Visitor Flow Forecasting in Heritage Sites: A Predictive Module Integrated into a Territorial Digital Twin for Decision Support

Kimberley VIARON[1], Philippe RAUFFET[1], Nathalie JULIEN[1], Christine PETR[2‡]

Abstract

Forecasting visitor flows is a critical yet underexplored capability for decisional Territorial Digital Twins (TDT) dedicated to heritage tourism. Heritage sites typically operate under strong data constraints, high sensitivity to exogenous factors, and require forecasts at fine temporal granularity to support regulation decisions. Within the JUNITER project, dedicated to the UNESCO Megalithic Landscapes of Morbihan, this paper addresses the challenge of hourly visitor flow forecasting across multiple decision horizons. Based on a structured review of statistical, machine learning, and deep learning approaches, we identify scientific barriers specific to data-sparse heritage contexts: limited historical depth, noisy demand signals, and heterogeneous decision objectives. In response, we propose a horizon-adaptive forecasting architecture that aligns model families and feature spaces with operational (D+1), tactical (D+7), and strategic (D+90) decision-making needs. The contribution is twofold: (i) a methodologically grounded model selection strategy under data scarcity, and (ii) an explicit data and control flow specification linking forecasts to regulation actions in the broader TDT, including a pre-registered evaluation protocol. This positions forecasting as a structural component of a decisional TDT, enabling the transition from reactive monitoring to proactive and sustainable heritage site regulation.

1. INTRODUCTION

The Digital Twin (DT) concept has progressively established itself as a central approach for the modeling, analysis, and control of complex systems. Initially developed within the industrial sector [1, 2], it integrates three complementary dimensions [3]: observation, understanding through models, and anticipation of future states. This hybridization of data and models [4] makes forecasting a structural component of the DT [5, 6], enabling the transition from reactive management to proactive regulation through *ex ante* evaluation of decisions and exogenous disturbances [7].

Extending the DT to territories shifts the paradigm toward open socio-cyber-physical systems [8]. A Territorial Digital Twin (TDT) orchestrates continuous interactions between physical infrastructures, human dynamics, and digital components; its value lies not in geometric fidelity but in its ability to federate these heterogeneities to link observation with decision-making [9]. TDTs frequently operate within constrained environments characterized by the “Big Data illusion”: limited datasets, sparse sensor networks, and high variability of human behavior driven by exogenous factors [10]. In heritage contexts, noise is non-stationary and exogenous-driven (weather, events, structural breaks), invalidating industrial assumptions and motivating parsimonious, data-sparse forecasting architectures.

2. The JUNITER Project and Use Case Structuring

The JUNITER project develops a TDT dedicated to the Megalithic Landscapes of Morbihan, recently inscribed on the UNESCO World Heritage List. The territory must reconcile growing tourism attractiveness with the preservation of sensitive archaeological sites. The TDT is structured around four complementary Use Cases (UC) covering the entire data value chain: observation through sensor deployment optimization (UC1) and near-real-time pressure monitoring (UC2); anticipation through hourly visitor flow forecasting (UC3); and dynamic optimization with prescription of corrective or preventive actions (UC4).

UC3 occupies a pivotal position, ensuring the transition from a descriptive twin (monitoring) to a decisional twin (proactive control). It is explicitly designed under

*[1]University of South Brittany, Lab-STICC (UMR CNRS 6285), Lorient, France

†[2]University of South Brittany, Laboratoire LEGO, Vannes, France

‡Corresponding author: kimberley.viaron@univ-ubs.fr

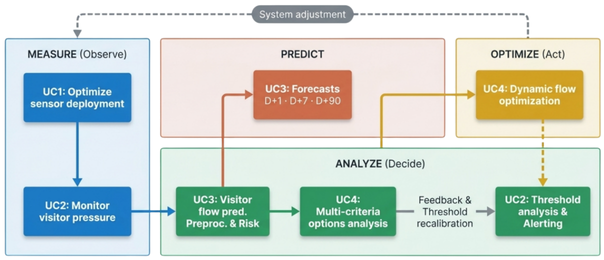

Figure 1. Functional architecture of the JUNITER Digital Twin tailored to the Use Cases (UC).

strong data constraints: sparse historical attendance series, calendar variables (school holidays, events), short-horizon meteorological forecasts, and contextual exogenous signals whose predictive relevance varies across time horizons. Forecasting choices are therefore data-driven, and the nature, reliability, and temporal validity of each source justify horizon-specific prediction strategies.

