

Ontologically Constrained Ensemble Learning Architecture for Complex System Dynamics Predictive Analysis

Evgenii Sergeevich Mityakov

Dr. Sci. (Econ.), Professor, HOD
KB 9, Subject Oriented Information Systems,
Cybersecurity and Digital Technologies Institute,
MIREA – Russian Technological University,
78 Vernadsky Ave., Moscow, 119454, Russia
E mail: mityakov@mirea.ru.
Scopus Author ID 55960540500,
RSCI SPIN code 5691 8947,
[https://orcid.org/0000 0001 6579 0988](https://orcid.org/0000_0001_6579_0988).

Andrey Ivanovich Ladynin

Dr. Sci. (Econ.), Professor
Informatics Department, Cybersecurity
and Digital Technologies Institute, MIREA
Russian Technological University,
78 Vernadsky Ave., Moscow, 119454, Russia
E mail: ladynin@mirea.ru.
Scopus Author ID 57202347343,
RSCI SPIN code 8762 8534,
[https://orcid.org/0000 0001 7659 2581](https://orcid.org/0000_0001_7659_2581).

Natalia Nikolaevna Karpukhina

Dr. Sci. (Econ.), Associate Professor,
HOD, Innovation Management,
Institute of Management Technologies, MIREA
Russian Technological University,
78 Vernadsky Ave., Moscow, 119454, Russia
E mail: karpukhina@mirea.ru.
Scopus Author ID 57190409353,
RSCI SPIN code 7442 9985,
[https://orcid.org/0000 0003 3378 5230](https://orcid.org/0000_0003_3378_5230).

Abstract

Predicting dynamic behavior in complex organizational-technical systems such as industrial ecosystems, energy grids, and transportation networks faces substantial challenges. These include pronounced nonlinearity, inherent uncertainty, and acute data scarcity. Traditional machine learning methods deliver high predictive accuracy but remain largely uninterpretable, whereas ontological models provide semantic clarity yet suffer from static representations. A hybrid synergy proves necessary, one that effectively bridges inductive learning with deductive logical inference. This study proposes a hybrid approach for predictive modeling and scenario analysis, wherein an ontological model functions as an active semantic constraint for ensemble methods resilient to small-sample settings. Our architecture incorporates a three-tier semantic filter: an ontological layer formalizing system structure, axioms, and causal rules; a predictive layer that leverages an ensemble comprising Bayesian Ridge regression, Gaussian Process Regression, and Elastic Net and a scenario layer designed to generate alternative trajectories. In addition, a consistency filter that validates all outputs against ontological restrictions proposed. Using leave-one-out cross-validation we performed experimental validation on annual data in between 2012–2024 drawn from three Russian high-technology sectors: pharmaceuticals, electronics, and automotive. The hybrid model retains the pure ensemble's accuracy with average MAPE ~9.9% and vastly outperforms lagged ARX regression on nonlinear datasets. Under stress-test scenarios, the semantic filter eliminates roughly 20% logically contradictory projections, effectively tightening confidence intervals. Forecasts extending to 2030 demonstrate coherent trajectory differentiation. Sensitivity analyses concerning temporal lag further substantiate our methodology's robustness. Our framework aims towards interpretability and logical consistency in forecasts without compromising accuracy, an indispensable property for industrial, energy, and transportation decision support systems. These findings establish a foundational basis for developing intelligent hybrid simulation platforms.

Keywords: Hybrid Modeling; Ontological Modeling; Semantic Filter; Ensemble Machine Learning; Time-Series Forecasting; Complex Organizational-Technical Systems; Industrial Ecosystems; Small Samples; Bayesian Ridge Regression; Gaussian Processes; Elastic Net; Scenario Analysis; Inertia; Causal Relationships.

Acknowledgements: The article has been prepared with the support of the Russian Science Foundation (project № 23-78-10009-II).

Introduction

Modeling complex organizational-technical systems, such as industrial ecosystems, energy grids, transport networks, urban agglomerations, and beyond remains a pressing challenge for contemporary applied informatics. Such systems exhibit nonlinear dynamics, multiple feedback loops, delay effects, heterogeneous elements, and profound uncertainty (Duan et al., 2022; Burström et al., 2024). Moreover, such systems have a strong impact on economic development and national innovative ecosystem (Shpak, 2026). Traditional simulation methods like system dynamics, discrete-event simulation, and agent-based approaches demonstrate proven effectiveness, yet they exhibit fundamental limitations: they require complete and homogeneous statistical data, adapt poorly to changing conditions, and cannot effectively integrate expert knowledge with machine learning within a unified predictive architecture (Demartini et al., 2021; Xue et al., 2025).

Recent years have witnessed active developments in machine learning for forecasting complex system behavior. Ensemble and regularization-based methods achieve high predictive accuracy even with incomplete and noisy data (Lv et al., 2022; Galicia et al., 2019; Gao et al., 2022). Nevertheless, non-interpretable models substantially constrain their deployment in decision-support systems. Decision-makers cannot interpret the generated forecasts, verify their logical consistency, or comprehend why the model produced a particular outcome (Mohammed & Kora, 2023; Mienye & Sun, 2022). This limitation proves especially critical for industrial policy, where recommending demonstrably unrealistic measures remains unacceptable (Irwin, 2020).

Concurrently, ontological modeling has matured, offering formal machine-interpretable representations for domain structures, causal relationships, and permissible system states (Benjamin et al., 2006; Beverley & Tolk, 2025; Lei et al., 2020). Ontologies ensure semantic clarity, interoperability, knowledge reuse, and automated modeling processes (Beck et al., 2010; Al-Mohusain et al., 2026). In digital twins and co-simulation environments, ontologies help maintain linkages among assets, processes, data sources, and simulation logic (May et al., 2022; Bjørnskov & Jradi, 2023; Jones et al., 2025). However, ontological models remain inherently static and cannot learn from data, which restricts their adaptivity (Courtot et al., 2011; Livet et al., 2010).

This context creates an objective necessity for hybrid approaches, ones that combine ontological modeling advantages including interpretability, logical consistency, semantic structuring with machine learning strengths like adaptivity, data-driven learning, high forecast accuracy. As Ghidalia et al. (2024) note, hybrid artificial intelligence aims to bridge inductive learning with deductive logical inference rather than employing them separately. Similar conclusions emerge from Von Rueden et al. (2020), Procopio et al. (2023), and Wang et al. (2022): neither purely knowledge-driven nor purely data-driven methods prove sufficient when systems combine known causal structures with hidden dependencies in large datasets.

Within modeling and simulation, hybrid approaches constitute a family of integration patterns where each component assumes a distinct role. Ontologies structure meaning and constrain the search space; machine learning extracts patterns from data; simulation encodes system dynamics or causal process logic (Beverley & Tolk, 2025; Schweidtmann et al., 2023; Rajulapati et al., 2022). Several studies employ ontologies to organize variables and subsequently guide feature selection for machine learning, thereby reducing feature space and improving predictive accuracy (De Coste et al., 2023). In systems biology, specific ontologies describe model components, simulation algorithms, and dynamic behavior, thereby creating an algorithmic semantic layer (Courtot et al., 2011).

