Causal Structure Learning for Wildfire Risk in the Maule Region, Chile

Efraín Campusano
Facultad de Ingeniería
Universidad Andres Bello
Santiago, Chile
e.campusanomelendez@uandresbello.edu

Paulo Quinsacara
Facultad de Ingeniería
Universidad Andres Bello
Santiago, Chile
p.quinsacarajofr@uandresbello.edu

Pablo Schwarzenberg
Facultad de Ingeniería
Universidad Andres Bello
Santiago, Chile
pablo.schwarzenberg@unab.cl

Hernán Astudillo
Facultad de Ingeniería
Universidad Andres Bello
Viña del Mar, Chile
hernan.astudillo@unab.cl

Carla Taramasco
Facultad de Ingeniería
Universidad Andres Bello
Viña del Mar, Chile
carla.taramasco@unab.cl

***Abstract*—Wildfires in central Chile have intensified in frequency and impact, yet many operational analyses still rely mainly on associative predictors, which can limit interpretation and decision support. This work-in-progress study applies causal structure learning to wildfire event records from the Maule Region, Chile, with the objective of identifying plausible dependency pathways among meteorological, topographic, spatial-context, and burned-area variables. We use the official CONAF incident dataset, preprocess continuous, categorical, circular, and skewed variables, and learn a causal graphical structure using Greedy Equivalence Search (GES) with a BIC score under a light domain-informed partial ordering. The resulting CPDAG is interpreted as a hypothesis-generating representation of conditional dependence constraints rather than as a uniquely identified causal mechanism. The learned graph shows a modular organization, separating a terrain–meteorology component from a burned-area accounting component. This separation is informative because it distinguishes contextual and hazard-related covariates from compositional impact variables whose dependencies are partly driven by definitional aggregation. A permutation-based falsification analysis indicates that the observed graph is more compatible with the data than graphs obtained under the tested permutation nulls. The results suggest that causal graph learning can provide an interpretable framework for wildfire risk analysis in Maule, while also highlighting the need for future work involving larger permutation schemes, sensitivity analyses across scores and constraints, and validation of intervention-oriented decision scenarios.**

***Index Terms*—Wildfire, Wildfire risk forecasting, Causal discovery, Causal structure learning**

I. Introduction

The Maule Region in central Chile is recurrently affected by wildfires due to the coupled influence of (i) a Mediterranean climate with warm dry summers that promotes seasonal fuel desiccation, (ii) spatially continuous fuels associated with productive forestry and mixed land covers, and (iii) strong anthropogenic pressure related to rural activities and the wildland urban interface. In addition, the 2010 to 2015 "megadrought" intensified hydroclimatic deficits and vegetation water stress in central Chile, increasing fuel flammability and preconditioning the landscape for severe fire seasons [1].

These interacting drivers were evident during the extreme 2017 fires in Mediterranean Chile, where antecedent drought, fire-weather extremes, and land-use patterns jointly contributed to unusually large burned areas and substantial impacts [2]. Evidence from Maule-specific spatio-temporal analyses further supports the presence of strong spatial structure in fire occurrence and affected area, motivating models that explicitly address confounding by geography, seasonality, and exposure [3].

Accordingly, this study is framed as a causal analysis rather than a purely predictive exercise. Our objective is to identify plausible causal pathways linking meteorology, topography, fuel composition, and suppression/response timing to wildfire severity (e.g., total burned area), and to distinguish robust causal signals from spurious associations induced by latent spatial and temporal structure.

The remainder of this article is structured as follows: Section II surveys related work on causal discovery, graphical models, and wildfire risk analysis; Section III presents causal graphical models, Greedy Equivalence Search (GES), and permutation-based falsification; Section IV describes the dataset, preprocessing, causal structure learning setup, learned graph, and falsification results; Section V discusses the learned causal structure in the wildfire risk setting, including graph modularity, spatial and administrative dependencies, intervention-oriented interpretation, reporting bias, and limitations for future validation; and Section VI summarizes and concludes.

II. Related work

Score-based causal discovery offers a principled route to estimate (up to Markov equivalence) directed acyclic causal structures from observational data, which is attractive for disentangling coupled biophysical and anthropogenic drivers of wildfire occurrence and impacts [4], [5]. Greedy Equivalence Search (GES) traverses equivalence-class space using locally score-decomposable operators and, under standard assumptions, provides large-sample guarantees for recovering the Markov equivalence class; scalable variants such as FGES extend this paradigm to higher-dimensional settings through optimized scoring and caching [4], [6].

In Earth system sciences, causal graph inference has been increasingly advocated to move beyond association toward mechanistic hypotheses and counterfactual reasoning in complex, confounded systems [5]. Within the wildfire literature, causal and probabilistic graphical modeling—often operationalized as (causal) Bayesian networks—has been proposed to encode causal knowledge for risk assessment and scenario reasoning, while complementary work learns Bayesian-network structure and parameters from historical fire records and geo-environmental covariates to analyze interactive relations among ignition causes and conditioning factors, including BN–GIS integration and data-driven wildfire behavior modeling [7]–[10].

