

# Sequential and Temporal Data Analysis: Foundations, Bayesian Procedures, Modern Methodologies, and Cross-Domain Applications

## Abstract

Sequential data—spanning domains from finance and health to robotics and climate science—forms the backbone of modern statistical, computational, and operational systems. This survey presents a comprehensive synthesis of the motivation, methodologies, and evolving challenges in sequential data modeling, motivated by the field’s unprecedented expansion in scale, heterogeneity, and complexity. We systematically review foundational theoretical frameworks, encompassing classical frequentist and Bayesian paradigms, and detail advances in deep learning, generative modeling, and hybrid approaches. The discussion emphasizes robust preprocessing, adaptive data partitioning, and feature engineering strategies critical for fair and generalizable model evaluation. We analyze core modeling families, including state space models, recurrent and attention-based neural architectures, graph and hypergraph neural networks, matrix/tensor-based techniques, and compositional models, clarifying their strengths and limitations across diverse analytical tasks such as forecasting, anomaly detection, generative simulation, and adaptive experimentation.

The paper elucidates the integration of robust evaluation metrics, interpretability techniques, and ensemble-based robustness strategies, highlighting emerging needs in adversarial defense, fairness, and security. Cross-domain benchmarking, transparent evaluation protocols, and automated pipelines are identified as vital for reproducibility and translational impact. We articulate open challenges regarding scalability, explainability, standardization, and domain adaptation, offering a future outlook that advocates the fusion of classical statistical rigor with data-driven learning, automated model selection, causal reasoning, and responsible AI deployment. This survey serves as both a reference and a roadmap, equipping researchers and practitioners to address the theoretical, methodological, and societal imperatives at the frontier of sequential data analysis.

## ACM Reference Format:

. 2025. Sequential and Temporal Data Analysis: Foundations, Bayesian Procedures, Modern Methodologies, and Cross-Domain Applications. In . ACM, New York, NY, USA, 21 pages. <https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).  
*Conference’17, Washington, DC, USA*

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.  
ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM  
<https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>

## 1 Introduction

### 1.1 Overview of Sequential Data and Its Significance

Sequential data—defined as ordered collections of observations indexed by time, space, or other systematic dimensions—serves as a fundamental structure across diverse scientific, technological, and societal domains. The importance of sequential data analysis is evident in a wide spectrum of applications, including clinical trial monitoring, financial econometrics, engineering systems, environmental and climate modeling, public health surveillance, human-robot interaction, computer vision, natural language processing, scientific discovery, e-commerce, and adaptive experimentation [4, 6, 7, 9, 11–16, 19, 21, 33, 34, 36, 40–42, 47, 53, 63–66, 70, 74–76, 78, 81, 82, 86]. The pervasiveness of sequential data stems not only from its ubiquity but also from its ability to encode system dynamics, temporal dependencies, regime shifts, and causal mechanisms, making it indispensable for both scientific inference and operational decision-making.

For example, high-frequency multivariate volatility modeling in finance can uncover latent interconnections between assets and guide risk management strategies [82]; in clinical trials, real-time patient monitoring and adaptive interim analyses accelerate knowledge discovery and promote safety assurance [4, 64]. Advances in sensor technology, digitization, and automation have simultaneously amplified the scale, complexity, and heterogeneity of sequential datasets, creating increased demand for robust and scalable analytic frameworks.

Sequential data modalities are diverse, encompassing numerical, categorical, spatial, textual, image-based, and even molecular or biological sequence formats. This diversity calls for flexible methodological paradigms. In engineering and environmental sciences, spatio-temporal datasets are crucial for modeling complex systems—such as climate processes or energy consumption—where both stationarity assumptions and latent dependency structures present substantial modeling challenges [9, 12, 14, 19]. In modern commerce and public health, real-time sequence analytics underpin adaptive experimentation, optimization, and resource allocation, as seen in A/B testing for e-commerce platforms and spatial-temporal cluster detection in epidemiological surveillance [33, 40, 76].

Recent technological breakthroughs—including automated experimentation, foundation models, and integrated multi-modal methodologies—have further extended the reach of sequential data analysis, unlocking discoveries in structural biology, drug design, and materials science [47, 70, 74, 75].

- The breadth of sequential data applications highlights the universality and adaptive utility of sequential data methodologies.
- The growth in available data and complexity requires continual methodological innovation.

- Cross-domain exchange of analytic tools often fuels progress and inspires novel frameworks.