Despite active hybrid approach development, several unresolved problems persist. Integrating ontologies and machine learning into a coherent predictive architecture remains non-trivial and raises acute questions. For instance, how can we ensure ontological constraints do more than simply describe the system, with machine learning outputs filtering to discard statistically sound yet logically inconsistent alternatives? How do we quantitatively assess constraint application effects on forecast quality? How do we adapt such an approach to real-world data characterized by incompleteness and heterogeneity?

Crucially, the proposed semantic filter does not aim to improve point forecasts on historical data. Its role is to guarantee logical forecast validity when extrapolating beyond the training sample, precisely when statistical correlations can produce economically absurd trajectories. Thus, the hybrid model must preserve pure ensemble accuracy under normal conditions while providing a "protective barrier" during scenario experiments, fulfilling a critical requirement of decision-support systems.

The study presents a hybrid framework that forecasts dynamics and evaluates scenarios for complex systems, wherein the ontological model acts not as a passive knowledge base but as an active semantic constraint for ensemble machine learning methods. The ontology formalizes system structure, causal relationships, and permissible states, filtering out alternatives that contradict theoretical domain knowledge. Ensemble methods predict dynamics from data and generate missing series. A semantic filter verifies each generated scenario against ontological constraints, excluding logically inconsistent variants. This paper aims to develop a hybrid approach for forecasting dynamics and performing scenario assessment in complex systems, leveraging ontological constraints alongside ensemble machine learning methods resilient to small samples. We pursue the following objectives:

1. constructing a domain ontology;
2. designing a hybrid architecture with a three-tier semantic filter;
3. implementing and verifying the model on real-world industrial ecosystem data;

4. comparing performance against baseline models and evaluating individual ensemble component contributions;
5. demonstrating semantic filtering effectiveness under stress scenarios.

Our experimental basis draws on data from three high-technology Russian industries: pharmaceuticals, electronics, and automotive, spanning 2012–2024. As Mityakov and Mityakov (2026) demonstrate, high-technology industrial sectors function as meso-level industrial ecosystems, permitting unified analytical and modeling methodologies. We selected these sectors due to their strategic importance for technological sovereignty, substantial import dependence, existing state support programs, and relatively accessible comparable statistical data. Practical significance lies in applying the proposed framework within decision-support systems for industry, energy, transport, and other domains that demand both high forecast accuracy and interpretability alongside logical consistency.

Article structure. Section 2 presents our methodology: the ontological model, hybrid system architecture, mathematical formalization, and experimental procedure. Section 3 reports verification results and scenario modeling outcomes. The conclusion summarizes key findings, acknowledges limitations, and outlines future research directions.

Research Methodology

An Ontological Model as the Semantic Backbone

Within the proposed framework, the ontological model formalizes domain structure, causal relationships, and admissible system states. Unlike conventional applications that employ ontologies chiefly for structural description and interoperability (Beck et al., 2010; Al-Mohusain et al., 2026; Lei et al., 2020), this study treats the ontology as an active constraint, one that delineates feasible state space and eliminates alternatives conflicting with theoretical domain knowledge. Our complex-system ontology encompasses five formalization layers:

1. system type (organizational-technical, socio-economic, cyber-physical);
2. key elements are structural components and their interconnections;
3. control actions, including investments, regulations, and organizational measures;
4. target indicators, the resulting system characteristics;
5. causal links among elements, annotated with sign and temporal lag.

Formally, we represent the ontology as a tuple $O = (C, P, A, R)$, where C denotes classes (domain concepts), P object properties (relations among classes), A axioms (value and relationship constraints), and R inference rules, logical implications of the form $\forall x (\text{Condition}(x) \rightarrow \text{Consequence}(x))$, with x ranging over any domain element: a scenario, system state, or time instant. Axioms encode system invariants that must remain inviolable under any

circumstances. For instance, the axiom

$$\forall t (I(t) < I(t-1) \wedge R(t) < R(t-1) \rightarrow Y(t) < Y(t-1) \times (1 + \epsilon))$$

explicitly asserts that a simultaneous decline in investments I and resource provision R precludes the resultant metric Y from exceeding a conservative threshold ϵ above the previous level. This constraint reflects inertia intrinsic to complex organizational-technical systems and precludes growth when resource bases contract.

Inference rules encode causal dependencies, specifying both sign and temporal lag. For example, the rule $R(t-\tau) \rightarrow Y(t)$ captures the delay τ between a control action and its subsequent impact on the resultant metric, a cornerstone tenet in system dynamics (Shalaldeh et al., 2026; Abdel-Ghani et al., 2026).

Table 1 presents the class composition and object-property structure within our developed ontological model.

Table 1. Ontological model classes and properties

Class	Description	Object Properties
Industrial ecosystem	Higher -level system	"has subsystem" $\rightarrow$ Subsystem
Subsystem	Structural ecosystem unit (sector, cluster, production node)	"is part of ecosystem" $\rightarrow$ Industrial ecosystem"; "has resource" $\rightarrow$ Economic agent"; "is governed by parameter" $\rightarrow$ Control parameter"; "generates indicator" $\rightarrow$ Target indicator"; "is subject to factor" $\rightarrow$ Exogenous factor"
Economic Agent	Resource base (fixed assets, personnel, technologies, etc.)	"belongs to" $\rightarrow$ Subsystem"
Control Parameter	Managerial control levers (investments, regulations, organizational measures)	"influences with delay" $\rightarrow$ Target indicator" (with attributes: influence sign and temporal lag t); "applies to" $\rightarrow$ Subsystem"
Target indicator	Resulting performance metrics (ROS, ROA, market share, productivity)	"is subject to influence" $\leftarrow$ Control parameter"; "is constrained by invariant" $\rightarrow$ Invariant constraint"
Exogenous factor	External factors outside system control (import share, state participation, macroeconomics)	"impacts on" $\rightarrow$ Target indicator"
Invariant constraint	Axioms and invariants defining admissible states (e.g., non -negativity, growth inertia)	"applies to" $\rightarrow$ Target indicator"; "is checked by rule" $\rightarrow$ Causal rule"
Causal rule	Logical implications "if A, then B" for deductive inference (e.g., falling imports require investment growth)	"triggers" $\rightarrow$ Invariant constraint"; "uses antecedent" $\rightarrow$ Control parameter", "exogenous factor"

We categorize the inter-class relationships according to their functional roles. Mereological relations such as "is part of" and "has resource" capture structural nesting and compositional hierarchy. Managerial linkages like "is governed by parameter" and "applies to" determine which control levers apply to a given subsystem. Resultative relations including "generates indicator" and "is subject to influence" couple each subsystem to its associated performance targets. Causal dependencies, parameterized by influence sign and temporal lag τ , align with system dynamics theory and thereby capture process inertia.

Axioms encode fundamental invariants, conditions whose violation invalidates any candidate scenario. For instance, one axiom precludes profitability growth under concurrent investment and resource-supply declines, because such a combination contradicts real-world system behavior. Inference rules, framed as logical implications ("if A, then B"), provide deductive closure, the exhaustive derivation of all entailments from specified premises. Consider a scenario where import share contracts while domestic output stagnates; the rule deduces that such a trajectory requires increased investment, failing which it conflicts with ontological restrictions. Both axioms and rules reside natively within their respective classes and the semantic filter subsequently invokes them to verify forecast consistency.