For south-central Chile, recent analyses attribute exceptional fire seasons to compound extreme weather and climate influences interacting with human ignitions [11], and urban-interface case studies (e.g., the Concepción metropolitan area) highlight strong anthropic biophysical coupling in mapped wildfire risk [12]. Spatio-temporal analyses in Maule further emphasize land cover and human accessibility as correlates of fire occurrence and burned area [13].

In summary, these studies motivate applying GES/FGES-style causal structure learning to operational wildfire datasets in Chile to better separate robust driver effects from spatial and seasonal confounding [4], [6].

III. Background

A. Causal models

Causal inference distinguishes statistical association from the effect of an intervention. Two complementary frameworks are commonly used. The potential outcomes framework defines, for each unit, counterfactual outcomes under alternative treatments and formalizes estimands such as average treatment effects, together with identification conditions such as consistency, exchangeability, and positivity [14]–[16]. The structural causal model framework represents the data generating process through a directed acyclic graph (DAG) and structural equations of the form $X_i \leftarrow f_i(\mathrm{Pa}(X_i), U_i)$, where $\mathrm{Pa}(X_i)$ are the parents of X_i and U_i are exogenous disturbances [17]–[19]. Interventions are defined operationally using the $\mathrm{do}(\cdot)$ operator by replacing the structural equation of the intervened variable, yielding an interventional distribution that generally differs from observational conditioning [17]. Within DAG semantics, d-separation links graph structure to conditional independences, and graphical identification criteria such as back door adjustment formalize how to control for confounding and estimate causal effects from observational data under explicit assumptions [17], [18]. Importantly, without interventions or additional functional assumptions, observational data typically identify only a Markov equivalence class of DAGs, not a unique causal graph, which motivates careful interpretation and the use of domain constraints when available [17], [19].

B. Causal structure discovery algorithms

Causal structure discovery aims to infer a causal graph, or more generally its Markov equivalence class, from observational data. Most approaches rely on the Causal Markov condition and a faithfulness type assumption that connects population conditional independences to graphical separation properties [17], [19]. Under causal sufficiency, discovery can return a completed partially directed acyclic graph (CPDAG) that represents the set of DAGs compatible with the observed independences. Methods differ primarily in how they use the data. In this work we adopt a score based approach and use Greedy Equivalence Search (GES) because it directly optimizes a global penalized likelihood score over equivalence classes, which can be advantageous in finite samples where conditional independence testing may be unstable [4]. The resulting output is interpreted as a hypothesis space of plausible causal structures to support subsequent effect estimation and robustness checks, rather than as a uniquely identified causal mechanism.

GES (Greedy Equivalence Search) is a score based causal discovery algorithm that searches over Markov equivalence classes of DAGs, typically represented as CPDAGs [4]. GES assumes a score that is score equivalent, meaning all DAGs in the same equivalence class obtain the same score, and decomposable, meaning the total score can be written as a sum of local node scores determined by each node and its parent set [4]. A common choice is the Bayesian information criterion (BIC), which combines the log likelihood with a complexity penalty and is consistent for model selection under standard regularity conditions [20].

GES has two greedy phases:

- **Forward**: it starts from the empty graph and repeatedly applies the single admissible edge insertion that yields the largest improvement in the score while maintaining acyclicity at the equivalence class level.
- **Backward**: it starts from the forward solution, and repeatedly applies the single admissible edge deletion that most improves the score, removing superfluous edges and reducing overfitting [4].

Under causal sufficiency, faithfulness, and a consistent score, GES is consistent for recovering the true Markov equivalence class in the large sample limit [4]. In applied settings, the learned CPDAG is used to identify which directions are supported by the data and the assumptions, which directions remain ambiguous within the equivalence class, and which covariates are plausible confounders or mediators for downstream causal effect estimation.

C. Falsification testing based on permutations

Because causal discovery from observational data can produce edges driven by finite sample artifacts, strong seasonality, or residual confounding, it is good practice to include falsification tests that quantify how often the learned structure could arise under a null where the relevant dependence is destroyed. Permutation testing provides a nonparametric way to construct such null distributions by reassigning values across units while preserving the marginal distribution of each permuted variable [21], [22]. In our context, a permutation falsification protocol can be defined as follows: (i) fit GES to the original dataset and record target summaries such as the total score, the number of edges, and the presence of specific edges or orientations; (ii) generate B permuted datasets by permuting one or more variables across events, or by permuting rows, depending on which dependence mechanism is being falsified; (iii) refit GES to each permuted dataset using the same score and hyperparameters; and (iv) compare the original summaries against their permutation distributions to obtain empirical p values or tail probabilities. This approach supports edge level and graph level validation, for example by testing whether an observed edge appears substantially more often than under permutation, or whether the achieved score improvement over the empty graph exceeds what is typical under the null. Permutation based testing has also been studied directly in Bayesian network structure learning as an alternative to parametric tests for dependence assessment, providing further precedent for resampling based validation of learned graphical structures [23]. When data exhibit temporal or seasonal dependence, permutations should be constrained to respect the dependence structure being treated as nuisance, for example by permuting within time strata, so that the falsification target remains meaningful for the scientific question.