Preliminary empirical characterization. Approximately 25 months of hourly eco-counter data (July 2022–August 2024) from three instrumented pilot sites (Pen Hap, Île-aux-Moines; Kernous, Le Bono; Kermaillard, Sarzeau) were analyzed to characterize the operational regime. Aggregated to daily granularity, the series exhibit pronounced concentration: Gini coefficients range from 0.43 to 0.71, June–August captures 47–54% of annual visits, and peak days (98th percentile) reach 8 to 89 times the off-season median depending on the site. Saturation-alert candidates (days exceeding twice the site mean) account for 7–17% of observed days. Analysis of the merged eco-counter and weather records further confirms the exogenous sensitivity of the series: daily visitor counts correlate positively with maximum temperature up to 25°C and negatively above this thermal optimum, a non-linear bioclimatic relationship that justifies the Heat Index as a composite exogenous feature rather than raw temperature alone. This regime, characterized by short history, strong seasonality, extreme exogenous-driven peaks and sparse instrumentation, matches the data-sparse heritage context for which UC3 is designed and motivates the horizon-adaptive architecture detailed below.

3. Visitor Flow Predictive Methods: State of the Art and Scientific Barriers

This section reviews forecasting methodologies applicable to heritage visitor flows. The goal is not to declare a single "best" model, but to understand the contextual strengths and limitations of each methodological family, essential for context-specific model selection in UC3.

3.1. The Statistical Approach: Classical Time Series Models

Grounded in Box and Jenkins [11], Seasonal AutoRegressive Integrated Moving Average (SARIMA) models have long served as a reference for univariate time series exhibiting strong seasonal patterns, with empirical validation in tourism applications [12]. Their popularity stems from parametric simplicity, robustness under limited data, and analytical interpretability through trend and seasonality components. However, their linear and largely univariate formulation limits their effectiveness in contexts characterized by abrupt structural breaks and strong exogenous dependence (e.g., post-pandemic regime shifts). They suit stable seasonal dynamics and long-term baselines, but exhibit reduced performance for short-term operational regulation in data-sparse, exogenous-driven heritage environments.

3.2. Machine Learning: Tree-Based Ensembles for Tabular Time Series

Gradient boosting algorithms such as eXtreme Gradient Boosting (XGBoost) and Light Gradient Boosting Machine (LightGBM) have emerged as state-of-the-art methods for structured tabular data, capturing non-linear interactions between heterogeneous variables without strong distributional assumptions. Their effectiveness on noisy demand series was confirmed in the M5 forecasting competition, where tree-based ensembles outperformed both classical statistics and deep learning on exogenous-driven demand forecasting [13]. By contrast, their reliance on data-space partitioning makes them poorly suited for extrapolating trends beyond the training range, limiting their use for long-term strategic forecasting.

3.3. Deep Learning: From Recurrent Networks to Transformers

Deep Learning architectures learn hierarchical temporal representations. Long Short-Term Memory (LSTM) networks [14] solved the vanishing gradient problem through gating mechanisms. Hybrid neural-statistical approaches extend this paradigm: feedforward Artificial Neural Networks (ANN) coupled with non-linear least-squares regression for trend extraction [15] and seasonal autoregressive ANN architectures [16] demonstrate that careful hybridization extends the operational range of neural approaches under irregular and seasonal conditions, an observation directly relevant to heritage demand. The Temporal Fusion Transformer (TFT) further advances this paradigm with attention and variable-selection networks for interpretable multi-horizon forecasts. However, with limited historical observations, these models frequently underperform simpler statistical baselines.

3.4. State-of-the-Art Architectures and Emerging Paradigms

The 2023–2025 literature has seen the emergence of architectures prioritizing efficiency over sheer parameter count: patch-based Transformers (PatchTST) enable attention over longer historical windows [17], while N-BEATS and TimeMixer demonstrate that purely Multi-Layer Perceptron (MLP) architectures can match Transformers [18, 19].

Attention-based Spatio-Temporal Graph Convolutional Networks (ASTGCN) [20] explicitly model spatial structure as graphs whose nodes correspond to locations and edges encode connectivity. Both classes achieve their best results on large datasets with smooth long-range dependencies or dense sensor networks, conditions rarely satisfied by heritage sites.

