Time-Series Forecasting with SARIMAX for Intent Prediction

Nagham Hachem, Manh Cuong Nguyen, and Éric Renault

Original scientific article

***Abstract*—By converting high-level user objectives into workable settings, intent-based networking makes autonomous and flexible network administration possible. However, without the ability to proactively predict network intents across different temporal granularities, its full potential is still constrained. Using practical operational datasets, this study examines the use of SARIMAX (Seasonal AutoRegressive Integrated Moving Average with eXogenous variables) for multi-scale intent prediction. With a fixed 1-day forecast horizon, we perform extensive trials under two training regimes—8-day and 10-day historical data—across prediction windows of 5, 10, 15, 20, and 25 minutes. Our findings show a consistent trade-off between granularity and stability: longer windows (20–25 minutes) produce smoother forecasts at the cost of increased lag and decreased sensitivity to sudden changes, while shorter windows (5–10 minutes) offer greater responsiveness to real-time fluctuations but are prone to noise. There were clear bias-variance trade-offs between the two training durations, with the 5-minute window achieving the lowest MAE and the 25-minute window minimizing RMSE. The 10-minute setup reliably balanced responsiveness and stability despite regularly high MAPE values (over 280%), making it operationally appropriate for IoT service orchestration and intent-driven 5G slice management. The study lays the groundwork for future machine learning and hybrid model integration to improve intent prediction in dynamic network contexts and emphasizes the crucial role temporal aggregation plays in forecast success.**

***Index Terms*—SARIMAX, ARMA family, Intent, Forecasting.**

I. Introduction

Intent-based networking's full potential relies on predicting user intent at various time granularities, from long-term planning to real-time adaption [1], [2], [3]. While reactive approaches ignore temporal connections, time series forecasting allows proactive operation over particular spans.

This work uses the SARIMAX model [4], [5] for multi-scale intent prediction in order to handle non-stationary, seasonal, and exogenously generated intent data. SARIMAX combines autoregressive, moving average, differencing, seasonal, and exogenous components to capture short-term fluctuations like hourly QoS variations, mid-term demand evolution, and long-term periodic trends. With a focus on cross interval dependence and external factors like network congestion, the model is assessed against ARMA family baselines [4], [5], [6], [7], [8], [9], [10], [11].

This paper builds on our previous conference study [12] by conducting an innovative experimental investigation using the NILE framework and realistic operational data. This journal submission constitutes a substantial extension with the following novel contributions:

- *New Realistic Dataset:* The introduction and analysis of a second, realistic operational dataset (IWIB-5GNET) that was not part of the conference version.
- *Multi-Horizon Evaluation:* A systematic evaluation across five prediction windows (5, 10, 15, 20, and 25 minutes) under two training regimes (8-day and 10-day historical data).
- *Quantitative Trade-off Analysis:* A new, in-depth analysis of the bias-variance trade-off and granularity-stability relationship as a function of temporal aggregation.
- *Controlled Experimental Design:* A more robust experimental methodology that isolates the effects of window size and training data volume on forecast performance.

These additions provide results and insights not found in existing literature, demonstrating how temporal aggregation—rather than model complexity—is the primary factor influencing forecast behavior in intent-based networking scenarios.

This approach fills the temporal flexibility gap found by standardization organizations like GSMA and ETSI by combining intent-based networking with multi-interval SARIMAX forecasting. The suggested method supports real-world scenarios like 5G slice elasticity and IoT service chain orchestration and permits proactive resource alignment in closed loop systems.

This research advances intent-based forecasting by:

- SARIMAX with endogenous and exogenous variables for multiscale intent prediction.
- A comprehensive and unique experimental investigation employing practical operational data that goes beyond [12].
- To determine the ideal time forecasting interval, quantitative error analysis is used.
- Systematic comparison over several time periods using MAE, RMSE, and MAPE.
- Future learning-based extensions will benefit from new experimental findings.

This paper is organized as follows. In Section II, standardized perspectives are reviewed and the concept of intent

Manuscript received February 3, 2026; revised February 27, 2026. Date of publication March 31, 2026. Date of current version March 31, 2026. The associate editor prof. Pascal Lorenz has been coordinating the review of this manuscript and approved it for publication.

This paper was presented in part at the International Conference on Software, Telecommunications and Computer Networks (SoftCOM) 2025.

Authors are with the LIGM, Univ. Gustave Eiffel, CNRS, ESIEE Paris, 93162 Marne-la-Vallee, France (e-mails: firstname.lastname@esiee.fr).

Digital Object Identifier (DOI): 10.24138/jcomss-2025-0298

is defined. The function of intent forecasting in proactive networking is discussed in Section III. The ARMA family is introduced and the usage of SARIMAX is encouraged in Section IV. The datasets and preparation procedure are explained in Section V. Results and comparisons from experiments are presented in Section VI. The work is concluded and future research directions are outlined in Section VII.

II. What is Intent?

In simple terms, intent is an abstraction that describes what a user wants the network to accomplish rather than how the system ought to do it [2]. This is a fundamental difference between user objectives and low-level configuration, which forms the basis of automation in modern networks. Intents are communicated by:

- Primitives with structure, like

  `<domain, attribute, operation>`
- Inputs in natural language.
- Templates for policies.

Furthermore, intent engines compute them and display them as machine-executable settings [2], [1]. One of the essential components of this paradigm is closed-loop verification, which uses real-time telemetry to verify that the present network state matches to the stated aim [2], [3]. When the system detects misalignment, it can automatically correct the behavior. This paradigm makes use cases easier, like:

- With 5G network slicing, objectives define slice-specific specifications such QoS [1]
- Domain-spanning IoT orchestration, wherein intents create and oversee complex service chains [3]
- More flexibility and less manual intervention are the end results.

The definitions of intent provided by the major standards bodies are displayed in Table I, and their areas of emphasis differ significantly. These groupings provide several ways to understand intent:

- GSMA [13]: Intents are high-level, outcome-oriented statements that use AI and machine learning to guide network configurations and actions, simplifying network management and automation.
- ETSI [14]: Intent is a set of expectations that directs network management and orchestration, moving the focus from particular configurations to overarching network aims and objectives.
- IETF [15]: Intent is a set of declaratively specified operational objectives that enhances efficiency, scalability, and adaptability by enabling autonomous network configurations that correspond with operator intents.
- TM Forum [16]: Intent concentrates on system goals and commercial objectives without going into detail about how it will be achieved.
- 3GPP [17]: Intent is a high-level description of system goals that can be decomposed into network management rules and actions. It defines expectations, requirements, and restrictions.