## 1.2 Motivation for Accurate and Robust Analysis in Sequential Modeling

The imperative for precise and robust analysis of sequential data is anchored in its pivotal role in both high-stakes scientific research and real-world operational contexts. Sequential data streams, such as time series, frequently exhibit patterns including recurring structures, abrupt regime shifts, long-range dependencies, non-stationarity, noise, and significant heterogeneity. These attributes challenge conventional statistical techniques and necessitate ongoing evolution in analytic methodologies [6, 34, 36, 64, 65, 74, 75, 81, 82].

The following considerations underscore the importance of robust methodologies for sequential data:

- **Error Propagation:** In domains like financial risk modeling, healthcare monitoring, environmental forecasting, and automated industrial control, analytic errors can propagate through interconnected systems, resulting in substantial consequences for risk, safety, or resource optimization.
- **Adaptivity:** Financial applications require methods capable of not only accurately forecasting volatility and tail risk but also adaptively updating models in the face of evolving market relationships and latent shocks [34, 82].
- **Real-time Requirements:** Biomedical and clinical settings demand robust sequential methods for early detection of adverse events, optimal decision timing, and real-time adaptation to new evidence, often under constrained or noisy data scenarios [4, 64, 81].
- **Emerging Challenges:** Novel settings—including self-driving vehicles, cyber-physical infrastructure, and large-scale adaptive experimentation—impose new demands: real-time inference, adversarial robustness, interpretability, and scalability to massive, high-dimensional sequence streams [65, 74, 75].

In response, developments in neural network architectures, Bayesian approaches, and hybrid probabilistic-automated paradigms have accelerated, driven by the simultaneous need for accuracy, stability, interpretability, and adaptability [36, 65, 74, 81]. As sequential data becomes more complex and heterogeneous, the necessity for principled model diagnostics, data-driven model selection, and rigorous benchmarking—in both simulated and real-world scenarios—is increasingly paramount [36, 81, 82].

## 1.3 Survey Structure and Thematic Organization

This survey addresses the expansive significance of sequential data by systematically traversing theoretical, methodological, and application-focused aspects of sequence modeling. The organization is as follows:

- **Foundational Theories:** Exploration of core statistical and probabilistic concepts underpinning both classic and contemporary approaches to temporal dependence, stationarity, and inference.
- **Methodological Innovations:** Synthesis of advances in statistical learning, automated model discovery, Bayesian

and probabilistic methodologies, deep learning and neural architectures (with a focus on foundation and generative models), and adaptive learning strategies [14, 65, 74, 75, 81].

- **Representation and Robustness:** Dedicated attention to advances in data splitting, representation learning, robust trend estimation, and adaptive nonparametric frameworks suited for real-world, noisy, and high-dimensional sequences [11, 74, 75].
- **Application Landscapes:** Thematic organization of the sequence modeling literature by domain, including financial econometrics, health streams, climate science, engineering, commerce, structural biology, and scientific discovery platforms [4, 7, 12, 36, 40–42, 47, 63, 70, 74–76, 82].
- **Interpretability and Evaluation:** Critical discussion of interpretability, uncertainty quantification, robustness to data drift and adversarial inputs, and the design of benchmarking protocols for model generalization [65, 74, 75, 81].
- **Outlook:** The review synthesizes open questions and future directions aimed at advancing theoretical, algorithmic, and domain-bridging research in sequential data analysis and modeling.

This structured approach is intended to equip readers with both a broad and deep perspective on the rapidly evolving field of sequential data analysis, encouraging informed engagement with the statistical, computational, and application-driven challenges at its frontier.

## 2 Theoretical Foundations and Core Definitions

### 2.1 Data Types, Modalities, and Structure

A rigorous examination of sequential, temporal, and spatio-temporal data structures forms the foundation of recent advances in statistical modeling and machine learning. Fundamentally, time series are organized as sequences of observations indexed by time, with intervals that may be regular or irregular. Although “temporal” and “sequential” are often used interchangeably to describe unidimensionally indexed series, contemporary applications require extending these frameworks to encompass:

- **Spatio-temporal series:** Observations indexed jointly over both time and space;
- **Compositional or constrained data:** Each observation is a vector constrained structurally, such as compositional vectors whose elements sum to one;
- **Irregular or unstructured modalities:** Including point clouds, graph-based sequences, or data with variable-length observations.

The form of time series data varies widely:

- *Binary series*, e.g., daily indicators of patient mortality;
- *Categorical series*, e.g., discrete event types per interval;
- *Continuous-valued series*, e.g., sensor measurements.

Specialized methodologies are necessary for binary, categorical, and compositional time series. For compositional data, conventional approaches give way to models such as Bayesian Dirichlet or autoregressive conditional proportion processes, which capture the inherent dependencies and constraints [6, 43, 73, 75].