IV. Experiments

A. Dataset and preprocessing

We use an event level wildfire dataset for the Maule Region obtained from the official CONAF cause and origin reporting system [24]. this dataset contains 6,610 wildfire events and 22 variables. Time information is not used as a causal feature. It is used only to report the observational window of the original CONAF export, from 2011-09-29 to 2021-06-21. All date and time fields and all geographic coordinates were removed before causal structure learning.

The dataset includes four variable blocks. First, burned-area variables (hectares) by fuel and land cover: pine_child, pine_adult, pine_senior, eucalyptus, others_plants, wooded, scrub, pastureland, agricultural, and waste. The file also provides three aggregated subtotals, subtotal_plantations, subtotal_natural_vegetation, and subtotal_other_surfaces, plus total_area_ha. Second, meteorological variables: temperature, humidity_percent, wind_speed, and wind_direction (degrees). Third, terrain and physiography: earring (slope percent), exposition (aspect degrees), and topography (categorical class). Fourth, an administrative unit code, cut_reg_prov_com, is used to account for spatial heterogeneity without using geographic coordinates.

Preprocessing is designed to support score-based causal discovery with a BIC score. Burned-area variables are strongly right-skewed, so we apply the transformation $\log(1 + x)$ to total_area_ha and to each burned-area component. To handle circular variables, wind_direction and exposition are represented using sine and cosine mappings, that is, $(\sin\theta, \cos\theta)$. The administrative unit code cut_reg_prov_com and the topography class are treated as categorical variables and encoded using one-hot indicators. Finally, all continuous inputs are standardized to zero mean and unit variance. The cleaned file contains no missing values, so no imputation is required.

B. Causal structure learning and falsification setup

We learn causal structure with Greedy Equivalence Search (GES) using a BIC objective [4], [20]. Let X denote the matrix of preprocessed covariates from incendios_dataset_limpio_k6_v3.csv and let Y denote the severity response derived from total_area_ha. GES is applied to the joint variable set $\{X, Y\}$ and returns a completed partially directed acyclic graph (CPDAG) that represents the Markov equivalence class supported by the data under the modeling assumptions [4]. The BIC score is computed from local likelihood contributions, so each candidate move is evaluated through a local score difference, enabling efficient greedy search in equivalence class space [4], [20].

To keep the search largely data driven while respecting domain knowledge, we impose a light partial ordering on admissible edge directions. Spatial and terrain variables, including cut_reg_prov_com, topography, earring, and exposition, are treated as exogenous. Meteorological variables are allowed to depend on spatial and terrain context, but not on burned-area outcomes. Burned-area variables, including total_area_ha and land-cover components, are treated as downstream outcomes. Candidate graph moves are therefore restricted to directions consistent with this ordering, while adjacencies within admissible blocks remain determined by the BIC score.

We quantify whether the discovered structure could be explained by chance alignments rather than systematic dependence using falsification based on permutations [21], [22]. For B replicates, we generate null datasets that preserve marginal distributions while breaking targeted dependencies, rerun GES with identical settings, and compare summary statistics to the observed graph. We use two null schemes that do not use time information. The first permutes the outcome Y across events, which destroys all covariate severity links while keeping the joint distribution among covariates intact. The second permutes Y within cut_reg_prov_com strata, which preserves local severity distributions and therefore controls for spatial heterogeneity while eliminating within stratum covariate outcome relations. For each run we record graph level quantities, such as BIC improvement relative to the empty graph and the number of adjacencies, and edge level quantities, such as the selection frequency of specific adjacencies across replicates. Empirical p values are computed from the permutation distribution using the standard add one correction. This procedure follows general permutation testing principles [21], [22] and is aligned with permutation based validation ideas in Bayesian network structure learning [23].