TABLE I
Comparison of Intent Characteristics across Standards Organizations [12]

Characteristic	GSMA	ETSI	IETF	TMF	3GPP
Declarative interface	✓	✓	✓	✓	✓
Business objectives	×	△	×	✓	△
Tech implementation	✓	✓	✓	×	✓
Customer/supplier	×	✓	×	✓	✓
Automation support	✓	✓	✓	×	✓
AI/ML integration	✓	△	△	×	×
Policy decomposition	×	✓	✓	×	✓
5G-specific context	×	×	×	×	△
Legend: ✓=Full, △=Partial, ×=None					

III. Intent Forecasting

Future values or actions are predicted using data from the past, present, and present. Among the many ways it helps with decision-making is intent-based networking. Methods range from more traditional linear methods like exponential smoothing, ARMA, and ARIMA [18] to more complex nonlinear methods like recurrent neural networks (RNNs), Transformers, and diffusion models. By examining telemetry and operational patterns, these models anticipate user objectives and network requirements before they are formally stated [2]. By spotting trends and autocorrelations in performance metrics, linear models manage both stationary and nonstationary data, whereas deep learning algorithms deal with nonlinear patterns of intent evolution. Better IoT service orchestration [3] or dynamic 5G network slice modification based on anticipated QoS [17], [13] are examples of proactive resource management made feasible by this. Intent decoupling enhances this approach by separating high-level goals from implementation details [2]. Intent-predictive closed-loop verification systems reduce setup errors and improve network agility. Despite progress, full integration of predictive models into standardized intent frameworks is still being developed, according to GSMA and ETSI [13], [14].

IV. The ARMA Family of Models

A. A Unified Overview

Time series forecasting is challenging. No single model works well in every circumstance. Real data is noisy, nonlinear, and susceptible to external effects. Forecasts are skewed by seasonality, trends shift, and patterns appear. The ARMA (AutoRegressive Moving Average) family provides a trustworthy baseline [4], [5], [6], [7], [8], [9], [10], [11]. These models employ both the autoregressive and moving average components to identify trends in time-dependent data. They are easy to understand, have solid mathematical underpinnings, and are widely used in numerous domains. They are usually easy to code and have respectable baseline performance. ARMA-based models provide structure and flexibility for anticipating future behavior, action, or system demand in intent forecasting. When there is predictive information in past behavior, they work best. By extending them, you can adapt to increasingly complex data patterns. The structural elements of the main ARMA-family models—autoregression (AR), moving average (MA), integration (I), seasonality (S), and exogenous variables (X)—are compared in Table II.

Each element is associated a specific function:

TABLE II
MODEL/COMPONENT STRUCTURE USING SLASHBOX

Model \ Component	MA	AR	I	Exog. var.	Seas.
ARMA	✓	✓			
ARMAX	✓	✓		✓	
ARIMA	✓	✓	✓		
ARIMAX	✓	✓	✓	✓	
SARIMA	✓	✓	✓		✓
SARMAX	✓	✓	✓	✓	✓

- The link between the present value and its past values is described by the Autoregressive (AR) component. It assumes that the present value is a linear combination of previous values and takes into account temporal relations in the data.
- The Moving Average (MA) term accounts for the relationship between the forecast errors for the current point and the previous point. The accuracy of the model is improved by fixing its forecast errors using prior forecast errors.
- The Integrated (I) word refers to differentiating the time series in order to remove trends and make the series stationary. The differencing process will ensure that the time series have consistent statistical properties, such as a constant mean and variance, which are necessary for stable modeling.
- Periodic cycles or recurring patterns in the data are estimated by the seasonal component. It possesses:
  - P: Seasonal autoregressive term count.
  - D: The quantity of seasonal variations.
  - Q: The quantity of words in the seasonal moving average.

  The model can incorporate periodic fluctuation at predetermined intervals thanks to these factors.
- Exogenous variables are external factors that may have an impact on the time series. By supplying context-relevant information, exogenous factors—such as weather, economic trends, or other variables that impact the time series—are added to models like SARIMAX to improve forecast precision.

We use the SARIMAX model [4], [5], which includes all of the key elements of the ARMA family. It combines exogenous variables that impact the target series and captures autoregressive and moving average dynamics by employing differencing to account for both seasonal and non-seasonal trends. It is therefore a thorough model framework suitable for handling complex, real-world time series data. Its flexibility and comprehensive structure allow for a better comprehension and prediction of intent by simultaneously integrating internal patterns and external information. While recent hybrid frameworks, like Cognitive Digital Twins (CDT), which integrate CNN-LSTM and transformers, have shown good forecasting performance in power systems [19], SARIMAX provides a lighter and easier-to-understand alternative that is appropriate for real-time intent prediction with little historical data.

B. *Comparative Analysis of ARMA-Family Models*

Six ARMA-family models—ARMA, ARIMA, SARIMA, ARMAX, ARIMAX, and SARIMAX—were thoroughly evaluated under the same conditions utilizing 60-minute prediction windows and training on 8 days of historical data with 5-minute prediction steps. With the lowest MAE (0.0107 GHz) and RMSE (0.0223 GHz), ARMAX performed the best overall in terms of accuracy. SARIMAX came in second in MAE (0.0163 GHz) and third in RMSE (0.0272 GHz). SARIMAX displayed the highest MAPE (451.46%), suggesting possible overfitting or sensitivity to seasonal components in this dataset, while SARIMA recorded the lowest MAPE (152.66%).

Due to the expense of estimating seasonal factors, SARIMA and SARIMAX took significantly longer (323.1 and 302.4 seconds), but ARIMAX finished forecasts the quickest (24.0 seconds). With a moderate computation time (40.1 seconds) and great precision, ARMAX provided a good compromise. ARMAX's better performance implies that exogenous variables—like CPU and memory metrics—offer greater predictive value for short-term forecasting than seasonal adjustments or differencing. However, because SARIMAX's combined seasonal and exogenous modeling capabilities enable it to efficiently capture both internal trends and external impacts, it continues to be useful for longer-term predictions or datasets with distinct periodic patterns.