Modalities diversify not only by the characteristics above, but also along other structural axes:

- *Univariate (single-channel) vs. multivariate (multi-channel) series;*
- *Stationary vs. non-stationary dynamics;*
- *Regular vs. irregular sampling schemas;*
- *Discrete, continuous, or compositional value domains;*
- *Higher-order or structured objects: point clouds, graphs, and tensors.*

Correspondingly, domains such as robotics and autonomous systems introduce sequential point cloud data, demanding explicit modeling of spatial relationships across temporal frames [74].

The nature of data representation—e.g., univariate versus multivariate, regular versus irregular sampling—profoundly impacts downstream analysis. For instance, spatio-temporal phenomena may be regularly indexed in time, yet sampled at irregular spatial locations, requiring frameworks that jointly address spatial and temporal autocorrelation [73, 74]. Diverse data types and observation structures—ranging from high-dimensional tensors to graphs and point clouds—necessitate flexible representations and new algorithmic strategies.

A spectrum of statistical challenges arises across data types:

- **Missing data:** Commonplace due to sensor dropouts or irregular sampling;
- **Nonstationarity:** Evolving distributional properties over time;
- **Seasonality:** Periodic trends requiring deseasonalization or specialized modeling;
- **Heteroscedasticity:** Nonconstant variance, often mitigated via transformation or robust modeling;
- **Outliers and noise:** Necessitating robust statistical and machine learning frameworks.

Nonstationarity, in particular, poses persistent obstacles in economic, climate, or biomedical analysis; addressing it may involve robust trend estimation (e.g., via wavelet shrinkage), seasonality removal, or flexible probabilistic models [11, 64, 73].

Interpretability adds a further layer of complexity: in applications such as healthcare and finance, transparency in mappings from input data to predictions is often as critical as predictive accuracy itself [7, 41, 73]. Cutting-edge neural architectures seek to address these requirements by integrating mechanisms for missing data imputation and post hoc interpretability—yet, balancing complexity, informativeness, and transparency remains a critical research challenge.

From a probabilistic standpoint, three main paradigms for uncertainty quantification in sequential analysis have emerged:

- **Frequentist:** Emphasizing asymptotic properties, error metrics such as mean squared prediction error (MSPE), and hypothesis tests;
- **Bayesian:** Allowing full uncertainty quantification via posterior inference and interval estimation;
- **Information-theoretic:** Utilizing entropy-based or mutual information measures for model selection and evaluation.

Bayesian and information-theoretic tools, such as entropy power ratios and mutual information, have proven especially valuable in

high-dimensional or model-misspecified regimes, as they enable principled model assessment and selection [17, 25, 29, 41, 75].

Central theoretical bottlenecks in modern sequential data analysis include high dimensionality, heterogeneity, and irregular observation structures—collectively often referred to as the “curse of dimensionality.” As the number of dependent series or features increases, so too does computational and statistical complexity; this reality necessitates dimension reduction (e.g., dynamic functional principal components, factor models) and robust, scalable neural encodings [4, 7, 16, 32, 35, 38, 40, 41, 49, 51, 58, 72, 74, 84]. Recent approaches such as dynamic neural embeddings are engineered to flexibly aggregate variable subsets and impute missing data, addressing challenges of irregularity and heterogeneity [41, 84]. Emerging data structures—e.g., high-order tensors, spatio-temporal graphs—demand new inductive biases, representation learning frameworks, and highly scalable algorithms. Finally, robustness to distributional shifts and outliers remains essential, especially where classical assumptions (e.g., Gaussianity, regularity) do not hold [7, 41, 74].

Thus, the field has transitioned decisively from its classical roots in univariate, regularly sampled processes to a multifaceted landscape dominated by heterogeneity in data type, modality, structure, and noise, all of which drive both methodological advances and open new theoretical challenges.

## 2.2 Core Statistical and Machine Learning Paradigms

Statistical and machine learning approaches to sequential data are fundamentally rooted in the frequentist and Bayesian traditions, most notably in modeling, inference, and hypothesis testing for time-indexed phenomena [15, 41, 75].

**Frequentist models**, including autoregressive (AR), moving average (MA), and state space approaches, provide interpretable and tractable recursive formulations, with error quantification and model selection procedures enriched via asymptotic theory.

**Bayesian frameworks** enable explicit uncertainty quantification—particularly invaluable when data is limited, highly noisy, or nonstationary—and support adaptive model selection and sequential decision-making. Notable applications include Bayesian state space and hierarchical Dirichlet models for compositional and discrete-valued time series [6, 41, 75], as well as Bayesian sequential testing in clinical and business analytics.