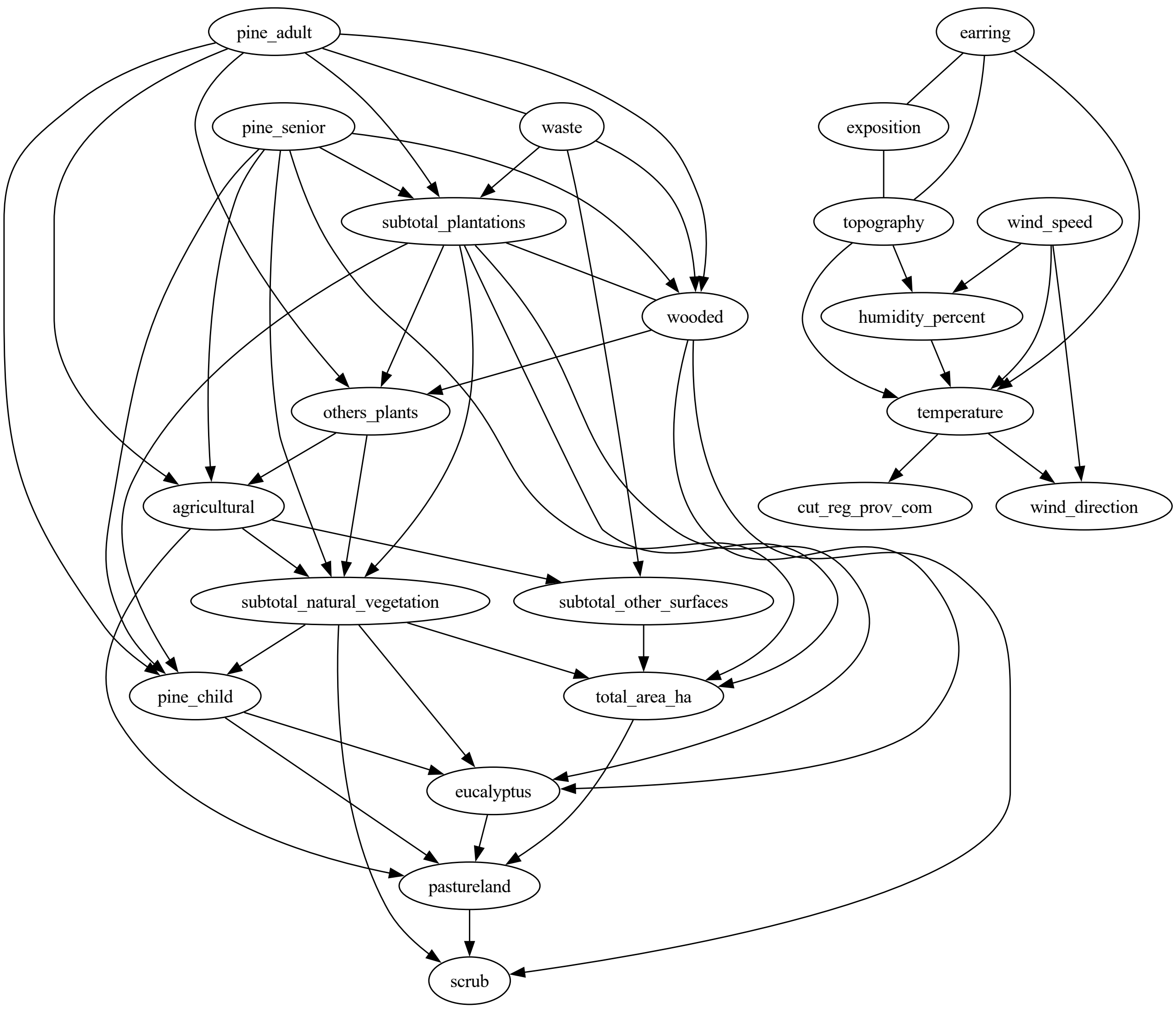

Fig. 1: Causal graph generated by the GES algorithm.

C. Application of algorithms for structure causal discovery

Fig. 1 shows the directed acyclic graph learned with the GES algorithm from the CONAF Maule wildfire dataset. The resulting structure provides a compact, data-driven representation of conditional dependencies among event-level drivers and fire-impact variables. The graph separates into two main modules: a terrain and meteorology block that links slope and physiographic descriptors (earring, exposition, topography) to atmospheric conditions (humidity_percent, temperature, wind_speed, wind_direction), and an impact and fuel-composition block that organizes burned-area components and their aggregates (subtotals and total_area_ha). This modular organization is useful for interpretation because it isolates the internal dependence structure among contextual covariates from the compositional structure of the impact measurements. In particular, the impact block reflects deterministic accounting relationships between land-cover components and total burned area, while the terrain and meteorology block highlights coherent associations among site conditions and weather descriptors. Overall, the learned graph offers an interpretable summary of plausible pathways supported by the observed conditional dependence patterns and can guide subsequent effect estimation under explicit assumptions.