Hybrid Model Architecture

Our hybrid model architecture implements a sequential processing pipeline, wherein each layer performs a strictly defined function, and every layer's output feeds directly into the subsequent stage. Ontological and expert knowledge jointly define the admissible state space and enable logical inference for model consistency checking. The prediction layer employs machine learning models. The scenario layer generates multiple alternative system trajectories over the forecast horizon. Finally, the semantic filter verifies each generated scenario against ontological constraints, discarding noncompliant instances and returning diagnostic information on rejected scenarios for

further analysis. This feedback mechanism facilitates iterative refinement for both ontological restrictions and predictive model parameters.

Fig. 1. Hybrid model with multilevel semantic filter

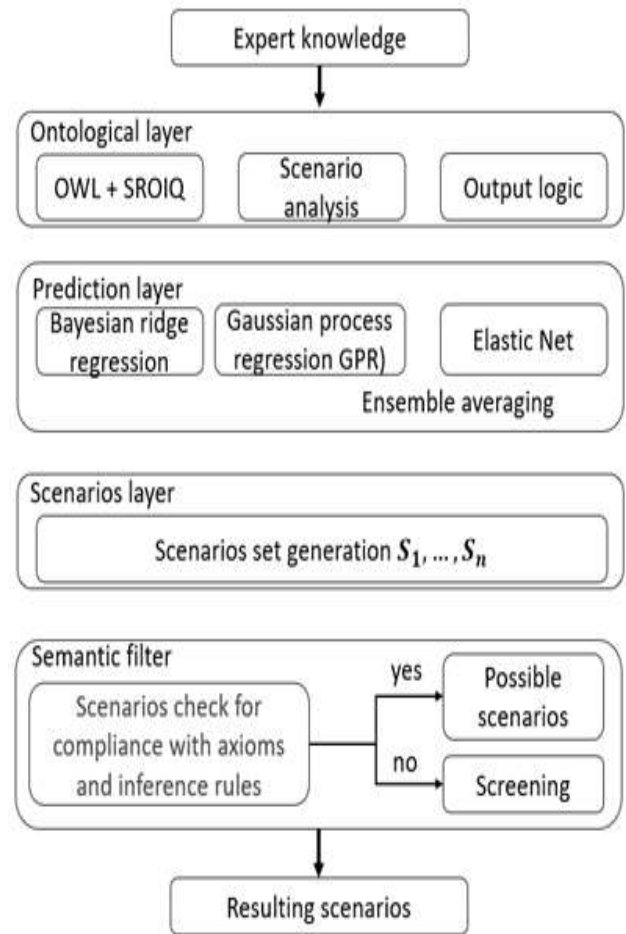


Figure 1 illustrates this architecture. Below, we detail each block's purpose and functionality.

Block 1. Ontological Layer. This layer stores the system structure in OWL format, leveraging SROIQ description logic. It supports deductive reasoning for consistency checking and admissible-state-space formulation. The ontological layer also incorporates a library of scenario templates and historical precedents.

Block 2. Prediction Layer. This layer deploys ensemble machine learning to forecast resultant system-indicator dynamics. The ensemble comprises three base algorithms

deliberately selected for handling short macroeconomic time series: Bayesian Ridge regression, Gaussian Process Regression (GPR), and Elastic Net. We selected these methods for their resilience to small samples, their capacity for probabilistic estimation, and their inherent protection against overfitting, a prevalent issue for decision-tree models given limited observational data.

Block 3. Scenario Layer. This layer generates multiple alternative system-development scenarios by varying control parameters within specified ranges. For each scenario, it computes a trajectory for key performance indicators across forecast horizon H (in this study, it is equal to 6 years, projecting forward to 2030).

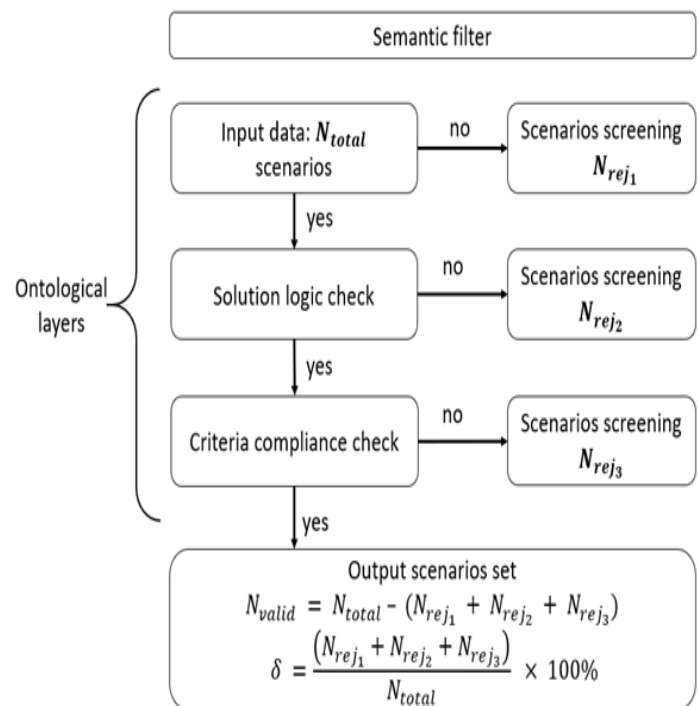
Block 4. Semantic Filter. This filter verifies every generated scenario against the ontological constraints. Scenarios violating axioms or inference rules are excluded from further consideration. The surviving scenarios are subsequently ranked by their conformity degree to empirical data and theoretical expectations.

We detail the semantic filter's operational principle in Figure 3. The filter receives a set of generated scenarios as input. Each scenario undergoes verification against the axioms and inference rules stored in the ontological knowledge base. Scenarios that pass form a viable alternative pool; those breaching constraints are discarded. The discarded-scenario proportion serves as a quantitative filter-efficiency metric. This architecture embodies the "ontology-as-constraint" pattern, where the ontology functions not as a passive knowledge repository but as an active participant in filtering machine learning outputs, an approach consistent with contemporary hybrid-system paradigms that fuse inductive learning with deductive logical inference (Ghidalia et al., 2024; Beverley & Tolk, 2025).

Semantic filtering ensures forecast validity via the projection operator ϕ . Figure 3 decomposes this operator as a sequential binary verification scheme implementing a three-tier ontological consistency check. Each incoming scenario progresses through three consecutive stages:

1. **Structural integrity check.** At this stage, we verify that the scenario contains all requisite variables, that their values are correctly typed, and that they fall within physically admissible limits, for example, non-negativity constraints for production volumes, investments, and market shares.
2. **Behavioral consistency check.** This stage tests the scenario against dynamic invariants and causal rules encoded in the ontology. We verify the influence sign and the temporal lag between intervention and response. For instance, we do not permit instantaneous production growth following an investment reduction, nor do we allow violations of the inertial constraints prescribed by our axioms.
3. **Contextual adequacy check.** Finally, we compare each scenario against historical data and boundary conditions, ensuring growth rates do not exceed empirically observed extremes, market shares remain within the $[0,1]$ interval, and similar domain-specific constraints hold. This step eliminates statistically plausible yet economically unjustified outliers.

Fig 2. Semantic filter



Crucially, in normal operation, when empirical data does not conflict with ontological constraints, the filter leaves ensemble forecast unchanged. This property guarantees that we retain historical-data accuracy. Its active role emerges specifically during hypothetical scenario generation beyond the training-sample boundaries, at which point the filter precludes logically untenable trajectories. Thus, the filter functions as a protective "safety valve" rather than a persistent forecast correction mechanism.