The comparative performance of the six models is shown in Table III. While SARIMA exhibits low MAPE but more inaccuracy in other metrics, ARMAX consistently exhibits the best MAE and RMSE. This demonstrates that evaluating overall model performance requires taking into account a variety of measures, such as MAE and RMSE, while MAPE alone may be deceptive. SARIMAX was chosen as the main model despite ARMAX's superior performance in this setup due to its extensive ARMA-family structure, which offers flexibility to capture a variety of intent patterns. For longer-term projections and recurring traffic patterns, its seasonal components are very helpful. The ARMAX results highlight the usefulness of SARIMAX's capacity to integrate external data in addition to internal dynamics, underscoring the significance of exogenous variables.

TABLE III
PERFORMANCE COMPARISON OF ARMA-FAMILY MODELS FOR INTENT PREDICTION

Model	MAE (GHz)	RMSE (GHz)	MAPE (%)	Time (s)
ARMA	0.0197	0.0309	237.43	63.6
ARIMA	0.0224	0.0346	191.22	43.5
SARIMA	0.0313	0.0752	152.66	323.1
ARMAX	**0.0107**	**0.0223**	278.49	40.1
ARIMAX	0.0158	0.0266	424.22	24.0
SARIMAX	0.0163	0.0272	451.46	302.4

Note: The target variable for all models is predicted bandwidth allocation, measured in GHz.

V. DATA PREPARATION

A. *Baseline Dataset*

We keep the baseline dataset from our previous conference publication [12] to guarantee reproducibility and facilitate direct comparison. To reproduce actual patterns in service-oriented networks, we developed a dataset of 8,353 synthetic network intents, each formatted in accordance with the NILE (Network Intent Language) specification [20]. The dataset, which includes measures of CPU usage, memory pressure, security alerts, and traffic fluctuations gathered across various services and time periods, is generated from Google's publicly accessible operational traces and workload analyses. These operational logs offer a practical foundation for simulating service needs and network behavior, even in the absence of a standardized dataset for network intentions.

We transformed the unstructured data into structured NILE intents by employing mapping methods that translate usage patterns into network objectives. For example, observed CPU or traffic surges were mapped to requests for more bandwidth allowance, and scheduled tasks or service tier changes were translated into changes or deletions of network slices. Each synthesized intent was designed to simulate a practical management activity that an orchestrator or network operator may do in a dynamic, intent-based network.

Each intent includes the following structured fields:

- Unique identifier and action type: allocation, modification, or deletion
- Target network resources: RAN, Core, Transport, or other domains
- Operational parameters: memory allocation, bandwidth requirements, execution time windows, and security levels

As NILE [20] is a specialized, high-level declarative language, used to convey network intents, it serves as a semantic bridge between machine-executable settings and human-readable service needs. Users can express operational goals like bandwidth constraints, security requirements, or service chaining without having to explain low-level details due to its well-organized design.

NILE's main features include:

- Human readability: Easy for network operators and non-programmers to understand
- Formal grammar enforcement: Ensures syntactic and semantic correctness
- Abstract expressiveness: Supports a wide range of intent types such as QoS, policy routing, and service chaining
- Execution-ready structure: Can be compiled into actionable configurations for programmable infrastructure
- Interoperability: Compatible with diverse network technologies and platforms

"Add firewall and intrusion detection from the gateway to the backend for client B with at least 100 Mbps of bandwidth, and allow HTTPS only" [20], can be expressed in NILE as show in Fig. 1.

Forecasting systems can comprehend evolving service requirements thanks to this clear, high-level terminology. All created records in our dataset adhere to this structure and terminology, ensuring accurate purpose encoding for SARIMAX predictive modeling.

The primary target variable for our forecasting models is the predicted bandwidth requirement for each intent, measured in Gigahertz (GHz). This metric serves as a key indicator of network resource consumption, derived from the traffic patterns in the operational traces.

```
define intent qosIntent:
    from endpoint('gateway')
    to endpoint('database')
    for group('B')
    add middlebox('firewall'),
        middlebox('ids')
    set bandwidth('min', '100', 'mbps')
    allow traffic('https')
```

Fig. 1. An example of intent with NILE.

The mapping from operational traces to structured NILE intents was governed by a set of explicit, reproducible rules:

- *CPU Threshold:* A CPU usage spike above 80% sustained for more than 5 minutes was mapped to a NILE intent to "increase bandwidth" for the affected service by 50 Mbps.
- *Scheduled Jobs:* A scheduled batch job identified from workload logs was mapped to a NILE intent to "allocate" a new network slice with specific QoS parameters for the job's duration.
- *Memory Pressure:* A gradual increase in memory pressure exceeding 20% over 30 minutes was mapped to an intent to "modify" an existing slice's resource allocation.
- *Traffic Surges:* A sudden increase in packet rate beyond 2 standard deviations from the mean was mapped to an intent requesting "additional bandwidth" for the affected service chain.
- *Security Alerts:* The presence of DDoS attack signatures in the IWIB-5GNET dataset triggered intents to "add firewall" and "enable traffic filtering" middleboxes.

B. Extended Realistic Dataset

The iwib5gnet_v1.csv dataset [21] is provided by the IWIB-5GNET (Infrastructure-Wide and Intent-Based Networking Dataset) [21], [22] to facilitate research on network automation, intent-based management, and security in 5G and beyond networks. 1,000 instances with 72 features encompassing various network layers, such as device, interface, flow, and control layers, are included in this CSV.

Among the features are:

- Packet counts, transmitted and received bytes, traffic control rules engaged, and inter-packet statistics are examples of numerical metrics.
- Boolean indicators that show when Open vSwitch (OVS), Traffic Control (TC), and iptables are activated
- Information on experimental parameters and system setups

Three important characteristics were varied in 120 controlled tests that produced the dataset:

- 2–16 user devices and 1-2 edge nodes make up the network topology.

- 32–1024 bytes make up a packet (includes 34, 40, 41, 43, 44, 128, 131, 132, 251, 256, 489, 511, 512, 960, 1022, 1023, 1024).
- 50–100 packets per second per user is the packet rate.

Distributed Denial of Service (DDoS) attacks based on UDP were purposefully included in a number of tests to mimic malicious traffic. Every record records network control replies, traffic performance, and both typical and attack-related activity.

Its design makes it possible to analyze network performance under pressure, identify anomalies, and assess autonomous, intent-driven 5G network settings in great detail. Although it lacks precise NILE intent representations, the dataset shows the results of purpose-driven rules.