D. Assessing DAG Validity through Falsification Testing

We evaluate the plausibility of the learned causal structure using a permutation-based falsification strategy that checks

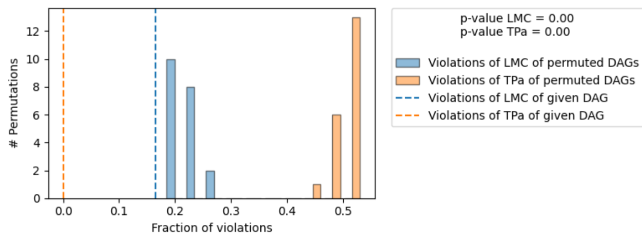

Fig. 2: Permutation-based falsification results for the GES-learned graph. The histograms show the distribution of violation rates under permuted datasets for two sets of falsifiable implications (LMC and TPa). Vertical dashed lines indicate the violation rates of the graph learned from the original data.

whether the graph implies conditional independences that are compatible with the observed data. The key diagnostic is the set of Local Markov Condition (LMC) implications: for each variable, the graph asserts that it is conditionally independent of its non-descendants given its parent set. We operationalize this by running a collection of conditional independence tests implied by the graph and summarizing the fraction of rejected implications as an LMC violation rate. To contextualize this rate, we construct an empirical null distribution via permutations that preserve marginal distributions while breaking targeted dependencies, re-estimate the structure under each permuted replicate, and compare the observed violation rate to the resulting null distribution.

Fig. 2 summarizes the permutation-based falsification test for the GES-learned structure on the CONAF Maule wildfire dataset. Under the adopted permutation null, none of the 20 permuted runs produced a graph in the Markov equivalence class of the observed graph, yielding an empirical p-value reported as 0.00 after rounding, that is, below the resolution allowed by the number of permutation replicates. In addition, the observed graph violates $79/1629$ local Markov conditions (LMCs) and achieves a lower LMC violation rate than all permuted graphs, corresponding again to a rounded p-value of 0.00. Consistently, for the TPa criterion, the observed graph shows essentially zero violations, whereas the permuted graphs exhibit substantially larger violation rates. At the 0.05 significance level, these results indicate that the discovered structure is informative relative to the permutation null and is not rejected by the falsification procedure.

Because the present study is framed as a work-in-progress contribution, the number of permutation replicates was kept limited. Therefore, the reported empirical p-values should be interpreted as descriptive falsification evidence under the tested nulls rather than as definitive inferential guarantees. A larger permutation scheme would improve the resolution of the empirical tail probabilities and is identified as a priority for future validation.

V. Wildfire Risk Interpretation of Causal Structure

A. *Interpretation of the GES estimated causal graph*

Fig. 1 displays the structure estimated by Greedy Equivalence Search (GES) as a CPDAG, which represents the Markov equivalence class supported by the data under the BIC scoring assumptions [4], [20]. Consequently, (i) directed edges should be interpreted as orientations that are identifiable given the learned equivalence class and model assumptions, and (ii) the overall topology is best read as a structured summary of conditional dependence constraints rather than a uniquely identified physical mechanism. A salient feature of the estimated graph is its modular organization into two disconnected components. The first component groups terrain and meteorological descriptors, capturing coherent dependence patterns among slope and physiography (earring, exposition, topography) and atmospheric variables (humidity_percent, temperature, wind_speed, wind_direction), together with an administrative proxy (cut_reg_prov_com). The second component concentrates burned-area measurements (land-cover components, subtotals, and total_area_ha). This separation is useful for wildfire risk interpretation because it isolates context and hazard-related covariates from the internal accounting structure of impact variables, reducing the risk of over-interpreting definitional relationships as causal pathways.

Within the terrain and meteorology component, the learned edges are consistent with the notion that physiography constrains local atmospheric conditions and wind exposure at the event scale. However, directions involving cut_reg_prov_com must be interpreted conservatively, because administrative identifiers typically act as proxies for latent spatial drivers (human pressure, accessibility, land management, and fuel continuity) and can enter the graph through residual confounding rather than direct physical causation. Overall, the estimated structure provides an interpretable dependency map that can be used to guide subsequent effect estimation, for example by identifying plausible adjustment sets and by highlighting where orientation uncertainty remains due to Markov equivalence [4].

B. *Temporal and operational structure in incident surveillance variables*

Time and operational response variables are intentionally excluded from causal structure learning. Calendar time is used only to define the observation window of the dataset, not as a causal feature. This design choice follows standard causal reasoning: operational milestones and suppression-related timings are downstream of ignition and early spread, and they are likely influenced by the same upstream factors that affect severity. Including them as nodes can open post-treatment pathways or induce collider bias when estimating effects of weather, terrain, or context on burned area. For wildfire risk analysis, this implies that the present graph should be interpreted as describing relationships among event context, hazard descriptors, and impact accounting, not as a model of suppression performance. A natural extension is to perform stratified discovery or falsification by season, because south-central Chile exhibits strong seasonal clustering of large fires, and extreme fire seasons are associated with compound meteorological anomalies and high ignition pressure [11]. In that setting, time acts as a regime index rather than a causal driver, enabling regime-specific structures without introducing temporal nodes that complicate causal interpretation.