3.5. Comparative Synthesis and Conditional Performance

A cross-analysis of M4/M5 competitions and recent surveys confirms that no single family is universally dominant; performance is conditional on data regime, temporal granularity, and the relative importance of exogenous variables [13]. SARIMA and Exponential Smoothing State Space (ETS) remain competitive baselines for short univariate series with stable seasonal patterns [21]; gradient boosting dominates noisy heterogeneous tabular data [13]; patch-based Transformers and TSMixer-type MLPs dominate large-scale continuous signals with long-range dependencies [17, 19]; spatio-temporal GNNs require dense sensor networks [20]. Figure 2 summarizes the taxonomy of forecasting methods examined.

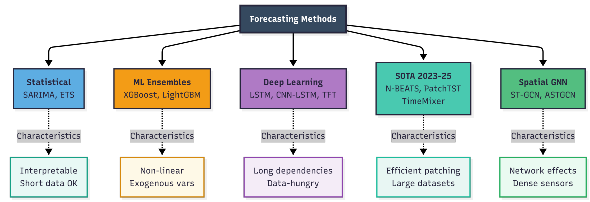

Figure 2. Taxonomy of time series forecasting methods organized by methodological family with key characteristics and representative models.

3.6. Synthesis and Methodological Positioning for UC3

For the JUNITER UC3 module, four findings guide our architectural choices.

- **F1: Data scarcity constrains high-parameter architectures.** Under restricted historical data, SARIMA achieves a Mean Absolute Percentage Error (MAPE) of 3.81% vs. 6.81% for LSTM [22]. The gap is attributable to overfitting in low-data regimes, also confirmed by Lionetti et al. [23].
- **F2: Hourly granularity demands operational-level precision.** Most tourism research operates at daily/monthly aggregation; no major study has examined hourly visitor flows in protected areas, yet hourly forecasts are indispensable to prevent parking saturation and manage intra-day peaks.
- **F3: Exogenous variable dominance favors tabular ML.** Heritage flows depend strongly on bioclimatic indices (Heat Index), calendar effects, and Google Trends signals, aligning with the regime where gradient boosting consistently outperforms deep learning [13].
- **F4: Multi-horizon heterogeneity requires adaptive architecture.** D+1, D+7 and D+90 serve distinct decisions; high-frequency exogenous signals lose skill beyond 7–10 days while strategic forecasts rely on stable seasonal patterns. This invalidates the universal-model assumption.

4. Proposed Methodology: Horizon-Adaptive Forecasting Architecture for UC3

This section formalizes the architecture, specifies the data and optimization pipeline, the evaluation protocol, and the UC3–UC4 coupling.

4.1. Conceptual Framework: Horizon-Stratified Modeling

UC3 forecasts must be explicitly stratified by horizon. Let y_t denote the visitor count at time t and $\hat{y}_{t+h}$ the forecast issued at time t for horizon $h \in \{1, 7, 90\}$ (in days). We define horizon-specific forecasting functions

$$\hat{y}_{t+h} = \mathscr{F}_h\left(\mathscr{X}_t^h, \theta_h\right),$$

where $\mathscr{F}_h$ is a forecasting model tailored to horizon h (possibly from a different methodological family), $\mathscr{X}_t^h$ is the feature set deemed relevant for predicting $\hat{y}_{t+h}$, and θ_h are parameters estimated by minimizing a horizon-specific loss. Crucially, $\mathscr{X}_t^h$ differs across horizons: high-frequency exogenous signals for $h = 1$, calendar effects and aggregated trends for $h = 7$, structurally stable components for $h = 90$. This motivates three specialized pipelines (D+1, D+7, D+90).

4.2. Data Pipeline and Feature Engineering

The data flow is structured in three stages. The *ingestion* layer aggregates four hourly sources: (i) eco-counter and ticketing data (UC2 output), (ii) calendar features (day-of-week, holidays, public events, European Heritage Days), (iii) meteorological observations and short-range forecasts (Heat Index, hourly precipitation), and (iv) Google Trends queries on site-specific keywords. The *feature engineering* layer constructs lag features (y_{t-1}, y_{t-24}, y_{t-168}), bioclimatic transformations, and moving averages. The *model input* layer normalizes the feature matrix and routes each instance to the appropriate model.