For the extended experiments using the IWIB-5GNET dataset, the target variables are raw packet counts and byte counts. These metrics are derived directly from network traffic and are presented in their original, unscaled form. Consequently, the error metrics (MAE, RMSE) for these experiments appear as large absolute numbers, reflecting the raw volume of network data.

VI. Experimental Results

A. Baseline Experiments

In this subsection, we perform a systematic search to identify the optimal SARIMAX parameters (p, d, q)(P, D, Q, s) for each prediction window. This approach allows us to evaluate the best possible performance for each temporal granularity.

SARIMAX is an extension of the ARIMA model that accounts for external factors (exogenous variables) that could affect the dependent variable while capturing both seasonal and non-seasonal trends in time series data. The SARIMAX model is particularly useful when there are clear seasonal trends or when external factors can improve prediction accuracy.

Figures 2, 3, and 4 demonstrate how the SARIMAX model functions differently over time, with distinct advantages and disadvantages associated with each configuration. The 5-minute interval (p=1, d=1) in Figure 2 shows remarkable sensitivity to abrupt changes, making it ideal for real-time anomaly identification. However, due to visible overfitting to short-term noise, its forecasts are unstable for consistent predictions, as demonstrated by dense residual patterns. The 15-minute period (p=3, d=3) in figure 4, on the other hand, produces smoother forecasts suitable for long-term trends; but, systematic residual errors and significant latency during rapid demand fluctuations limit its effectiveness for dynamic network improvements. The 10-minute period (p=2, d=2) in figure 3 is the best-performing configuration because it strikes the perfect mix between responsiveness and stability. This model effectively filters out minor changes while accurately capturing underlying patterns, as evidenced by its well-distributed, almost negligible residuals. While the p=2 autoregressive order sufficiently forecasts short-term trends without being unduly complicated, the d=2 differencing provides the best trend stability. Statistical measurements, which show lower error rates than those of other intervals, validate its superior performance. Most importantly, this configuration is perfect for applications that require precise forecasting timeframes of 10 to 15 minutes, such as IoT service orchestration and 5G slice management. Additionally, it perfectly fits intent-based networking's operational requirements. Because of its balanced performance, which outperforms the lagging 15-minute version and the too sensitive 5-minute model in all significant evaluation criteria, it is the best choice for general-purpose network forecasting.

The SARIMAX prediction errors for various time intervals are displayed in Figure 5. With a minimum Mean Absolute Error (MAE) of 25,331, Root Mean Square Error (RMSE) of 31,397, and Mean Absolute Percentage Error (MAPE) of 5.92%, the model exhibits peak performance during the 10-minute period. The errors increase dramatically as the predicting window is extended, especially at 25 minutes, when the RMSE is 1,581,841,425 and the MAE is 962,584,663. To avoid confusion about the forecasting delay, the last column has been removed.

Although SARIMAX permits exogenous inputs, the current configuration mostly employs metrics like CPU and RAM. In later study, we plan to distinguish and evaluate the unique impacts of specific external signals (e.g., CPU usage or network congestion) on intent prediction. This will increase the responsiveness of the model and measure its predictive power.

B. Extended Experiments on Realistic Data

To isolate the effect of temporal aggregation and training data volume, we fix the SARIMAX parameters to (p=1, d=1)(P=0, D=0, Q=0, s=0) for all experiments in this section. This controlled setup ensures that any performance differences can be attributed directly to the prediction window size rather than to variations in model complexity.

By adding seasonal effects and exogenous variables, SARIMAX expands on the ARIMA model, making it appropriate for forecasting time series impacted by outside variables like resource usage or network measurements. We performed two tests with the same SARIMAX settings (p=1, d=1) but different training periods and prediction window sizes to assess the model's sensitivity to training duration and temporal aggregation.

1) First Experiment: Using eight days' worth of historical data, we first trained the SARIMAX model to forecast the next day over a variety of temporal windows (5, 10, 15, 20, and 25 minutes). This setup investigates granularity effects while testing the model's resilience with little training data.

Different behavioral patterns across window sizes can be seen by visually examining the forecast charts. The 5-minute window (Fig. 6) is appropriate for real-time anomaly detection due to its great responsiveness to short-term variations. Its residual plots, on the other hand, reveal dense, non-random patterns that suggest sensitivity to transitory noise. While maintaining a respectable level of responsiveness, the 10-minute window (Fig. 7) exhibits modest smoothing. Increased lag during traffic transitions is seen in the 15-minute window (Fig. 8). At the expense of responsiveness, the 20-minute

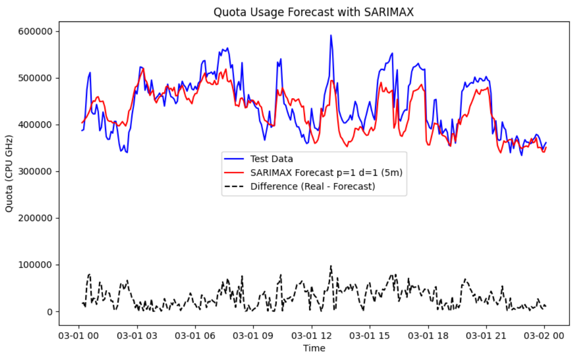

Fig. 2. SARIMAX prediction in 5 min. The bottom subplot shows the *absolute difference (|Real - Forecast|)* between the actual and predicted values.

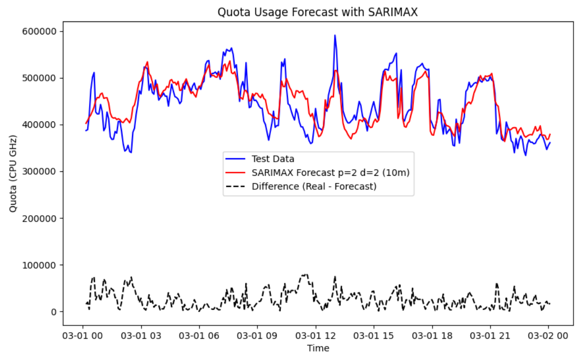

Fig. 3. SARIMAX prediction in 10 min. The bottom subplot shows the *absolute difference (|Real - Forecast|)* between the actual and predicted values.

