point response. Reviewer comments are summarized for clarity, followed by our responses and the revisions made where feasible within the manuscript's scope and length constraints.

Comments from Reviewer 1: The manuscript introduces a methodology for constructing causal loop diagrams (CLDs) by adapting Ishikawa diagrams. The limitations of current CLD construction approaches are clearly articulated, and the literature review is comprehensive. The decision framework within the methodology, which selects statistical tests based on the properties of each variable (Section 3.2), is particularly noteworthy.

I have the following minor suggestions.

First, the data description for validation is insufficient. The sources and processing steps for the time-series data in Section 4.8.4 are not systematically documented. Likewise, Section 4.3 lacks details on the data specifications used for causality testing. Please add a dedicated data section or clearly state the data specifications within Section 4.

Response:

> *We have added a dedicated data section (Section 4.1, "Data Sources, Preprocessing, and Testing Dataset") that consolidates the data specifications previously dispersed across the manuscript. It documents the secondary sources for both economic-context and automotive-market variables (DANE, Banco de la República, FEDESARROLLO, ANDI, FENALCO, ANDEMOS, among others), the monthly coverage period (2005–2016), the five selection criteria used to reduce the candidate indicators to the final 18 variables, and the preprocessing steps (temporal alignment, unit consistency, missing-data and comparability checks) applied before causality testing. This section now supplies the data specifications underlying both Section 4.3 (causality testing) and Section 4.8.4 (validation), addressing the documentation gap noted.*

Second, the frequent use of short, consecutive paragraphs makes the argumentation feel fragmented. Some sections also repeat the same points, leading to redundancy. Strengthen the logical connections between paragraphs and reorganize the text into more cohesive structures.

Response

> *We have revised the manuscript to improve cohesion: short consecutive paragraphs were consolidated into larger units organized around a single argument, explicit transitions*

were added to strengthen the logical flow between paragraphs, and redundant passages that restated the same points across sections were removed or merged. The text now reads as a more continuous and connected argument.

Third, abbreviation usage is inconsistent. Some abbreviations revert to full spelling after being defined. There are typographical errors, such as "GDP Shikawa Diagram." Additionally, Section 4.5 refers to "four principal feedback loops," but five are described. Please review the text for consistency and accuracy.

Response

We have corrected the typographical error ("GDP Shikawa Diagram" → "Ishikawa diagram") and reconciled the loop count in Section 4.5, which now consistently reports five feedback loops in agreement with the figures and surrounding text. We also performed a full pass on abbreviation usage, ensuring each term is defined once at first occurrence and used in its abbreviated form consistently thereafter. These corrections have been applied throughout the document.

Comments from Reviewer 2: This paper proposes a five-step methodology that integrates structured hypothesis generation, based on transformed Ishikawa diagrams, with formal temporal causality testing in order to construct empirically validated causal loop structures. The authors claim that the framework exhaustively enumerates plausible causal relationships, tests temporal precedence and predictive relevance using appropriate causality methods, and determines causal polarity to distinguish reinforcing from balancing feedback loops.

Because the classical Ishikawa diagram is inherently static, its use in this context requires not only the collection of time-series data for the variables of interest, but — more importantly — careful treatment of delays and accumulations. These are central features of system dynamics modeling, yet the authors do not adequately address these challenges.

Response

We respectfully note that this comment, while raising important system dynamics principles, may not fully reflect the methodological contributions already present in the paper, and furthermore conflates the role of the Ishikawa diagram in our framework with the framework as a whole.

In our methodology, the transformed Ishikawa diagram serves exclusively as a hypothesis generation tool — its purpose is to exhaustively enumerate candidate causal relationships in a structured, auditable manner. It is explicitly not used for causal validation, delay estimation, or structural formulation. The reviewer is correct that the classical Ishikawa diagram is static; this is precisely why our framework does not rely on it for dynamic inference. The subsequent steps — stationarity testing (Step 2), Granger/Toda-Yamamoto/Transfer Entropy causality testing (Step 3), and polarity determination (Step 4) — are inherently temporal and directly address the dynamic features the reviewer identifies as missing.

