AN AGENTIC FRAMEWORK FOR CAUSAL DISCOVERY AND FORECASTING IN OIL AND GAS TIME SERIES

Ahmed Esmin[1,2], **Alexandre Mello Ferreira** [1] **& Anderson Rocha** [1]
[1]Artificial Intelligence Lab., Recod.ai
Instituto de Computação (IC)
Universidade Estadual de Campinas (UNICAMP)
Av. Albert Einstein, 1251, Cidade Universitária, 13083-852, Campinas, SP, Brazil.
{aesmin,amferrei,arrocha}@unicamp.br

[2]Departamento de Ciência da Computação (DCC)
Universidade Federal de Lavras (UFLA)
Av. Central, 3037, Centro, 37200-900, Lavras, MG, Brazil.

ABSTRACT

Industrial time-series analysis requires both accurate forecasting and actionable explanations, particularly in production systems where interventions propagate with delays across interconnected assets. We present an ongoing applied project that develops an agentic framework integrating time-series causal discovery, temporal dependency analysis, and foundation-model forecasting into a unified workflow for operational diagnosis and decision support. Engineers interact through a conversational interface to (i) localize operational events and regime changes, (ii) estimate causal links and time lags among injector and producer variables, and (iii) generate short- and medium-horizon forecasts using a time-series foundation model (TimesFM). We demonstrate the approach on realistic simulated oil and gas production datasets generated with Unisim, capturing injector–producer interference patterns. Early observations indicate that combining causal structure with foundation-model forecasting improves interpretability and supports faster investigation than forecasting-only baselines, while maintaining competitive predictive performance. We summarize practical lessons from integrating causal tools and forecasting into an engineer-facing workflow and outline next steps toward multi-scenario causal discovery and interactive deployment.

Track: Industry & Applications

1 INTRODUCTION AND MOTIVATION

Oil and gas production operations rely on analyzing multivariate time series to relate operational changes (e.g., injection adjustments) to delayed responses in producers (e.g., oil-rate decline or salinity increase). In practice, engineers manually inspect dashboards to correlate events across wells and variables. This process is time-consuming, difficult to scale across large fields, fragmented across tools, and dependent on tacit expertise.

Several analytical techniques address individual aspects of this problem. Change-point detection methods enable the identification of regime shifts in time-series data Truong et al. (2020). Causal discovery algorithms estimate directional relationships and time delays between variables Runge et al. (2019); Castro et al. (2023) and can also support feature representation and selection for predictive models Souza et al. (2026). More recently, foundation models have demonstrated strong forecasting performance across diverse time-series datasets Das et al. (2024); Nie et al. (2024). In parallel, research on LLM-based agents has shown that language models can orchestrate complex analytical pipelines by dynamically invoking specialized tools Yao et al. (2023); Schick et al. (2023); Wang et al. (2024).

However, these capabilities are rarely integrated within a single interactive workflow for industrial analytics. In operational environments such as reservoir management, engineers must manually combine event detection, causal reasoning, and forecasting across multiple systems.

To address this gap, we develop an agentic framework, as part of an ongoing applied project, that integrates causal discovery, change-point detection, and foundation-model forecasting into a unified workflow for interactive time-series analysis. The system introduces an intent-driven tool-routing mechanism that enables natural-language interaction with analytical pipelines. A prototype implementation of realistic oil and gas production scenarios demonstrates improved interpretability and reduced analysis complexity compared to conventional workflows, based on preliminary evaluations.

2 System Overview

The proposed framework operates as an LLM-orchestrated assistant that routes engineer queries to specialized analysis tools for causal discovery, change-point detection, visualization, and forecasting. Natural-language queries are interpreted by an open-source LLM (e.g., GPT-OSS) deployed locally via Ollama, which invokes domain-specific analytical modules and aggregates their outputs into interpretable graphs, aligned plots, and structured summaries.

Figure 1 illustrates the system architecture. A conversational interface forwards user queries to the LLM orchestrator, which interprets the request, retrieves contextual information, and dynamically invokes analytical tools. The outputs of these tools are then combined into structured explanations and visual representations to support engineering analysis.

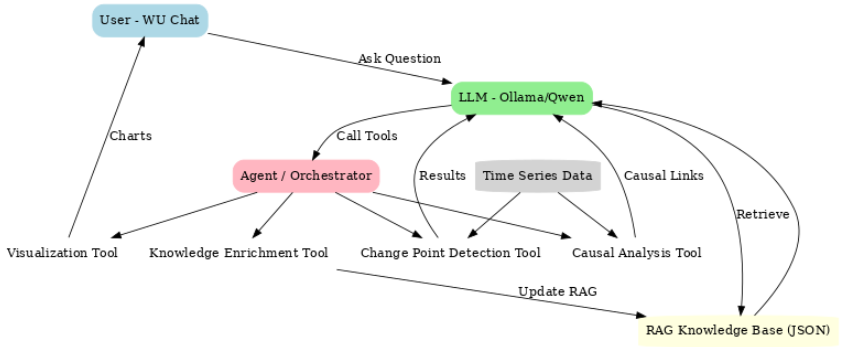

Figure 1: Agentic framework architecture. The LLM routes user queries to analysis modules and aggregates results into explanations and plots.