C. Fire progression and severe outcomes

The burned-area component of Fig. 1 is dense and highly directional because the dataset contains compositional impact variables. In particular, subtotals and total_area_ha are algebraic aggregates of land-cover components, which creates near-deterministic dependencies. As a result, many edges inside this block primarily reflect definitional constraints and redundancy among measurements rather than distinct causal mechanisms. From a wildfire risk standpoint, the informative interpretation is not that one burned-area component causes another, but that the dataset encodes severity as a decomposable impact profile that partitions total damage across plantations, natural vegetation, and other surfaces. This is consistent with Maule-scale evidence that burned area and fire occurrence exhibit strong structure across cover types and concentrate in specific zones associated with human activity and accessibility [13]. For causal attribution of severity to upstream drivers, the recommended interpretation is that total_area_ha should be treated as the primary outcome and the compositional components should be excluded or collapsed into a minimal representation; otherwise, score-based learning will preferentially explain severity through deterministic bookkeeping rather than through weather and terrain influences.

D. Administrative and spatial dependencies

The node cut_reg_prov_com summarizes spatial heterogeneity without explicit coordinates and is therefore best interpreted as a proxy for latent, spatially structured determinants of wildfire occurrence and impacts. Prior analyses in Maule show that many fires concentrate in the coastal zone and central valley and are strongly associated with road networks and proximity to populated areas, highlighting the role of human accessibility and land use patterns in shaping fire regimes [13]. In the estimated graph, dependencies between cut_reg_prov_com and meteorological variables should therefore be read as a spatial mixture effect: different administrative units tend to occupy different microclimatic and land-cover contexts, and the graph captures these systematic co-variations. For decision support, this suggests that administrative strata can function as effect modifiers and confounders. Although conditioning on administrative context may stabilize estimation of meteorological associations with severity by reducing spatial confounding, it may also mask mechanisms if the administrative code absorbs unmeasured drivers. This motivates reporting both pooled and within-stratum analyses, consistent with regional wildfire risk modeling studies that emphasize the joint role of anthropic and biophysical factors in central Chile [12].

E. Implications for intervention-oriented wildfire risk analysis

From an intervention perspective, not all variables in the learned graph are equally actionable. Terrain descriptors such as slope, aspect, and topography are not directly controllable, but they can be used for stratification and prioritization of prevention resources. Administrative context, represented by cut_reg_prov_com, is also not a physical intervention variable, but it can proxy spatial differences in accessibility, land management, exposure, and reporting practices. In contrast, variables related to fuel management, human accessibility, preparedness, and response allocation are more plausible candidates for policy intervention, even if they are only indirectly represented in the present dataset.

For future causal effect estimation, meteorological variables such as wind_speed, wind_direction, temperature, and humidity_percent should be treated as hazard drivers whose effects on total_area_ha may be confounded by spatial context, seasonality, terrain, and fuel conditions. A plausible adjustment strategy would therefore condition on administrative strata, topography, slope, aspect, and fuel or land-cover proxies, while avoiding adjustment for downstream burned-area components that are part of the severity outcome definition. This interpretation positions the learned CPDAG as a guide for defining adjustment sets and intervention hypotheses, rather than as a final policy model.

F. Cause determination and potential reporting bias

CONAF maintains a structured cause and origin determination system, organized into broad cause groups and a detailed taxonomy of general and specific causes [24]. Although cause variables are not part of the cleaned feature set used for structure learning here, the investigation and reporting process still matters for interpretation because it can introduce systematic measurement and selection effects. For example, the probability that an event is detected, documented with complete covariates, and assigned an investigated cause can depend on severity, accessibility, and proximity to infrastructure. Such severity-dependent ascertainment can induce apparent dependencies in observational datasets even when biophysical mechanisms are unchanged. Therefore, edges that link administrative context and hazard descriptors should be interpreted as conditional dependencies in the recorded dataset, potentially influenced by heterogeneous reporting practices across communes and by completeness thresholds for small events. In practice, this motivates robustness checks that restrict analysis to subsets with higher data completeness or that incorporate missingness indicators as potential selection proxies, while keeping the causal interpretation anchored on upstream variables.

G. Global assessment supported by falsification

The permutation-based falsification results in Fig. 2 provide a global check that the learned structure is not a trivial artifact of random alignment under the tested nulls. The observed graph exhibits a lower fraction of LMC violations than the distribution obtained from permuted datasets: the learned structure is near 0.17 violations, while permuted runs concentrate at larger violation rates, indicating that the conditional independence implications of the estimated graph are more compatible with the data than those arising under permutation. For the TPa diagnostic, the observed violation rate is near zero, whereas permuted runs show substantially larger values, further separating the learned structure from the null distribution. With $B = 20$ permutations, p-values reported as 0.00 should be interpreted as being below the permutation resolution (at most $1/(B+1)$), which is sufficient to reject the tested null at conventional significance levels. Importantly, falsification does not prove causal correctness, but it strengthens the claim that the estimated CPDAG captures reproducible conditional dependence constraints present in the dataset and therefore serves as a defensible hypothesis generator for wildfire risk interpretation and follow-on causal effect estimation [21]–[23].