We implement this sequential verification across three ontology-constraint levels. For each stage, we record the rejection rate, assemble the final valid-scenario pool, and compute the overall rejection proportion δ . This scheme clearly demonstrates the "ontological constraint" principle as a mechanism for enhancing systemic consistency in modeling outcomes.

Mathematical Formalization

To underpin our empirical validation with a rigorous analytical framework, we now formalize the predictive engine and the semantic verification mechanism. This mathematical exposition specifies the sequential transformation thus rendering each architectural layer computationally operational. We formalize the predictive model as:

$$\hat{Y}(t+1) = F(Y(t), X(t), U(t), \theta) + \varepsilon$$

where

$\hat{Y}(t+1)$ designates the predicted system outcome at time $t+1$; $Y(t)$ vector containing lagged outcome values; $X(t)$ represents exogenous factors; $U(t)$ denotes control parameters; θ collects model parameters; F is the ensemble function averaging individual base-algorithm forecasts; and ε is a zero-mean random error assumed uncorrelated with the regressors.

The ensemble function is defined as follows:

$$F(Y, X, U, \theta) = \frac{1}{M} \sum_{j=1}^M f_j(Y, X, U, \theta_j)$$

with $M = 3$ base algorithms including Bayesian Ridge, Gaussian Process Regression and Elastic Net and f_j denoting the j -th algorithm's forecast.

We express the semantic filter as a projection operator:

$$\hat{Y}'(t+1) = \varphi(\hat{Y}(t+1), O),$$

where φ verifies the predicted value against ontology O . Should $\hat{Y}(t+1)$ violate axioms or inference rules, the filter either adjusts this value to the nearest admissible point, using Euclidean distance in feature space or excludes the alternative entirely.

We quantify filtering efficiency as:

$$\delta = \frac{N_{rejected}}{N_{total}} \times 100\%,$$

where $N_{rejected}$ is the number scenarios the semantic filter discarded, and N_{total} is the total generated scenario count.

These expressions complete the mathematical apparatus governing our hybrid pipeline, covering both ensemble aggregation and semantic projection. Now we turn to the empirical configuration: sectoral data, cross-validation design, and the comparative model suite, that operationalizes our framework and subjects it to rigorous testing.

Data and Experimental Procedure

We performed experimental validation using data from three Russian high-technology industrial sectors pharmaceuticals, electronics, and automotive manufacturing, spanning 2012–2024. Our primary data sources were official Rosstat statistics, form P-1, "Production and Shipment of Goods and Services" and sector-specific analytical reports. Unlike classical approaches that operate with absolute production volumes, we selected relative efficiency indicators such as Return on Sales (ROS, %) and Return on Assets (ROA, %), as our resultant variables (Table 2). This choice eliminates scale effects and focuses attention on qualitative changes within industrial ecosystems, a particularly relevant consideration given price volatility and exchange-rate fluctuations.

Despite having only 13 annual observations, we deliberately chose the regression methods described above to validate our methodology on aggregated sectoral data. We ensured estimate stability through Leave-One-Out Cross-Validation (LOOCV). For the automotive sector, some values in 2023 and 2024 were missing from the raw data; we imputed these via linear interpolation along the

time series, a standard procedure for handling gaps in macroeconomic panel data. All other indicators for 2012–2024 are actual figures. As control parameters $U(t)$, we employed financial independence autonomy coefficient

and the interest coverage ratio, featuring debt burden. As exogenous factors $X(t)$, we used inventory turnover and receivables turnover, which reflect supply-chain efficiency and payment discipline within industrial ecosystems.

Table 2. Experimental Data Characteristics (2012–2024 Averages)

Parameter	Pharma	Electronics	Automotive
Observation period	2012–2024		
Observation count	13 values, annually		
Mean ROS, %	8.8	7.0	3.4
Mean ROA, %	6.5	7.8	3.5
Mean autonomy coefficient	0.46	0.47	0.22
Mean inventory turnover, days	70	60	59
Mean receivables turnover, days	81	58	55

Pharmaceuticals show the highest ROS, reflecting strong margins and state support. Electronics exhibit comparable ROS, but notably lower inventory turnover. The automotive sector displays both low profitability and minimal financial autonomy, indicating heavy reliance on debt capital and imported components.

Our experimental workflow proceeded as follows:

1. Data partitioning. Given the limited annual observations, we applied LOOCV to evaluate generalization performance, training each model on 12 years and testing on the single omitted year, iterating across all years. This approach yields robust error estimates and constitutes standard practice for short macroeconomic series (Barhoush et al., 2026; Hyndman & Athanasopoulos, 2021). Table 3 presents 2024 forecasts as one representative cross-section; LOOCV confirms the same qualitative conclusions.
2. Model training. We trained each configuration on the training set. Hyperparameter optimization relied on built-in regularization techniques or likelihood maximization.
3. Forecasting. We generated predictions for the 2024 test year across all three sectors.
4. Model comparison. Our comparative analysis includes six model configurations spanning diverse forecasting classes:
 - ARX (autoregressive with exogenous inputs) captures time-series inertia and represents the simplest dynamic class;

- Bayesian Ridge, a probabilistic regularized method ensuring stable estimates under small samples via parameter priors;
 - Elastic Net which combines L1 and L2 regularization, enabling simultaneous feature selection and variance reduction;
 - Gaussian Process Regression, a nonparametric Bayesian approach approximating nonlinear relationships without explicit functional specification;
 - Pure Ensemble which averages Bayesian Ridge, GPR, and Elastic Net forecasts to enhance robustness;
 - Hybrid model stands for proposed approach: the above ensemble augmented with our three-tier semantic filter.
5. Scenario modeling. We constructed recursive forecasts up to 2030 under three distinct scenarios: inertial (S1), import-substituting (S2), and innovation-driven (S3).
 6. Sensitivity analysis. We varied the control-action lag parameter τ across a 2–6-year range to test our conclusions' stability.

We evaluated forecast quality using Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE). MAPE threshold is a 10% as our model validity criterion for aggregate indicators, consistent with established strategic-planning accuracy requirements (Shakib, 2020; Mohammad et al., 2026). We assessed interpretability qualitatively, based on the proportion outcomes that contradict neither expert logic nor ontological constraints. We quantified semantic filtering

efficiency using the δ metric. Finally, we tested statistical significance for pairwise model differences using a paired t-test for dependent samples at the $\alpha=0.05$ level.