Specifically regarding delays: the lag structure in the Granger causality framework (optimal lag selection via AIC/BIC/HQ criteria, Section 3.3) explicitly captures and quantifies the temporal delay between cause and effect. This is one of the methodology's *distinctive contributions: not only does it confirm whether a causal relationship exists, but it identifies at what lag the causal signal is statistically significant — information that directly informs delay parameterization in the stock-and-flow model.*

Regarding accumulations: the paper explicitly addresses stock-flow semantics in Section 4.8.1, where validated causal links involving stock variables are implemented as flows rather than direct influences, preserving System Dynamics accumulation principles. The methodology is designed for data-rich environments, where time-series observations of both stock proxies and auxiliary variables are available — and it is in precisely these environments that Granger-based lag detection and stock initialization from empirical data are most tractable.

We therefore respectfully suggest that the comment, while pointing to genuinely important SD principles, does not identify a gap in the methodology as presented — rather, it identifies concerns that are addressed across Steps 2 through 5 and in the hybrid model formulation. We added a subsection in the paper making this mapping more explicit.

The approach to defining "system variables" — restricted to those supported by available data — conflicts with fundamental principles of system dynamics. Key variables that are conceptually essential from a system dynamics perspective may be unjustifiably excluded simply because suitable data are not readily available.

Moreover, the number of hypothesized directional relationships grows quadratically with the number of “system variables.” For example, with 20 variables, this results in 380 hypothesized relationships. In system dynamics, parameter estimation is typically conducted after a stock-and-flow structure (SFD) has been established. Attempting instead to construct and calibrate a causal loop diagram (CLD) directly from an Ishikawa diagram — without distinguishing stocks from other variables and without accommodating delays — fails to meet the basic requirements of system dynamics modeling.

Response

We appreciate the reviewer's engagement with scalability and methodological rigor, but respectfully disagree that these represent fundamental limitations of the proposed framework.

On quadratic growth of hypothesized relationships: The reviewer correctly observes that the number of candidate relationships grows as n(n-1) with the number of variables. However, this is a property of exhaustive hypothesis generation — which is precisely the methodological virtue being claimed. Traditional expert-driven CLD construction is not more scalable; it is simply less transparent about which hypotheses were never considered. More critically, the causality testing pipeline (Steps 2–3) is highly amenable to automation. Modern computational environments execute Granger, Toda-Yamamoto, and Transfer Entropy tests across hundreds of variable pairs in seconds. The real bottleneck in traditional SD practice is not computation but undocumented hypothesis pruning — a problem our framework directly addresses. Furthermore, as explicitly noted in the Future Research Directions section of the paper, integration with Large Language Models and agentic AI frameworks provides a natural solution for large-scale problems: LLMs can automate the identification and classification of stocks, flows, and auxiliary variables from domain descriptions or data dictionaries, effectively serving as an intelligent pre-screening

layer that reduces the hypothesis space before statistical testing begins. The quadratic growth concern is therefore a tractable engineering challenge, not a methodological flaw.
On the claim that the methodology fails basic SD requirements: We respectfully but firmly disagree. The reviewer appears to conflate the scope of the methodology with its completeness as an SD workflow. The paper is explicit that the methodology produces a validated causal loop diagram — not a complete stock-and-flow diagram. The CLD is the upstream artifact from which the SFD is subsequently derived, consistent with standard SD practice (Sterman, 2000). The methodology intervenes specifically at the CLD construction stage, which is precisely where empirical grounding is most lacking in traditional practice.
Regarding the distinction between stocks and auxiliary variables: as clarified in the paper (and elaborated in the methodological note now added to Section 3), the Granger and Transfer Entropy frameworks operate on time-series observations irrespective of whether a variable is a stock or auxiliary. The stock/flow/auxiliary classification is then applied at the SFD construction stage, where validated causal links involving accumulation variables are correctly implemented as flows. The methodology does not ignore this distinction — it defers it to the appropriate stage of the modeling workflow, where structural and semantic judgment, rather than statistical testing, governs the decision.
The claim that the approach "fails to meet basic SD requirements" therefore mischaracterizes the methodology's scope. A framework that empirically validates causal structure, quantifies temporal delays, and provides complete traceability from data to CLD — in a domain where data are rich enough to support such analysis — advances SD practice rather than contradicting it.