Google Trends as a leading indicator. Empirical search-versus-visit lags in tourism consistently identify a 1–3 week lead reflecting the planning-to-execution gap. We therefore treat Google Trends as a *leading indicator* (lag features g_{t-7}, g_{t-14}, g_{t-21}) rather than a concurrent signal. This is decisive for D+1 and D+7; for D+90, the lead structure is incompatible with the horizon and Google Trends is excluded.

4.3. Horizon-Specific Forecasting Pipelines

D+1 (short-term operational). Targets parking-saturation anticipation and immediate regulation levers (dynamic signage, alert messages). Short-range weather forecasts retain skill, making Heat Index and hourly precipitation decisive. The preferred model is a *gradient boosting ensemble* (XGBoost or LightGBM), trained on (i) hourly and daily lag features, (ii) calendar variables, (iii) bioclimatic indices, and (iv) Google Trends lead features. This is consistent with the dominance of tree-based boosting on noisy heterogeneous tabular data.

D+7 (weekly tactical). Supports staffing, maintenance, and anticipatory shuttle activation. Numerical weather prediction skill degrades, while calendar effects (extended weekends, holidays) and 7-day aggregated trends dominate. A *hybrid SARIMA + XGBoost* is adopted: SARIMA captures weekly/yearly seasonality, gradient boosting corrects the baseline through residual correction using enriched calendar features. This design is methodologically continuous with hybrid neural-statistical approaches [16, 15], transposed to a tabular boosting backbone better suited to the data-sparse heritage regime.

D+90 (long-term strategic). Supports seasonal opening hours, infrastructure sizing, and preventive communication with tour operators. A *univariate SARIMA* or *Prophet* is preferred. Table 1 summarizes the architecture.

Justification of the univariate choice at D+90. Three converging arguments: (i) the skill of high-frequency exogenous variables (Numerical Weather Prediction, real-time search trends) collapses beyond two weeks, so injecting them into a 90-day model amplifies noise; (ii) in low-data regimes (2–3 years), a univariate SARIMA or Prophet fits a small number of robustly estimable structural parameters, whereas multivariate or deep architectures face identifiability issues; (iii) strategic decisions require calibrated confidence intervals provided natively by SARIMA and Prophet. Exclusion of exogenous variables at D+90 is a deliberate alignment with the information regime of the horizon, not a methodological limitation.

4.4. Optimization Framework and Evaluation Protocol

Unified hyperparameter optimization. Within each horizon, candidate models share a unified protocol ensuring methodological consistency. The objective is a horizon-specific loss: pinball loss for D+1 (calibrated quantiles), weighted Mean Absolute Error (MAE) for D+7 and D+90 (outlier robustness). Hyperparameters are searched via Bayesian optimization with Tree-structured Parzen Estimators over a constrained search space (tree depth, learning rate, regularization for boosting; orders $(p,d,q)(P,D,Q)_s$ for SARIMA) under a horizon-capped time budget. Cross-family selection uses a Diebold–Mariano test on horizon-specific losses to ensure that performance differences are statistically significant. The same protocol is applied to baselines, enabling fair comparison.

Table 1. Tri-horizon architecture of the UC3 predictive module and methodological instantiation by forecast horizon.

Horizon	Decision objective	Model family	Dominant features
D+1	Short-term operational regulation (parking, signage)	Gradient boosting (XGBoost / LightGBM)	Heat Index forecast; hourly rainfall; Google Trends (lead 7–21 d); hourly lags
D+7	Weekly tactical planning (staffing, shuttles)	Hybrid SARIMA + XGBoost (residual correction)	Calendar effects (holidays, weekends); 7-day aggregated trends; Google Trends (lead 14–21 d)
D+90	Strategic capacity planning (opening hours, infrastructure)	Univariate SARIMA or Prophet	Annual seasonality; long-term trend; known calendar events

Pre-registered evaluation protocol. Empirical validation adopts *walk-forward expanding-window cross-validation*: the model is retrained on $[t_0, t_k]$ and evaluated on $[t_k+1, t_k+h]$, with t_k advancing by one full season (3 months). Performance is assessed through (i) Root Mean Square Error (RMSE), penalizing large deviations critical for saturation prevention, (ii) MAPE for scale-invariant inter-site comparison, and (iii) Mean Absolute Scaled Error (MASE) using a seasonal naive benchmark, robust to zero-attendance periods. Directional accuracy is reported for D+1, as it conditions the operational utility of the alert system. Reference baselines (seasonal naive, SARIMA without exogenous variables, XGBoost without Google Trends, LSTM on the same feature space) isolate each architectural component's contribution.