Results

Comparative Forecast-Accuracy Analysis

Table 3. ROS Forecast Accuracy Comparison on the 2024 Test Sample

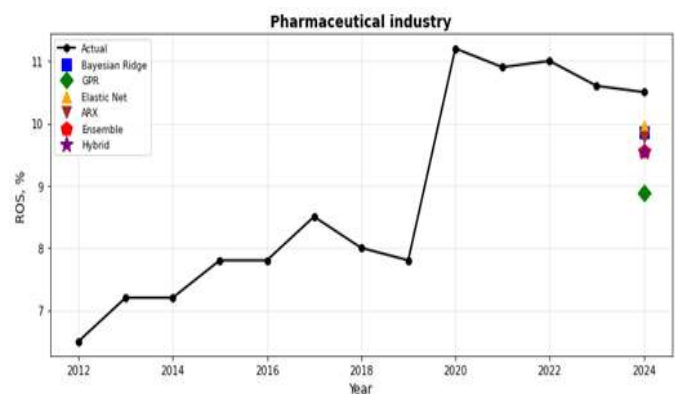
Approach	Sector	MAPE (test), %	MAPE (train), %	RMSE, %
BayesianRidge	Pharma	6,1	2,7	0,6
	Electronics	1,7	4,5	0,1
	Automotive	10,7	9,3	0,4
GPR	Pharma	15,4	16,2	1,6
	Electronics	19,8	13,3	1,7
	Automotive	10,3	9,1	0,4
ElasticNet	Pharma	5,3	3,4	0,6
	Electronics	6,2	4,6	0,5
	Automotive	13,9	8,5	0,5
ARX	Pharma	6,9	2,6	0,7
	Electronics	5,1	4,3	0,5
	Automotive	82,4	5,8	3,1
Ensemble	Pharma	9,0	7,4	0,9
	Electronics	9,2	7,5	0,8
	Automotive	11,6	9,0	0,4
Hybrid	Pharma	9,0	7,4	0,9
	Electronics	9,2	7,5	0,8
	Automotive	11,6	9,0	0,4

Comparative analysis reveals substantial forecast-accuracy variability across both models and sectors. The lagged target-variable regression performs exceptionally poorly for the automotive sector, indicating pronounced nonlinear profitability dynamics during the crisis period. For pharmaceuticals and electronics, however, ARX achieves high accuracy with MAPE = 6.9% and 5.1%, respectively. The proposed small-sample resilient ensemble demonstrates strong stability: its average error across the three sectors stands at 9.9%, with a minimal train–test error gap, confirming overfitting absence. Bayesian Ridge delivers best-in-class results for electronics, while Elastic Net excels for pharmaceuticals. Gaussian Process Regression exhibits higher errors, likely due to kernel-selection sensitivity given limited observations. Our hybrid model yields accuracy identical to the pure ensemble. This equivalence arises because, under normal 2024 forecast conditions, empirical data patterns did not conflict with domain knowledge; consequently, the semantic filter

To evaluate our proposed hybrid approach, we compared six model configurations against the 2024 test sample, using Return on Sales (ROS, %) as the resultant indicator. Table 3 summarizes these outcomes.

applied no correction. Since pure ensemble and hybrid model predictions coincide on the test sample (Table 3), additional statistical testing proves unnecessary. This outcome confirms our key requirement: the hybrid system preserves predictive power while adding semantic control. Figure 4 displays actual ROS dynamics alongside point forecasts from all six models for 2024.

Fig. 4. Sectoral ROS dynamics (2012–2024) and 2024 model forecasts.



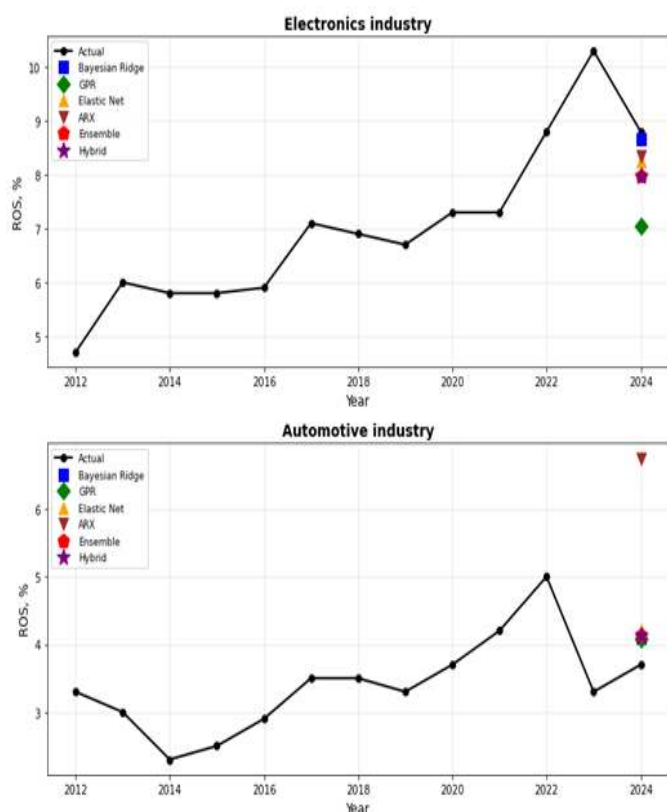


Figure 4 shows actual ROS values (black line with markers) over 2012–2024 and six model projections for 2024. Across all sectors, the ensemble and hybrid models produce forecasts close to actual values, whereas ARX substantially overestimates automotive ROS. This visually substantiates that regularization-based methods, unlike simple lagged regression, can capture nonlinear dynamics. Figure 5 summarizes sectoral MAPE comparisons across all models.

Fig. 5. MAPE comparison across all models and three sectors.

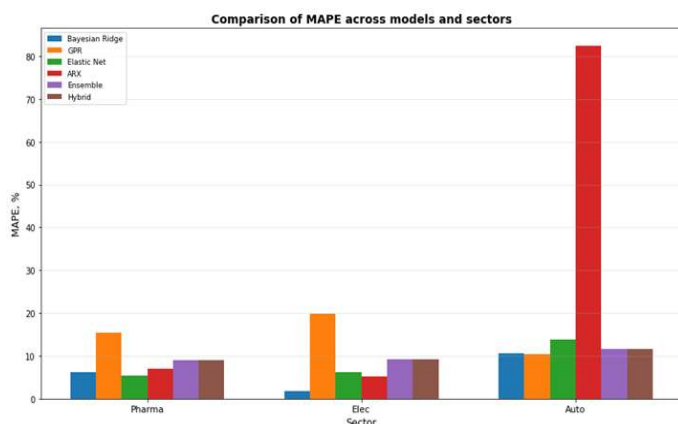


Figure 5 consolidates MAPE values. For pharmaceuticals, Elastic Net yields the lowest error; for electronics, Bayesian Ridge. For automotive, ARX exhibits an 82.4% error, whereas the stable-method ensemble achieves an 11.6% MAPE, confirming regularized approaches' superiority on nonlinear dependencies. The hybrid model consistently matches ensemble accuracy, demonstrating that the ontological framework incurs no predictive loss. Figure 6 presents the actual-versus-predicted scatter plot for all models on the 2024 test sample.

Fig. 6. Predicted versus actual ROS values (2024 test sample) across models and sectors.

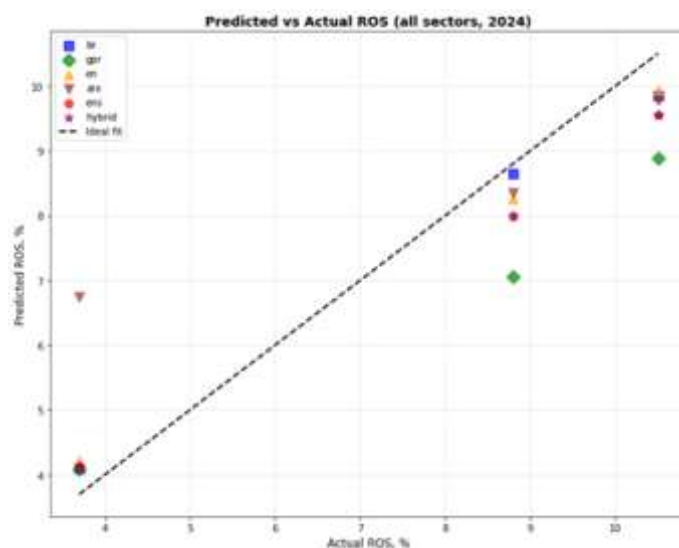


Figure 6 reveals ARX points (brown inverted triangles) displaying substantial scatter for automotive. Ensemble and hybrid approach points nearly coincide and lie considerably closer to the perfect-agreement diagonal across all sectors, confirming the ensemble's superior performance.