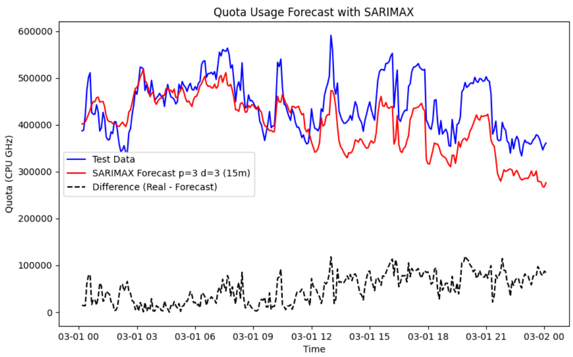

Fig. 4. SARIMAX prediction in 15 min. The bottom subplot shows the *absolute difference (|Real - Forecast|)* between the actual and predicted values.

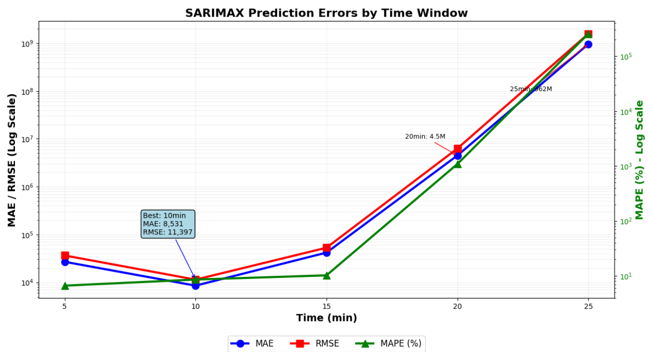

Fig. 5. SARIMAX prediction errors by time window.

window (Fig. 9) offers more stability. On the other hand, the 25-minute window (Fig. 10) yields the smoothest projections, although during sudden changes in traffic, there is a noticeable systematic lag.

The error metrics for each window size with consistent model parameters are summarized in Fig. 11. The findings demonstrate a distinct trade-off between mistake kinds at different temporal granularities. The best average prediction accuracy for short-term changes is demonstrated by the 5-minute window, which has the lowest MAE (6.02M). The 25-minute window, on the other hand, produces the lowest RMSE (17.26M), indicating that longer aggregation periods improve variance handling and outlier mitigation. The model's decreased percentage accuracy as forecasting granularity coarsens is reflected in the MAPE values, which rise steadily with window size from 299.39% at 5 minutes to 316.24% at 25 minutes.

These results demonstrate a basic trade-off between granularity and stability: longer windows (20–25 minutes) offer smoother, more stable forecasts at the cost of lag during abrupt transitions, while shorter windows (5–10 minutes) offer greater responsiveness to immediate network changes but are vulnerable to noise. The observed performance variations verify that temporal aggregation, not model complexity, is the main factor influencing forecast behavior in intent-based networking situations, even though the same SARIMAX parameters (p=1, d=1) were used for all windows.

2) Second Experiment: To evaluate the effect of more historical data, we then increased the training period to 10 days while keeping the same prediction horizon and window sizes.

Different behavioral patterns across window sizes can be seen by visually examining the forecast charts. The 5-minute window (Fig. 12) is appropriate for real-time anomaly detection due to its great responsiveness to short-term variations. Its residual plots, on the other hand, reveal dense, non-random patterns that suggest sensitivity to transitory noise. While maintaining a respectable level of responsiveness, the 10-minute window (Fig. 13) exhibits modest smoothing. Increased lag during traffic transitions is seen in the 15-minute window (Fig. 14). At the expense of responsiveness, the 20-minute window (Fig. 15) offers more stability. On the other hand, the forecasts produced by the 25-minute window (Fig. 16) are the smoothest, although there is a noticeable systematic lag during sudden changes in traffic.

The error metrics for each window size with consistent model parameters are summarized in Fig. 17. The findings show that window size and forecast accuracy have a non-linear relationship. The best average prediction accuracy for short-term changes is demonstrated by the 5-minute window, which

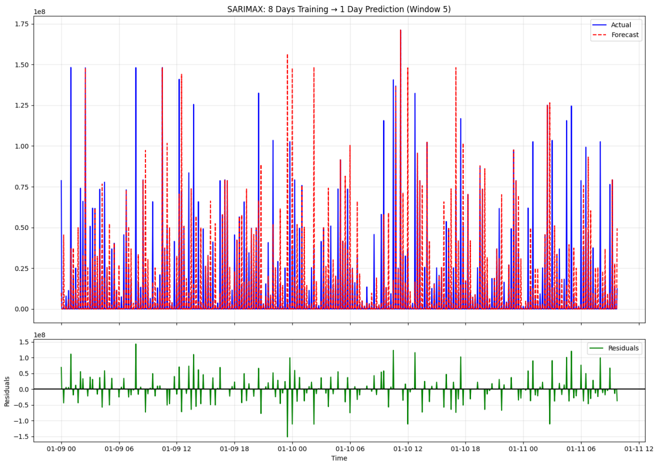

Fig. 6. SARIMAX prediction in 5 min.

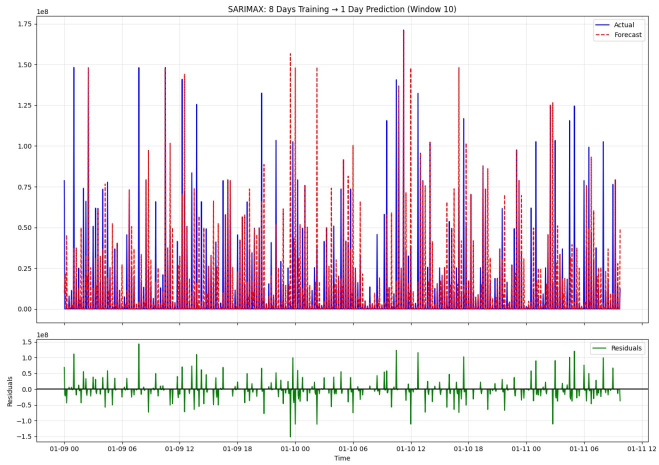

Fig. 7. SARIMAX prediction in 10 min.