H. Limitations and future validation needs

Several limitations should be considered when interpreting these results. First, the falsification analysis uses a limited number of permutation replicates, which restricts the resolution of the empirical p-values. Second, the learned structure may be sensitive to the chosen score, preprocessing decisions, and partial-order constraints. Third, the dataset is event-level and observational, so the inferred CPDAG should be interpreted as a hypothesis-generating structure rather than as proof of causal effects. Fourth, the inclusion of multiple burned-area components and their aggregates creates compositional dependencies that can dominate score-based structure learning. Finally, the present study does not yet evaluate whether the learned graph improves operational prediction, intervention planning, or policy prioritization relative to standard associative models. These issues define the next validation stage, including larger permutation schemes, sensitivity analysis across alternative scores and constraints, reduced severity representations, and explicit comparison against predictive or decision-support baselines.

VI. Conclusions

This work presented a causal, graph-based analysis of wildfire events in the Maule Region using the official CONAF incident records and Greedy Equivalence Search (GES) as the causal structure discovery method. The estimated CPDAG provides a compact representation of the conditional dependence constraints supported by the data under a BIC scoring criterion [4], [20]. A central empirical outcome is the modular organization of the learned structure, which separates a terrain and meteorology component from a burned-area accounting component. The terrain and meteorology block captures coherent dependencies among physiographic descriptors (earring, exposition, topography) and atmospheric variables (humidity_percent, temperature, wind_speed, wind_direction), together with an administrative proxy (cut_reg_prov_com) that encodes spatial heterogeneity. In contrast, the burned-area block is dominated by compositional relationships among land-cover components, subtotals, and total_area_ha, highlighting that deterministic accounting constraints and redundancy among impact variables can strongly shape score-based discovery. This distinction is critical for wildfire risk interpretation, because edges within the impact block primarily reflect how severity is measured and decomposed across land-cover classes rather than independent physical drivers.

The permutation-based falsification analysis provides quantitative support that the learned graph is informative relative to a permutation null. None of the 20 permutation replicates yielded a graph belonging to the Markov equivalence class of the observed structure, resulting in an empirical p-value reported as 0.00 (rounded, i.e., below the permutation resolution). Moreover, the observed graph violates $50/303$ tested local Markov condition implications and performs better than 100.0% of the permuted graphs, again producing a p-value reported as 0.00. At the 0.05 significance level, and because the graph is informative under the falsification criteria, we do not reject the learned structure. Although falsification does not prove causal correctness, it provides evidence that the estimated CPDAG encodes conditional independence constraints that are markedly more compatible with the observed data than those arising under permutation, supporting its use as a defensible hypothesis generator for causal interpretation and subsequent effect estimation [21]–[23].

From an applied perspective, the results emphasize that spatial heterogeneity and administrative context can act as proxies for latent anthropogenic and landscape drivers and should be treated carefully as potential confounders and effect modifiers when attributing severity to meteorology and terrain. In addition, the findings highlight a methodological consideration for operational wildfire datasets: including multiple burned-area components and their aggregates can dominate score-based discovery due to compositional and partially deterministic relationships, so driver-focused causal interpretation should prioritize upstream covariates while treating total burned area as the primary severity endpoint.

Future work will extend this study in several directions directly aligned with the limitations identified above. First, we will increase the number of permutation replicates to obtain finer empirical p-value resolution and stronger falsification evidence. Second, we will evaluate sensitivity to alternative scores, preprocessing schemes, and partial-order constraints, as well as compare GES with additional causal discovery algorithms. Third, we will test reduced outcome representations that treat total_area_ha as the main severity endpoint, minimizing the influence of deterministic burned-area accounting relationships. Fourth, we will connect the learned graph to intervention-oriented analyses by identifying adjustment sets for estimating the effects of meteorological drivers such as wind speed, temperature, and humidity on burned area. Finally, we will expand the framework to additional wildfire-risk zones across Chile to assess whether the inferred structures generalize across climatic, topographic, and land-use contexts.

References

[1] R. D. Garreaud, C. Alvarez-Garreton, J. Barichivich, J. P. Boisier, D. Christie, M. Galleguillos, C. LeQuesne, J. McPhee, and M. Zambrano-Bigiarini, "The 2010–2015 megadrought in central chile: impacts on regional hydroclimate and vegetation," *Hydrology and Earth System Sciences*, vol. 21, pp. 6307–6327, 2017.

[2] D. M. J. S. Bowman, G. J. Williamson, J. T. Abatzoglou *et al.*, "Human-environmental drivers and impacts of the globally extreme 2017 chilean fires," *Ambio*, vol. 48, no. 4, pp. 350–362, 2019.

[3] I. Díaz-Hormazábal and M. E. González, "Análisis espacio-temporal de incendios forestales en la región del maule, chile," *Bosque (Valdivia)*, vol. 37, no. 1, pp. 147–158, 2016.