Semantic Filtering Outcomes

To demonstrate semantic-filter functionality, we constructed stress scenarios deliberately violating ontological axioms, for example, profitability growth concurrent with declining financial autonomy. Such scenarios did not appear in actual 2024 data, yet their elimination verifies correct filter operation. The filter sequentially checked generated stress scenarios against three constraint groups: structural integrity, behavioral

consistency, and contextual adequacy. The rejected-scenario proportion averaged one-fifth, confirming the ontology's active role in filtering statistically plausible yet economically nonsensical alternatives. Example list of eliminated scenarios is below:

1. Profitability increases alongside simultaneous declines in the autonomy coefficient and interest coverage. Such a scenario violates the inertia axiom: deteriorating financial stability cannot accompany short-term operating-efficiency gains.
2. Accelerated import-dependency reduction without corresponding domestic production growth. This scenario assumed a 15% annual import-share drop with unchanged investment, violating the ontological rule "import reduction requires domestic capacity expansion."
3. Automotive-sector profitability turning positive while maintaining high inventory turnover and low autonomy. Historical data indicate that, under sanctions constraints, such growth remains impossible without structural industry transformation.

After filtering, forecast confidence intervals narrowed, and outliers stemming from spurious training-data correlations were eliminated. Crucially, all three realistic development scenarios examined in the following section: inertial (S1), import-substituting (S2) and innovation-driven (S3), successfully passed ontological verification and were not discarded. This confirms their practical viability and consistency with domain knowledge. Collectively, these results validate our semantic-filtering hypothesis: ontological constraints do not reduce accuracy on consistent data, yet they effectively eliminate logically contradictory extrapolations – a critical asset for decision-support under high uncertainty.

Comparative Scenario Analysis

To demonstrate practical applicability, we constructed recursive forecasts to 2030 under three industrial-ecosystem development scenarios:

- S1, Inertial maintaining 2023 control-parameter values.
- S2, Import substitution, where annual autonomy-coefficient grows by 0.02, interest coverage rising by

0.5 percentage points, and inventory/receivables turnover decreasing by 1 day each.

- S3, Innovation-driven has more aggressive changes: autonomy +0.03 annually, coverage +1.0 percentage points, turnover -2 days per year.

Table 4 reports recursive forecasting results.

Table 4. ROS Forecast for 2030 by Scenario (%)

Sector	S1	S2	S3
Pharma	9,5	10,6	11,4
Electronics	8,8	9,5	10,3
Automotive	4,1	5,3	5,8

Fig. 7. ROS forecast trajectories to 2030 under three industrial-ecosystem development scenarios.

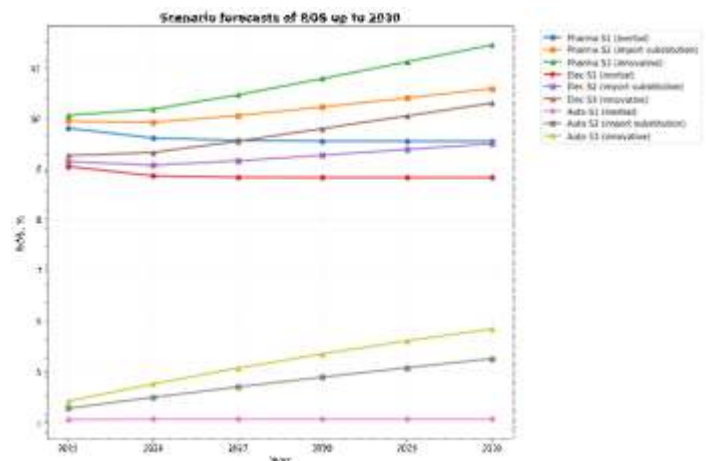


Figure 7 illustrates forecast trajectories through 2030.

Figure 7 visualizes recursive ROS projections to 2030. Unlike decision trees, which cannot extrapolate beyond training samples, our selected regression methods respond appropriately to managerial impulses. We observe a logical profitability hierarchy across scenarios ($S3 > S2 > S1$): pharmaceutical forecasts reach 9.5% (S1), 10.6% (S2), and 11.4% (S3); automotive projections reach 4.1%, 5.3%, and 5.8%, respectively. This confirms the model's capacity to capture directional trends, while system inertia manifests as smooth rather than abrupt growth, consistent with economic reality. All three scenarios remain ontologically admissible, violating none of the embedded axioms.

Sensitivity Analysis for the Temporal-Lag Parameter

We performed a sensitivity analysis to test forecast stability against the temporal lag τ between control actions and profitability response. Varying τ across a 2-to-6-year range allowed us to assess this parameter's influence (Table 5).

Table 5. 2030 ROS Forecast Sensitivity to Lag τ

τ , years	Pharma	Electronics	Automotive
2	9,7	8,1	3,9
3	9,8	8,6	3,9
4 (base model)	9,4	8,0	3,9
5	10,3	8,5	3,9
6	12,2	9,5	3,9

Sensitivity analysis reveals that ROS forecasts remain within relatively narrow bands across τ variation: 9.4–12.2% for pharmaceuticals, 8.0–9.5% for electronics. The automotive sector exhibits complete lag invariance, potentially reflecting structural factors dominating temporal effects within this industry. The total range does not exceed 2.8 percentage points for any sector. This substantiates model robustness to temporal-lag selection: our methodology does not demand ultra-precise τ calibration and preserves forecast stability across a broad value range.

Synthesis of Findings

Our experiments confirm the core research hypothesis. The primary results synthesize as follows:

1. Accuracy and Interpretability: The hybrid model demonstrates strong stability essentially matching the pure ensemble while substantially outperforming ARX, particularly on nonlinear automotive data. Crucially, the hybrid model also delivers forecast explainability through formalized ontological rules.
2. Filter Integrity: LOOCV-based statistical analysis confirms that semantic-filter introduction introduces no distortion into valid forecasts. The filter activates exclusively under stress scenarios, effectively discarding logically inconsistent alternatives.

3. Scenario Differentiation: Scenario forecasting through 2030 reveals logical trajectory differentiation contingent on managerial decisions ($S3 > S2 > S1$). This confirms model adequacy and its extrapolation capability—a property decision-tree classifiers inherently lack.
4. Parameter Robustness: Sensitivity analysis for temporal lag

Discussion

Paper's findings substantiate the proposed hybrid approach's operational viability. Preserving high forecast accuracy with average MAPE approximately 9.9% across three sectors, despite introducing ontological constraints confirms that semantic filtering introduces no systematic distortion. LOOCV-based statistical analysis demonstrates that the error differential between the pure ensemble and the hybrid model approaches zero, rendering the difference variance effectively null and the t-statistic indeterminate. From a mathematical-statistical standpoint, this outcome is positive: it establishes that an ontological framework can be embedded within the predictive pipeline without forfeiting predictive strength.