These structural shortcomings lead to flawed parameter estimation and misleading conclusions. For instance, the finding that "economic growth did not Granger-cause population changes at policy-relevant time scales" likely arises because the model neglects key mechanisms, such as the delay between fertility intentions and childbirth, the gradual effects of improvements in healthcare on life expectancy, and the fundamental fact that population is a stock variable.

Healthcare, life expectancy, and fertility intentions are outside the system boundary by design: for a system where quantitative macroeconomic variables are dominant, these are out-of-scope constructs, not neglected mechanisms. The finding that growth does not Granger-cause population at policy-relevant horizons is the validation step working as designed — it shows the coupling acts over longer horizons, consistent with R3 as a slow, long-run loop.

In summary, the paper inadvertently demonstrates that a causal loop diagram is a purely qualitative tool: it describes the architecture of a system in terms of feedback loops, but does not possess a mathematical structure suitable for quantitative estimation. Any empirical computation and validation in system dynamics must be grounded in a stock-and-flow framework.

We respectfully disagree. The main contribution of the paper is precisely a method for validating causal loops in data-rich environments, where formal temporal causality testing can be used to construct empirically grounded causal loop structures.

Only a few minor typographical errors were identified, including “methodoly,” “validated relationships. This substantial reduction,” and “GDP Shikawa Diagram.”

Response

We thank the reviewer for the careful reading. The typographical errors identified have been corrected throughout the manuscript: We also performed an additional proofreading pass to catch any remaining errors.

Comments from Reviewer 4: This study aims to develop a method for generating causal loop diagrams and empirically validating them. This work has the potential to fill an important gap in the SD modeling literature. I applaud the authors for this work and using a case study to demonstrate the approach and its effectiveness.

To strengthen the work, I suggest adding a couple of diverse cases, including one in which the data are rich (as in the case presented) but there is little theory about the relationships, and another with data available for some links but not for all.

We thank the reviewer. As suggested, the "Current Research" subsection (Section 5.1) of the conclusions now discusses two contrasting cases: a data-rich but theory-poor setting (an OEM producing PCBs with GPU components), where the framework also discovers new links and quantifies their delays; and a case with data for only some links (consumer preferences for new-product adoption), where the method applies only when preferences can be proxied by survey or panel data and otherwise requires adaptation. These cases make the methodology's boundary conditions explicit.

In Figure 6, the initial value of the simulated hybrid model is different from the actual data. Is it possible to adjust the initial stock values to address this? The parameters estimated using the regressions are appropriate, as they are estimated separately (i.e., partial calibration; Homer, 2006).

We agree that the regression-based parameters are appropriately estimated separately, consistent with the partial-calibration approach (Homer, 2006). Regarding the initial condition, we have adjusted the initial stock values so that the model's starting point aligns with the observed data, and Figure 6 has been updated accordingly. The revised figure now shows the simulated trajectory initialized at the actual initial value, removing the earlier discrepancy.

Could you elaborate on whether the method you suggested works for both stocks and auxiliary variables? Can they be treated the same?

We thank the reviewer; we have added a clarification in Section 3.1 ("A Note on Stocks and Flows"). In short: yes for testing, with a distinction at construction. The Granger causality and Transfer Entropy tests operate on time-series observations regardless of whether a variable is a stock or an auxiliary, so both are treated the same in the causality-

testing pipeline. They differ at the model-construction stage (Step 5): a validated link into a stock must respect accumulation semantics and be implemented as a flow, whereas auxiliary links translate directly into algebraic relationships.

The comments above have been made directly by the reviewers and have not been edited by the Program Committee. Reviewers are generally happy to receive feedback on their reviews and you are welcome to send them an anonymous email with any feedback you would like to provide. To send the email simply specify the reviewer number and fill in the body of the email you would like to send.