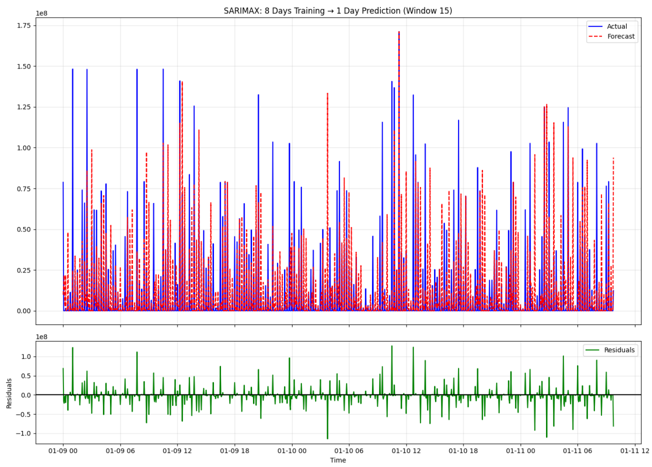

Fig. 8. SARIMAX prediction in 15 min.

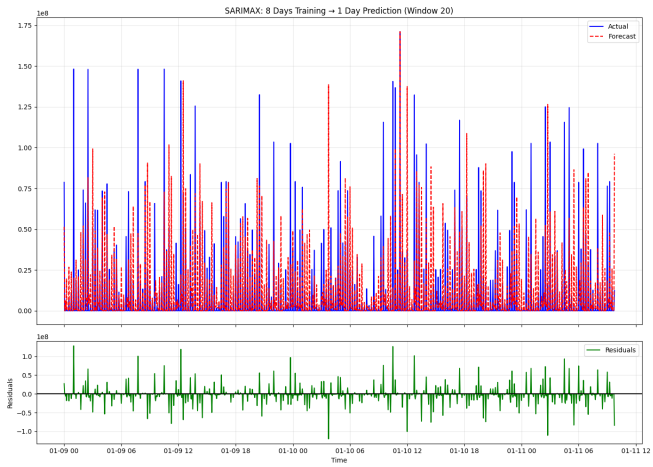

Fig. 9. SARIMAX prediction in 20 min.

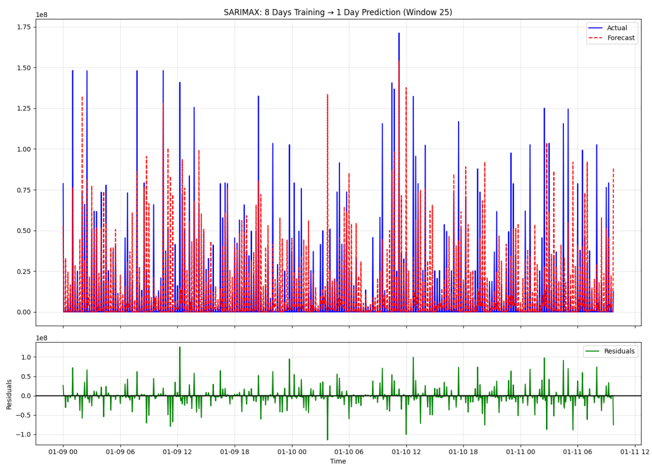

Fig. 10. SARIMAX prediction in 25 min.

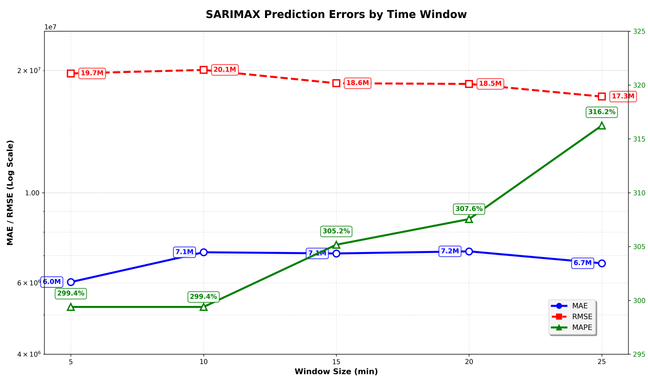

Fig. 11. SARIMAX prediction errors by time window.Note: The error metrics (MAE, RMSE) for this experiment are derived from raw packet/byte counts from the IWIB-5GNET dataset and are therefore presented as large absolute numbers, not in GHz.

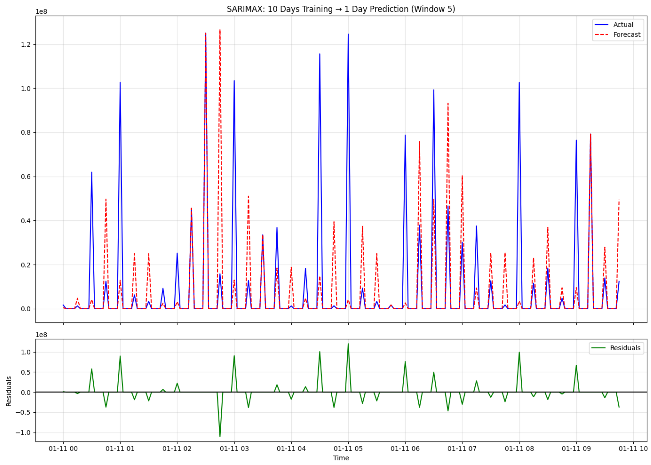

Fig. 12. SARIMAX prediction in 5 min.

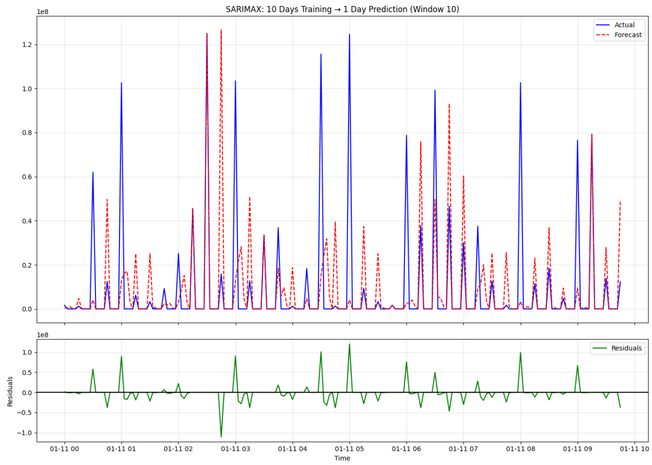

Fig. 13. SARIMAX prediction in 10 min.