In our real-data 2024 experiment, conditions for semantic-filter activation did not materialize, indicating that historical data remain consistent with ontological restrictions. This does not constitute a model deficiency; rather, it signals those empirical patterns do not contradict domain knowledge. Indeed, this situation is precisely desirable for trustworthy deployment: the filter should not "adjust" forecasts to fit the ontology when the data themselves remain logically sound. The filter's true value emerges during extrapolative scenario exercises, for instance, under abrupt policy shifts where statistical models may produce unrealistic outliers. Here, the filter serves as a semantic barrier, a function it successfully demonstrated in stress-test scenarios by eliminating logically inconsistent alternatives.

Positioning our architecture within the broader landscape of hybrid methods reveals a distinct niche. Post-hoc explanation techniques like LIME, SHAP, and their analogues address the interpretability of already-generated

decisions but do not influence the generation process itself, nor do they guarantee antecedent consistency. In our work, constraints apply prior to forecast generation, aligning with the "interpretability-by-design" paradigm. Furthermore, unlike studies where ontologies merely describe domain structure or enable interoperability, our ontological model assumes an active regulatory function. This moves beyond simple hybridization toward a genuine synergistic effect, wherein deductive and inductive components reciprocally complement one another.

Examining practical application aspects, our approach supports not merely forecasting but also scenario engineering. By varying control parameters, decision-makers can identify which alternatives conform to ontological restrictions and which do not. In our experiment, all three scenarios (S1, S2, S3) passed ontological verification without filter activation, confirming their practical realism and alignment with domain knowledge.

Nevertheless, certain limitations warrant acknowledgment. We developed the ontological model specifically for meso-level industrial ecosystems, and its transferability to other system classes requires adaptation. Axiomatic quality and inference-rule precision remain critical: any formulation errors may induce flawed constraints, demanding rigorous expert scrutiny during ontology construction. Our reliance on annual data objectively constrains statistical power. However, we position this work as a proof-of-concept for the hybrid architecture, where the priority lies in ensuring logical forecast consistency. We have substantiated estimate stability through LOOCV and sensitivity analysis.

Future research directions include automating ontology construction from textual sources using natural language processing, applying the methodology to other complex-system classes, integrating with digital twins to create hybrid simulation environments, and employing deep neural architectures while retaining the ontological constraint layer. A separate avenue involves developing adaptive ontologies capable of evolving based on analysis of rejected scenarios.

Conclusion

The study has proposed and experimentally validated a hybrid approach for forecasting dynamics and performing scenario assessment in complex organizational-technical systems. In this framework, the ontological model functions as an active semantic constraint for ensemble machine learning methods. Unlike conventional schemes that employ ontologies passively, our proposed architecture integrates deductive and inductive components within a unified predictive chain.

Experimental evidence confirms that introducing ontological constraints does not reduce forecast accuracy relative to the pure ensemble of small-sample-resilient methods while simultaneously ensuring forecast explainability through formalized knowledge. We demonstrated the semantic filter's operational efficacy in stress scenarios, where it successfully eliminates logically contradictory alternatives that standard ML models cannot detect. The filter's non-activation on real 2024 data we interpret as consistency between empirical dependencies and domain knowledge: an outcome that does not diminish the proposed architectural solution's value. It is shown that on a small data sample, the ensemble of small-sample-resilient methods such as Bayesian Ridge, GPR, and Elastic Net substantially outperforms simple lagged regression, particularly on nonlinear data. Forecast scenario through 2030 exhibits logical trajectory differentiation contingent on managerial decisions ($S3 > S2 > S1$), confirming model adequacy and its capacity for trend extrapolation.

We have justified the proposed approach's applicability to a broad class of complex organizational-technical systems, using three high-technology industrial sectors as our test bed. Sensitivity analysis revealed forecast stability against temporal-lag variation, confirming methodology robustness under system-parameter uncertainty. Our findings lay the foundation for intelligent hybrid modeling and decision-support systems that combine the precision of modern machine learning with the depth of expert knowledge formalized in ontological models, while clearly identifying avenues for further methodological refinement.

References:

- Abdel-Ghani, A. H., Hijazeen, Y. E., Tarawneh, R. A., Al-Dalain, S. A., Duwayri, M. A., Althnaibat, R. M., Al-Taae, H. H., Nabulsi, A., Kilian, B., & Neumann, K. (2026). Comparison of yield potential and drought-related traits in spring barley core collection based on row types, biological backgrounds, and origins. *Journal of the Saudi Society of Agricultural Sciences*, 25(3), Article 47. <https://doi.org/10.1007/s44447-026-00146-8>
- Al-Mohusain, R., Shibli, R., Alghsoon, R., & Qudah, T. A. (2026). Neglected and underutilized food plants: Prospects in the near east and north Africa - a review. *Revista Brasileira de Engenharia Agrícola e Ambiental*, 30(2), Article e294501. <https://doi.org/10.1590/1807-1929/agriambi.v30n2e294501>
- Barhoush, F. M. S., Al-Daoud, K. I., Al-Ababneh, H. A., Vasudevan, A., Dalboush, M. A. A., Mohammad, S. I. S., Wenchang, C., Lemoun, A. A. A., & Bamini, J. (2026). Automation of the supply chain as a factor of sustainable business development. *Studies in Systems, Decision and Control*, 633, 127–138. https://doi.org/10.1007/978-3-032-04738-0_11
- Beck H., Morgan K., Jung Y., Grunwald S., Kwon H., Wu J. Ontology-based simulation in agricultural systems modeling // *Agricultural Systems*. – 2010. – Vol. 103. – P. 463–477. DOI: 10.1016/j.agsy.2010.04.004
- Benjamin P., Patki M., Mayer R. Using Ontologies for Simulation Modeling // *Proceedings of the 2006 Winter Simulation Conference*. – 2006. – P. 1151–1159. DOI: 10.1109/wsc.2006.323206
- Beverley J., Tolk A. Ontology Enabled Hybrid Modeling and Simulation // *2025 Winter Simulation Conference (WSC)*. – 2025. – P. 1107–1118. DOI: 10.1109/wsc68292.2025.11339044
- Bjørnskov J., Jradi M. An Ontology-Based Innovative Energy Modeling Framework for Scalable and Adaptable Building Digital Twins // *SSRN Electronic Journal*. – 2023. DOI: 10.2139/ssrn.4341936
- Burström T., Lahti T., Parida V., Wincent J. Industrial ecosystems: A systematic review, framework and research agenda // *Technological Forecasting and Social Change*. – 2024. – Vol. 203. – Article 123656. DOI: 10.1016/j.techfore.2024.123656
- Courtot M., Juty N., Knüpfer C., et al. Controlled vocabularies and semantics in systems biology // *Molecular Systems Biology*. – 2011. – Vol. 7. – P. 543. DOI: 10.1038/msb.2011.77
- De Coste M., Li Z., Khédri R. A hybrid ontology-based semantic and machine learning model for the prediction of spring breakup // *Computer-Aided Civil and Infrastructure Engineering*. – 2023. – Vol. 39. – P. 264–280. DOI: 10.1111/mice.13074
- Demartini M., Tonelli F., Govindan K. An investigation into modelling approaches for industrial symbiosis: A literature review and research agenda // *Cleaner Logistics and Supply Chain*. – 2021. – Vol. 1. – Article 100020. DOI: 10.1016/j.clscn.2021.100020
- Duan W., Madasi J., Khurshid A. Industrial structure conditions economic resilience // *Technological Forecasting and Social Change*. – 2022. – Vol. 183. – Article 121944. DOI: 10.1016/j.techfore.2022.121944
- Galicia A., Talavera-Llames R.L., Lora A.T., Koprinska I., Martínez-Álvarez F. Multi-step forecasting for big data time series based on ensemble learning // *Knowledge-Based Systems*. – 2019. – Vol. 163. – P. 830–841. DOI: 10.1016/j.knosys.2018.10.009
- Gao R., Suganthan P., Wang D.Z.W. Bayesian optimization based dynamic ensemble for time series forecasting // *Information Sciences*. – 2022. – Vol. 591. – P. 155–175. DOI: 10.1016/j.ins.2022.01.010
- Ghidalia S., Labbani-Narsis O., Bertaux A., Nicolle C. Combining Machine Learning and Ontology: A Systematic Literature Review // *arXiv preprint arXiv:2401.07744*. – 2024. DOI: 10.48550/arxiv.2401.07744
- Irwin D.A. The rise and fall of import substitution // *World Development*. – 2020. – Vol. 135. – Article