4.5. UC3–UC4 Integration: Data and Control Flow

UC3 outputs feed UC4 through a structured data and control flow. For each monitored commune, UC3 publishes at each cycle a triplet $\left(\hat{y}_{t+h}, [\hat{y}^{\text{lo}}_{t+h}, \hat{y}^{\text{hi}}_{t+h}], h\right)$ comprising the point forecast and an 80% prediction interval. UC4 maps these to a discrete set of *regulation levels* via site-specific saturation thresholds calibrated on UC2 pressure indicators, triggering three action classes of increasing intensity: *informational* (signage updates, mobile-app notifications), *operational* (shuttle reinforcement, dynamic parking redirection), and *regulatory* (temporary access modulation, coordination with site managers).

Human-in-the-loop by design. UC4 operates with a human-in-the-loop for all operational and regulatory actions, surfaced to site managers and territorial actors (Grand Site, communal managers, Morbihan Tourisme) through a decision-support interface; only informational actions, with the lowest reversibility cost, are eligible for automation. Heritage decisions carry social, cultural, and economic stakes not fully encoded in the model; forecast uncertainty must remain visible to decision-makers; and territorial governance requires a legitimacy that fully automated regulation would erode. The architecture supports decision-augmentation rather than decision-automation.

5. Conclusion and Future Work

This paper has addressed visitor flow forecasting for heritage sites within a decisional TDT framework. We identified three scientific barriers: (i) data scarcity (2–3 years), (ii) hourly granularity for operational regulation, and (iii) heterogeneous decision objectives across horizons. We proposed a horizon-adaptive architecture (D+1, D+7, D+90) aligning model family, feature space, and decision objective at each scale, with a unified optimization framework, a pre-registered evaluation protocol (RMSE, MAPE, MASE, directional accuracy), and a human-in-the-loop UC3–UC4 coupling, contributing to the emerging field of decisional TDTs.

Limitations include distinct maintenance pipelines per horizon and empirical validation pending on JUNITER data. **Future work** targets: (i) empirical validation on three pilot sites (MAPE $<$15% at D+1, $<$25% at D+90, directional accuracy $>$80%); (ii) hierarchical reconciliation across horizons; (iii) transfer learning to data-sparse sites; and (iv) *preventive conservation* via coupling with physical-impact models of megalith wear. The eco-counter dataset (July 2022–August 2024) and evaluation protocol will be made publicly available upon publication. The horizon-stratified design is transferable to any UNESCO heritage site with basic visitor monitoring infrastructure.

ACKNOWLEDGMENT

The authors gratefully acknowledge the Paysages de Mégalithes du Morbihan association and the CRT Bretagne for their support throughout the JUNITER project, the Conseil Départemental du Morbihan for providing eco-counter data on the three pilot sites, and the anonymous reviewers for their constructive feedback that significantly improved this manuscript.

References

[1] F. Tao, J. Cheng, Q. Qi, M. Zhang, H. Zhang, and F. Sui, "Digital twin-driven product design, manufacturing and service with big data," *The International Journal of Advanced Manufacturing Technology*, vol. 94, no. 9-12, pp. 3563–3576, 2018.

[2] E. Glaessgen and D. Stargel, "The Digital Twin Paradigm for Future NASA and U.S. Air Force Vehicles," in *Proc. 53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conf.*, Honolulu, HI, USA, 2012, p. 1818.

[3] W. Kritzinger, M. Karner, G. Traar, J. Henjes, and W. Sihn, "Digital Twin in manufacturing: A categorical literature review and classification," *IFAC-PapersOnLine*, vol. 51, no. 11, pp. 1016–1022, 2018.

[4] A. Rasheed, O. San, and T. Kvamsdal, "Digital Twin: Values, Challenges and Enablers From a Modeling Perspective," *IEEE Access*, vol. 8, pp. 21980–22012, 2020.