Interaction workflow. The system follows a structured multi-step process that maps natural-language queries to coordinated analytical operations:

1. **Query interpretation:** the LLM parses the user request and identifies the analysis intent (e.g., event detection, causal analysis, forecasting).
2. **Context selection:** relevant variables and time windows are selected based on the query.
3. **Tool routing:** analytical modules are selected dynamically:
   - change-point detection for event localization,
   - causal discovery (e.g., the PCMCI algorithm—Peter and Clark Momentary Conditional Independence—or ensemble-based methods, including the Aleph model proposed by Castro et al. (2023)) for identifying directional relationships and associated time lags,
   - forecasting models (TimesFM) for prediction.
4. **Execution and aggregation:** outputs from each module are combined into aligned visualizations and structured summaries.

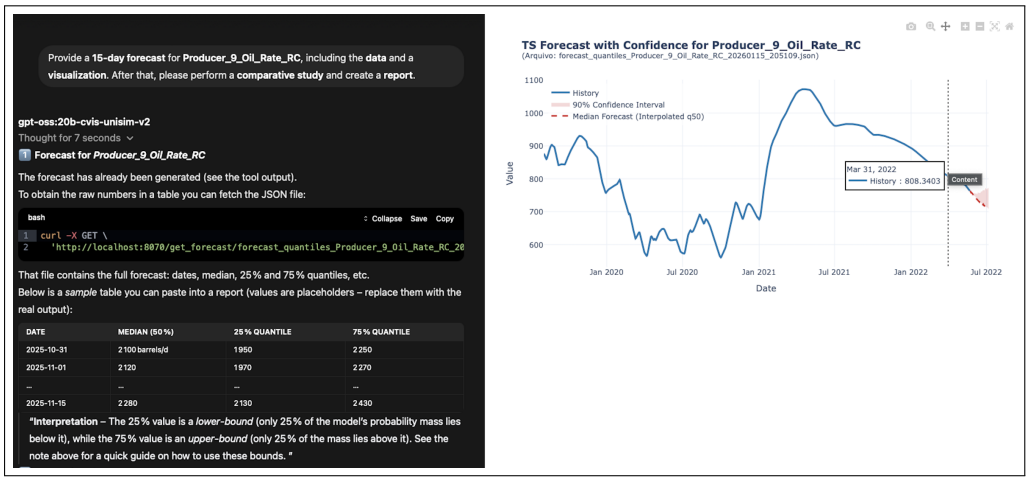

Figure 2: Prototype execution showing forecast trajectories with uncertainty bands for producer oil-rate signals.

The interaction workflow is formalized in Algorithm 1, which summarizes the mapping from user queries to tool selection and result aggregation.

Algorithm 1 Agentic Time-Series Analysis Workflow

1. Receive user query q
2. Infer intent $I \leftarrow \text{LLM}(q)$
3. Select variables and time window $D \leftarrow \text{Context}(q)$
4. Route tools based on I:
   - event $\rightarrow$ change-point detection
   - causal $\rightarrow$ causal discovery
   - forecast $\rightarrow$ TimesFM
5. Execute selected tools on D
6. Aggregate outputs into plots, graphs, and summaries

Together, these components enable interactive exploration of operational events and forecasting scenarios, as illustrated by the prototype output in Figure 2.

3 Case Study and Prototype Outputs

We evaluate the proposed framework on realistic simulated oil and gas production time series generated with **Unisim**[1], representing injector–producer interactions with delayed interference effects. This setup enables controlled operational scenarios (e.g., injection changes) while preserving key industrial characteristics such as noise, temporal lags, and multi-variable coupling.

Given a natural-language query (e.g., assessing the impact of an increase in injection for a specific injector), the system executes an intent-driven analytical pipeline. First, relevant time series are aligned and filtered based on the query context. Next, change-point detection is applied to localize operational events and define analysis windows Truong et al. (2020). Causal discovery is then performed using both feature-importance ensemble methods Castro et al. (2023) and constraint-based approaches such as PCMCI Runge et al. (2019), producing candidate causal links and associated time lags.

[1] https://www.unisim.cepetro.unicamp.br/benchmarks/en/

The inferred causal structure is used to identify a subset of injector variables acting as causal drivers, enabling the selection of a reduced set of predictor variables for forecasting. By focusing on these causally relevant inputs, the approach reduces the dimensionality of the feature space while preserving the most influential relationships. Short- and medium-horizon forecasts are then generated using the time-series foundation model TimesFM Das et al. (2024). Results are returned as aligned plots, causal graphs, and structured summaries to support engineering interpretation.