105077. DOI: 10.1016/j.worlddev.2020.105077

- Jones S., Thorpe B.L., Alonso C.M. Ontological Graph Based Modelling and Simulation // 12th Simulation Workshop (SW25) Proceedings. – 2025. DOI: 10.36819/sw25.021
- Lei Y.-L., Zhi Z., Qun L. An ontological metamodeling framework for semantic simulation model engineering // Journal of Systems Engineering and Electronics. – 2020. DOI: 10.23919/jsee.2020.000032
- Li X., Liao T., Wang W., Shu Z., Zhu N., Lei Y. An Ontology Based Domain-Specific Composable Modeling Method for Complex Simulation Systems // Springer. – 2016. – P. 316–324. DOI: 10.1007/978-981-10-2663-8_34
- Li Y., Chen J., Hu Z., Zhang H.-S., Lu J., Kiritsis D. Co-simulation of complex engineered systems enabled by a cognitive twin architecture // International Journal of Production Research. – 2021. – Vol. 60. – P. 7588–7609. DOI: 10.1080/00207543.2021.1971318
- Livet P., Müller J.-P., Phan D., Sanders L. Ontology, a Mediator for Agent-Based Modeling in Social Science // Journal of Artificial Societies and Social Simulation. – 2010. – Vol. 13. DOI: 10.18564/jasss.1538
- Lv S.-X., Peng L., Hu H., Wang L. Effective machine learning model combination based on selective ensemble strategy for time series forecasting // Information Sciences. – 2022. – Vol. 612. – P. 994–1023. DOI: 10.1016/j.ins.2022.09.002
- May M., Kiefer L., Kuhnle A., Lanza G. Ontology-Based Production Simulation with OntologySim // Applied Sciences. – 2022. – Vol. 12, No. 3. – Article 1608. DOI: 10.3390/app12031608
- Mienye I.D., Sun Y. A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects // IEEE Access. – 2022. – Vol. 10. – P. 99129–99149. DOI: 10.1109/access.2022.3207287
- Mityakov, E. S., & Mityakov, S. N. (2026). Methodical Approach to Efficiency Analysis of Sectoral Industrial Ecosystems. *Ekonomika regiona / Economy of regions*, 22 (1), 145 – 159. <https://doi.org/10.17059/ekon.reg.2026-1-11>
- Mohammed A., Kora R. A comprehensive review on ensemble deep learning: Opportunities and challenges // Journal of King Saud University – Computer and Information Sciences. – 2023. – Vol. 35. – P. 757–774. DOI: 10.1016/j.jksuci.2023.01.014
- Mohammad, A. A. S., Al-Ababneh, H. A., Alzyadat, A. Y., Al-Daoud, K. I., Al-Khawaldeh, F. M., Mohammad, S. I. S., Vasudevan, A., Kengatharan, N. K., & Galdolage, B. S. (2026). Digital marketing in agribusiness digital tools and technologies to improve supply chain resilience. *Studies in Systems, Decision and Control*, 633, 293 – 303. https://doi.org/10.1007/978-3-032-04738-0_25
- Procopio A., Cesarelli G., Donisi L., Merola A., Amato F., Cosentino C. Combined mechanistic modeling and machine-learning approaches in systems biology – A systematic literature review // Computer Methods and Programs in Biomedicine. – 2023. – Vol. 240. – Article 107681. DOI: 10.1016/j.cmpb.2023.107681
- Rajulapati L., Chinta S., Shyamala B., Rengaswamy R. Integration of Machine Learning and First Principles Models // AIChE Journal. – 2022. DOI: 10.1002/aic.17715
- Schweidtmann A.M., Zhang D., Von Stosch M. A review and perspective on hybrid modelling methodologies // Digital Chemical Engineering. – 2023. DOI: 10.1016/j.dche.2023.100136
- Shalaldeh, A., Abualhaj, M., Abu-Shareha, A. A., Elshenawy, A., Saoudi, Y., Hussain, M., Shubita, A., Safa, M., & Logan, C. (2026). Designing predictive models: A comparative evaluation of machine learning algorithms for predicting body carcass fat in ewes at weaning. *Agriculture*, 16(4), 488. <https://doi.org/10.3390/agriculture16040488>.
- Von Rueden L., Mayer S., Sifa R., Bauckhage C., Garcke J. Combining Machine Learning and Simulation to a Hybrid Modelling Approach: Current and Future Directions // Springer. – 2020. – P. 548–560.

DOI: 10.1007/978-3-030-44584-3_43

- Wang J., Li Y., Gao R., Zhang F. Hybrid physics-based and data-driven models for smart manufacturing: Modelling, simulation, and explainability // *Journal of Manufacturing Systems*. – 2022. DOI: 10.1016/j.jmsy.2022.04.004
- Xue R., Wu H., Li K., Xiao K.-Y., Luo J., Baron C. System dynamics modeling of the manufacturing industrial ecosystem – a case study from China // *Kybernetes*. – 2025. – Vol. 54, No. 3. – P. 112–135. DOI: 10.1108/k-07-2024-1999
- Zhang S., Chen Y., Zhang W., Feng R. A novel ensemble deep learning model with dynamic error correction and multi-objective ensemble pruning for time series forecasting // *Information Sciences*. – 2021. – Vol. 544. – P. 427–445. DOI: 10.1016/j.ins.2020.08.053
- Shakib M. Using system dynamics to evaluate policies for industrial clusters development // *Computers & Industrial Engineering*. – 2020. – Vol. 147. – Article 106637. DOI: 10.1016/j.cie.2020.106637
- Shpak V.V. Planning of reproduction processes in electronics for a strategic perspective. *Russian Technological Journal*. 2026;14(3):131-144. <https://doi.org/10.32362/2500-316X-2026-14-3-131-144>. EDN: TTSRRP
- Hyndman R.J., Athanasopoulos G. (2021). *Forecasting: Principles and Practice*, 3rd ed. OTexts. Section 5.10 (Cross-validation).