[4] D. M. Chickering, "Optimal structure identification with greedy search," *Journal of Machine Learning Research*, vol. 3, pp. 507–554, 2002. [Online]. Available: https://www.jmlr.org/papers/v3/chickering02b.html

[5] J. Runge, S. Bathiany, E. Bollt, G. Camps-Valls, D. Coumou, E. Deyle, C. Glymour, M. Kretschmer, M. D. Mahecha, J. Muñoz-Marí, E. H. van Nes, J. Peters, R. Quax, M. Reichstein, M. Scheffer, B. Schölkopf, P. Spirtes, G. Sugihara, J. Sun, K. Zhang, and J. Zscheischler, "Inferring causation from time series in earth system sciences," *Nature Communications*, vol. 10, p. 2553, 2019.

[6] J. Ramsey, M. Glymour, R. Sanchez-Romero, and C. Glymour, "A million variables and more: the fast greedy equivalence search algorithm for learning high-dimensional graphical causal models, with an application to functional magnetic resonance images," *International Journal of Data Science and Analytics*, vol. 3, no. 2, pp. 121–129, 2017.

[7] J. F. Carriger, M. Thompson, and M. G. Barron, "Causal bayesian networks in assessments of wildfire risks: Opportunities for ecological risk assessment and management," *Integrated Environmental Assessment and Management*, vol. 17, no. 6, pp. 1168–1178, 2021.

[8] V. Sevinç, Ö. Küçük, and M. Goltas, "A bayesian network model for prediction and analysis of possible forest fire causes," *Forest Ecology and Management*, vol. 457, p. 117723, 2020.

[9] Z. Konurhan, M. Yucesan, and M. Gul, "Investigating forest fire causes through an integrated bayesian network and geographic information system approach," *Natural Hazards*, vol. 121, pp. 12 933–12 958, 2025.

[10] K. Zwirglmaier, P. Papakosta, and D. Straub, "Learning a bayesian network model for predicting wildfire behavior," in *Safety, Reliability, Risk and Life-Cycle Performance of Structures and Infrastructures (Proceedings of ICOSSAR 2013)*, 2013. [Online]. Available: https://mediatum.ub.tum.de/doc/1191872/document.pdf

[11] T. Carrasco-Escaff, R. Garreaud, D. Bozkurt, M. Jacques-Coper, and A. Pauchard, "The key role of extreme weather and climate change in the occurrence of exceptional fire seasons in south-central chile," *Weather and Climate Extremes*, p. 100716, 2024.

[12] E. Jaque Castillo, A. Fernández, R. Fuentes Robles, and C. G. Ojeda, "Data-based wildfire risk model for mediterranean ecosystems – case study of the concepción metropolitan area in central chile," *Natural Hazards and Earth System Sciences*, vol. 21, pp. 3663–3683, 2021.

[13] I. Díaz-Hormazábal and M. E. González, "Análisis espacio-temporal de incendios forestales en la región del maule, chile," *Bosque*, vol. 37, no. 1, pp. 147–158, 2016.

[14] J. Splawa-Neyman, D. M. Dabrowska, and T. P. Speed, "On the application of probability theory to agricultural experiments. essay on principles. section 9," *Statistical Science*, vol. 5, no. 4, pp. 465–472, 1990.

[15] D. B. Rubin, "Estimating causal effects of treatments in randomized and nonrandomized studies," *Journal of Educational Psychology*, vol. 66, no. 5, pp. 688–701, 1974.

[16] M. A. Hernán and J. M. Robins, *Causal Inference: What If*. Chapman & Hall/CRC, 2024. [Online]. Available: https://miguelhernan.org/whatifbook

[17] J. Pearl, *Causality*. Cambridge university press, 2009.

[18] J. Peters, D. Janzing, and B. Schölkopf, *Elements of Causal Inference: Foundations and Learning Algorithms*. MIT Press, 2017.

[19] P. Spirtes, C. Glymour, and R. Scheines, *Causation, Prediction, and Search*, 2nd ed. MIT Press, 2000.

[20] G. Schwarz, "Estimating the dimension of a model," *Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.

[21] P. I. Good, *Permutation, Parametric, and Bootstrap Tests of Hypotheses*, 3rd ed. New York: Springer, 2005.

[22] D. Freedman and D. Lane, "A nonstochastic interpretation of reported significance levels," *Journal of Business and Economic Statistics*, vol. 1, no. 4, pp. 292–298, 1983.

[23] M. Scutari and A. Brogini, "Bayesian network structure learning with permutation tests," *Communications in Statistics - Theory and Methods*, vol. 41, no. 16-17, pp. 3233–3243, 2012.

[24] CONAF, "Determinación de causa y origen de incendios forestales," Institutional web resource, 2026, accessed: 2026-01-31. [Online]. Available: https://www.conaf.cl/incendios/prevencion-y-mitigacion/determinacion-de-causa-y-origen-de-incendios-forestales/