[5] A. M. Madni, C. C. Madni, and S. D. Lucero, "Leveraging Digital Twin Technology in Model-Based Systems Engineering," *Systems*, vol. 7, no. 1, p. 7, 2019.

[6] J. Lee, B. Bagheri, and H.-A. Kao, "A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems," *Manufacturing Letters*, vol. 3, pp. 18–23, 2015.

[7] M. Grieves and J. Vickers, "Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems," in *Transdisciplinary Perspectives on Complex Systems*, Springer, pp. 85–113, 2017.

[8] L. Deren, Y. Wenbo, and Y. Shaozhen, "Smart city based on digital twins," *Computational Urban Science*, vol. 1, no. 1, pp. 1–11, 2021.

[9] M. Batty, "Digital Twins," *Environment and Planning B: Urban Analytics and City Science*, vol. 45, no. 5, pp. 817–820, 2018.

[10] K. Viaron, P. Rauffet, N. Julien, P. Caudan, and C. Petr, "From Data to Decision: A Digital Twin Supporting Sustainable Multi-Criteria Decision-Making for Heritage Territories and Their Users," in *Service Oriented, Holonic and Multi-agent Manufacturing Systems for Industry of the Future (SOHOMA)*, 2024.

[11] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day, 1970.

[12] C. Lim and M. McAleer, "Time series forecasts of international travel demand for Australia," *Tourism Management*, vol. 23, no. 4, pp. 389–396, 2002.

[13] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "The M5 Competition: Background, Organization, and Implementation," *International Journal of Forecasting*, vol. 38, no. 4, pp. 1325–1336, 2022.

[14] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[15] F. Saâdaoui, H. Saadaoui, and H. Rabbouch, "Hybrid feedforward ANN with NLS-based regression curve fitting for US air traffic forecasting," *Neural Computing and Applications*, vol. 32, pp. 10073–10085, 2020, doi: 10.1007/s00521-019-04539-5.

[16] F. Saâdaoui, "A seasonal feedforward neural network to forecast electricity prices," *Neural Computing and Applications*, vol. 28, pp. 835–847, 2017, doi: 10.1007/s00521-016-2356-y.

[17] Y. Nie, N. H. Nguyen, P. Sinthong, and J. Kalagnanam, "A Time Series is Worth 64 Words: Long-term Forecasting with Transformers," in *Proc. ICLR*, 2023.

[18] B. N. Oreshkin, D. Carpov, N. Chapados, and Y. Bengio, "N-BEATS: Neural Basis Expansion Analysis for Interpretable Time Series Forecasting," in *Proc. ICLR*, 2020.

[19] S. Wang, H. Wu, X. Shi, T. Hu, H. Luo, L. Ma, J. Y. Zhang, and J. Zhou, "TimeMixer: Decomposable Multiscale Mixing for Time Series Forecasting," in *Proc. ICLR*, 2024.

[20] S. Guo, Y. Lin, N. Feng, C. Song, and H. Wan, "Attention Based Spatial-Temporal Graph Convolutional Networks for Traffic Flow Forecasting," in *Proc. AAAI*, vol. 33, pp. 922–929, 2019.

[21] R. J. Hyndman, A. B. Koehler, K. Ord, and R. Snyder, *Forecasting with Exponential Smoothing: The State Space Approach*. Berlin, Germany: Springer, 2008.

[22] D. Y. Prasojo and K. Muludi, *Passenger Density Prediction at the Airport Using LSTM and SARIMA: A Case Study at Radin Inten Airport, Lampung*. Journal of Applied Informatics and Computing (JAIC), vol. 9, no. 4, pp. 1955–1963, 2025.

[23] S. Lionetti, D. Pfäffli, M. Pouly, T. vor der Brück, and P. Wegelin, "Tourism Forecast with Weather, Event, and Cross-industry Data," in *Proc. 13th Int. Conf. Agents and Artificial Intelligence (ICAART)*, vol. 2, pp. 1097–1104, 2021.

[24] H. Rabbouch, H. Saadaoui, and F. Saâdaoui, "VMD-based Multiscaled LSTM-ARIMA to Forecast Post-COVID-19 US Air Traffic," in *Proc. 2022 Int. Conf. on Decision Aid Sciences and Applications (DASA)*, IEEE, 2022, pp. 1678–1683.