Figure 2 illustrates a prototype execution, which presents historical producer oil-rate signals along with forecast trajectories and their associated uncertainty bands. The figure highlights how predicted responses evolve under varying operational conditions (e.g., injection changes), allowing engineers to relate temporal patterns, forecast uncertainty, and potential causal drivers within a unified view.

In preliminary evaluations, causal filtering reduced input dimensionality and improved forecast stability, resulting in a lower MAE compared to models using all available variables. Previous studies reported a sensitivity of 0.74 and a specificity of 0.90 for the recovery of causal relationships in similar scenarios Castro et al. (2023). Future evaluation will include scale-independent forecasting metrics, such as MASE, and causal graph accuracy measures, such as the structural Hamming distance (SHD).

In addition to qualitative analysis, preliminary experiments indicate that causal filtering reduces the number of input variables, mitigates spurious correlations, and improves forecast robustness. By focusing on causally relevant drivers, the approach limits the influence of indirect or non-causal associations that can degrade predictive performance. In selected scenarios, this resulted in lower MAE and more stable prediction intervals compared to models using all available variables. These observations suggest that incorporating causal structure can improve both interpretability and predictive consistency.

4 Lessons Learned and Future Work

Lessons learned from prototype usage. The evaluation of the prototype in exploratory analysis scenarios revealed several key insights:

- **Causal filtering reduces investigation time:** Restricting forecasting inputs based on inferred causal relationships significantly reduced the number of variables engineers needed to inspect, streamlining the exploratory analysis process.
- **Interpretability drives adoption:** Engineers consistently relied on ranked causal connections, lag estimates, and aligned time-series visualizations to interpret operational changes, often prioritizing these explanations over raw predictive performance metrics.
- **Interactive workflows improve engagement:** The conversational interface enabled iterative refinement of queries (e.g., adjusting time windows and variable selections), accelerating hypothesis testing and increasing user engagement during analysis.

Practical considerations. The deployment of the proposed framework introduces challenges in integrating heterogeneous components—such as change-point detection, causal discovery, and forecasting—which require coordinated preprocessing, time alignment, and parameter consistency. While the agentic design simplifies orchestration, it depends on robust data pipelines and must balance automation with interpretability to support trust in operational environments. The framework is also sensitive to data quality, as noise, missing values, and nonstationary regimes can affect performance.

Limitations and future work. The current implementation is validated on simulated data, which does not fully capture real-world complexities such as unobserved confounders and operational variability. Future work will extend the framework to multi-scenario causal discovery, improve robustness to nonstationarity and confounding, incorporate stronger evaluation metrics (e.g., MASE and SHD), and validate the approach on real-world production data within decision-support systems.

Acknowledgments

This work is part of an ongoing R&D project with Shell Brasil Petróleo Ltda. (ANP No. 21373-6), funded under the Brazilian ANP R&D levy.

REFERENCES

M. Castro et al. Time series causal relationships discovery through feature importance and ensemble models. *Scientific Reports*, 13(1):11402, 2023. doi: 10.1038/s41598-023-37929-w.

Abhimanyu Das, Weihao Kong, Rajat Sen, and Yichen Zhou. A decoder-only foundation model for time-series forecasting. In *Proceedings of the 41st International Conference on Machine Learning*, ICML'24. JMLR.org, 2024.

Yuqi Nie, Nam H. Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. Time-llm: Time series forecasting by reprogramming large language models. *ICLR*, 2024.

Jakob Runge, Peer Nowack, Marlene Kretschmer, Seth Flaxman, and Dino Sejdinovic. Detecting and quantifying causal associations in large nonlinear time series datasets. *Science Advances*, 5 (11):eaau4996, 2019.

Timo Schick, Jane Dwivedi-Yu, Roberto Dessi, Roberta Raileanu, Maria Lomeli, Eric Hambro, Luke Zettlemoyer, Nicola Cancedda, and Thomas Scialom. Toolformer: Language models can teach themselves to use tools. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

B. Souza, M. Castro, A. Esmin, L. Machado, A. Ferreira, and A. Rocha. Causality-driven feature representation for connectivity prediction. *Frontiers in Artificial Intelligence*, 8:1686750, 2026.

Charles Truong, Laurent Oudre, and Nicolas Vayatis. Selective review of offline change point detection methods. *Signal Processing*, 167:107299, 2020. doi: 10.1016/j.sigpro.2019.107299.

Lei Wang, Chen Ma, Weizhi Zhang, et al. A survey on large language model based autonomous agents. *ACM Computing Surveys*, 2024.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. React: Synergizing reasoning and acting in language models. In *International Conference on Learning Representations*, 2023.