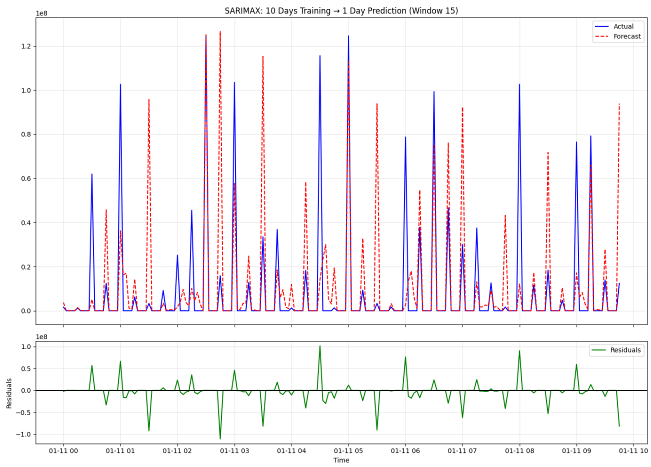

Fig. 14. SARIMAX prediction in 15 min.

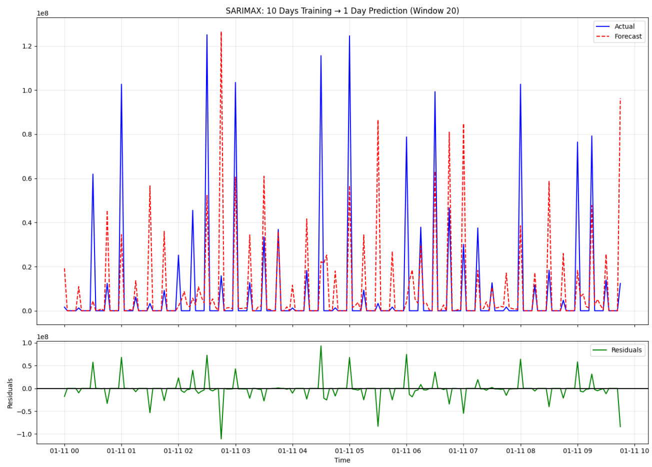

Fig. 15. SARIMAX prediction in 20 min.

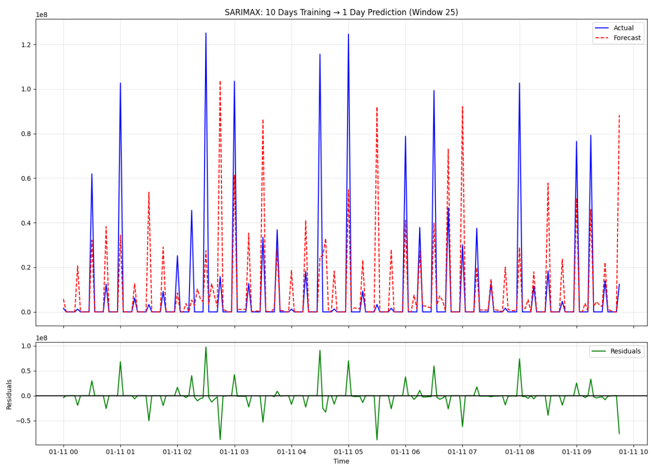

Fig. 16. SARIMAX prediction in 25 min.

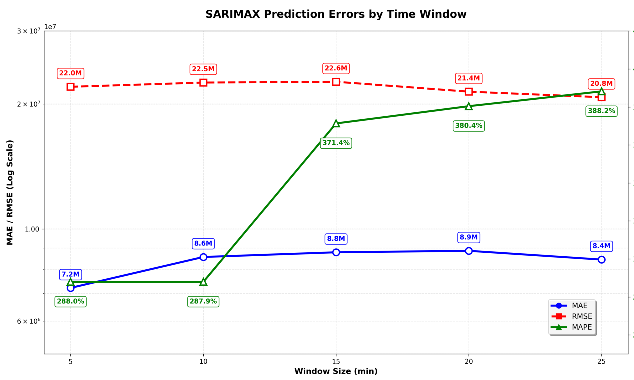

Fig. 17. SARIMAX prediction errors by time window. Note: The error metrics (MAE, RMSE) for this experiment are derived from raw packet/byte counts from the IWIB-5GNET dataset and are therefore presented as large absolute numbers, not in GHz.

has the lowest MAE (7.21M). The 25-minute window, on the other hand, produces the lowest RMSE (20.77M), indicating that longer aggregation periods handle volatility and extreme values better. The model's decreasing percentage accuracy as the forecast horizon lengthens is reflected in MAPE values, which rise significantly with window size from 287.96% at 5 minutes to 388.22% at 25 minutes.

These results point to a trade-off between stability and granularity: longer windows (20–25 minutes) smooth out volatility but lose sensitivity to abrupt changes, whereas narrower windows (5–10 minutes) are more sensitive to sudden changes but may overfit to noise. The observed performance variations demonstrate that temporal granularity has a considerable impact on forecast behavior in intent-based networking scenarios, even when the same SARIMAX parameters (p=1, d=1) are used for all windows.

VII. Conclusion

The application of SARIMAX for network intention forecasting across various temporal granularities has been thoroughly examined in this research. We showed that SARIMAX, with constant parameters (p=1, d=1), can successfully model intent-driven time series under different training periods and prediction windows through extensive testing on both synthetic and realistic datasets.

Forecast stability and temporal granularity are clearly traded off, according to the analysis. While shorter prediction intervals, like five minutes, offer greater responsiveness and are more appropriate for anomaly detection and real-time control, longer intervals, like twenty-five minutes, produce smoother predictions that help capacity planning and long-term trend analysis. Despite consistently high MAPE values brought on by data volatility, the 10-minute window strikes a reasonable compromise between responsiveness and stability, making it appropriate for operational 5G and IoT situations that call for short-term intent anticipation.

The findings also demonstrate that training data volume has less of an effect on prediction accuracy than temporal aggregation. Similar error levels during training periods of 8 and 10 days show that SARIMAX is still reliable in settings with little data. A logical way to deal with nonlinear intent dynamics is to expand the framework using hybrid statistical and learning-based models, building on this resilience. Deeper assessment of prediction accuracy in extremely dynamic network settings will be possible by integrating sequence models like LSTM and GRU, ensemble techniques, and supervised learners like support vector machines, random forests, and decision trees under identical datasets and metrics.

VIII. Acknowledgments

This work is supported by French government funding within the France 2030 framework through the INFLUENCE project.

References

[1] A. Leivadeas and M. Falkner, "A survey on intent-based networking," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 1, pp. 625–655, 2022.

[2] L. Pang, C. Yang, D. Chen, Y. Song, and M. Guizani, "A survey on intent-driven networks," *IEEE Access*, vol. 8, pp. 22 862–22 873, 2020.

[3] E. Zeydan and Y. Turk, "Recent advances in intent-based networking: A survey," in *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*. IEEE, 2020, pp. 1–5.

[4] A. Ampountolas, "Modeling and forecasting daily hotel demand: A comparison based on sarimax, neural networks, and garch models," *Forecasting*, vol. 3, no. 3, pp. 580–595, 2021.

[5] B. Artley, "Time series forecasting with arima, sarima and sarimax," *Towards Data Science*, 2022.

[6] N. Deretić, D. Stanimirović, M. A. Awadh, N. Vujanović, and A. Djukić, "Sarima modelling approach for forecasting of traffic accidents," *Sustainability*, vol. 14, no. 8, p. 4403, 2022.

[7] R. Milocco, P. Minet, E. Renault, and S. Boumerdassi, "Evaluating the upper bound of energy cost saving by proactive data center management," *IEEE Transactions on Network and Service Management*, vol. 17, no. 3, pp. 1527–1541, 2020.

[8] R. Milocco, P. Minet, É. Renault, and S. Boumerdassi, "Proactive data center management using predictive approaches," *IEEE Access*, vol. 8, pp. 161 776–161 786, 2020.

[9] R. Milocco, P. Minet, E. Renault, and S. Boumerdassi, "An evaluation method of optimal cost saving in a data center with proactive management," *Management of Data Center Networks*, pp. 105–127, 2021.

[10] ——, "Cost reduction bounds of proactive management based on request prediction," in *2019 International Conference on High Performance Computing & Simulation (HPCS)*. IEEE, 2019, pp. 864–871.

[11] B. M. Williams, "Multivariate vehicular traffic flow prediction: evaluation of arimax modeling," *Transportation Research Record*, vol. 1776, no. 1, pp. 194–200, 2001.

[12] N. Hachem, M. C. Nguyen, and É. Renault, "Time-series forecasting with sarimax for intent prediction," in *2025 International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*. IEEE, 2025, pp. 1–6.

[13] "Intelligence brief: Does intent matter in network automation?" https://www.gsmaintelligence.com/media-coverage/intelligence-brief-does-intent-matter-in-network-automation/.

[14] ETSI, "Zero-touch network and service management (zsm); intent-driven autonomous networks; generic aspects," European Telecommunications Standards Institute, Tech. Rep. GR ZSM 011 V2.1.1, Sep. 2024. [Online]. Available: https://www.etsi.org/deliver/etsi_gr/ZSM/001_099/011/02.01.01_60/gr_ZSM011v020101p.pdf

[15] "ietf," https://www.ietf.org/.

[16] "tmforum," https://www.tmforum.org/.

[17] "3gpp," https://www.3gpp.org/.

[18] J. Kim, H. Kim, H. Kim, D. Lee, and S. Yoon, "A comprehensive survey of time series forecasting: Architectural diversity and open challenges," *arXiv preprint arXiv:2411.05793*, 2024.

[19] X. Wu, Z. Chen, H. Jiang, S. Luo, Y. Zhao, D. Zhao, P. Dang, J. Gao, L. Lin, and H. Wang, "From forecasting to foresight: Building an autonomous o&m brain for the new power system based on a cognitive digital twin," *Electronics*, vol. 14, no. 22, p. 4537, 2025.

[20] A. S. Jacobs, R. J. Pfitscher, R. H. Ribeiro, R. A. Ferreira, L. Z. Granville, W. Willinger, and S. G. Rao, "Hey, lumi! using natural language for {intent-based} network management," in *2021 USENIX Annual Technical Conference (USENIX ATC 21)*, 2021, pp. 625–639.

[21] J. Andrade-Hoz, J. M. Alcaraz-Calero, and Q. Wang, "IW-IB-5GNET dataset," GitHub, 2024, accessed on 1 January 2026. [Online]. Available: https://github.com/jimenaandrade/iw-ib-5gnet

[22] J. Andrade-Hoz, Q. Wang, and J. M. Alcaraz-Calero, "Infrastructure-Wide and Intent-Based Networking Dataset for 5G-and-beyond AI-Driven Autonomous Networks," *Sensors*, vol. 24, no. 3, 2024. [Online]. Available: https://www.mdpi.com/1424-8220/24/3/783

Nagham Hachem is a PhD candidate in Computer Science at ESIEE Paris and a member of LIGM (UMR CNRS 8049) at Université Gustave Eiffel, France. She received an engineering degree in Telecommunications and Computer Engineering from the Lebanese University in 2021 and a Master's degree in Multimedia Networking from Paris-Saclay University in 2022. She is currently pursuing her PhD in Computer Science within the MSTIC doctoral school. Her research interests include 5G networks, network slicing, intent-based networking, machine learning for network management, and lightweight security. She has worked on research activities related to intent modeling, anomaly detection, and resource prediction, and contributes to collaborative projects such as INFLUENCE. She authored several articles in international conferences and journals.

Manh Cuong Nguyen completed a PhD in 2015 focused on voice capacity optimization over 4G LTE networks for security services (Télécom SudParis). He gained nearly eight years of experience in industrial R&D (AXA, LTU Tech, BNP Paribas Cardif). He applies Deep Learning methods (Computer Vision, OCR, LLM/RAG) to complex, large-scale problems. He is currently a Postdoctoral Researcher at LIGM/ESIEE Paris, focusing on Intent-Based Networking (IBN).

Éric Renault is Full Professor at ESIEE Paris and a member of LIGM (UMR CNRS 8049) at Univ. Gustave Eiffel, France. He received a MSc in Computer Engineering (diplôme d'ingénieur) from ISTY and a MSc in Computer Science (DEA) from UVSQ in 1995, a PhD in Computer Science from UVSQ in 2000 and a Habilitation thesis (HDR) from UPMC in 2011. His research interests include high-performance computing and messaging, compilation, virtualization, positioning and lightweight security for mobiles, sensors and vehicular networks. He has worked on several French national and European projects and serves as an expert for the evaluation of French national projects (ANR), private company research studies (MESRI) and Eiffel Excellence Scholarship Program applications (Campus France). He authored more than 150 articles in international journals and conferences.