While AR and MA models (and their extensions) remain foundational, the methodological landscape has been substantially transformed by the rise of machine learning-based generative and discriminative models. These encompass:

- **Autoregressive models** (linear and nonlinear), for capturing conditional dependencies;
- **Neural sequence models**, such as recurrent neural networks (RNNs), gated recurrent units, attention mechanisms, and MLP mixer architectures;
- **Generative models** employing deep learning for high-dimensional, heterogeneous series.

A comparative overview of major methodological families is provided in Table 1, highlighting their characteristic strengths and limitations.

**Table 1: Comparison of principal modeling paradigms for sequential data**

Paradigm	Key Models	Strengths	Limitations
Frequentist	AR, MA, ARIMA, State Space, Exponential Smoothing	Interpretability, tractability, well-characterized asymptotics, efficiency in low-dimensions	Limited expressivity; struggles with high-dimensional, nonstationary, irregular data
Bayesian	Bayesian AR, State Space, Hierarchical Dirichlet	Full uncertainty quantification, robust to noise and limited data, sequential adaptation	Computationally intensive, prior specification required
Generative (Machine Learning)	INGARCH, INAR, Dirichlet/Multiplicative, Generative NN	Models for discrete, count-valued, or compositional data, handles overdispersion	Often needs domain-specific customization, interpretability sometimes limited
Neural / Deep Learning	RNN, GRU, Attention, GNN, MLP Mixer	High capacity, non-linear modeling of complex, heterogeneous, multi-modal sequences	May overfit; interpretability and robustness can be challenging; requires large data
Hybrid / Ensemble	Combinations of above (e.g., ARIMA + NN, Boosting)	Enhanced accuracy and robustness, automated pipelines, handles diverse scenarios	Complex design, potential for overfitting, more intensive tuning

Best-practice frequentist models such as ARIMA, state space, and exponential smoothing, have proven to deliver robust, interpretable forecasting in low-dimensional, stationary regimes, benefiting from computational efficiency and strong theoretical underpinnings [31, 56, 74]. Limitations, however, arise in the face of high-dimensionality, nonstationarity, or irregular sampling [7, 41, 43]. Generative frameworks such as INGARCH or INAR extend autoregressive models to count-valued or overdispersed data, while advanced approaches for compositional series leverage Dirichlet or multiplicative dynamic processes [6, 43, 73, 75].

**Neural sequence models**—especially RNNs, gated architectures, attention mechanisms, MLP mixers, and graph neural networks (GNNs)—provide flexible, expressive representations of inherently non-linear and high-dimensional dependency structures [2, 7, 10, 12, 15, 16, 22, 24, 32, 43, 55, 61, 84]. Their effectiveness is particularly evident in domains demanding large-scale, non-linear, or cross-modal inference: for example, in retail forecasting, environmental monitoring, and traffic prediction [2, 24, 32, 43, 84]. Recent empirical work underscores, however, that classical linear models may outperform deep architectures in certain benchmarks—especially under noise misspecification or limited data—highlighting the enduring importance of careful model selection and adaptation [31, 32, 43, 56, 84].

Current research directions also foreground the incorporation of structured information—functional, spatio-temporal, graph, or point cloud representations—directly into neural or probabilistic sequence models. For instance, graph representation learning via GNNs captures topological and sequential relationships in dynamic systems such as traffic, protein networks, or sensor arrays [32, 41, 84]. Autoregressive models for point clouds and tensors have extended sequence forecasting to domains requiring geometric or multi-way dependencies, such as robotics and environmental surveillance [43, 74].

A salient trend is the systematic integration of classical statistical and deep learning approaches. Hybrid and ensemble models—combining ARIMA, exponential smoothing, neural networks, and boosting—have demonstrated top performance in forecasting competitions and underpin state-of-the-art automated time series solutions [22, 24, 43]. These pipelines now also automate data cleaning (handling missing values, outlier correction, deseasonalization), model and hyperparameter selection, and robust evaluation, broadening accessibility without sacrificing predictive power.

Finally, recent work synthesizes Bayesian and information-theoretic principles with neural methods: entropy-based regularization, uncertainty-aware neural predictions, and variational learning illustrate a convergence of statistical rigor and representational flexibility [10, 17, 25, 29, 41, 75]. Such interdisciplinary strategies are crucial for enhancing adaptability, interpretability, and robustness in the face of rapid data scale growth and emerging modalities. Accordingly, current methodological frontiers are increasingly characterized by

architectures and probabilistic frameworks that can accommodate the rich spectrum of time series types and representations encountered across scientific, environmental, and financial contexts—all while remaining computationally tractable and interpretable.

### 3 Bayesian Sequential Testing and Advanced Inference Procedures

#### 3.1 Bayesian Sequential Testing in Clinical and Experimental Trials

Bayesian sequential testing has increasingly demonstrated its value as a principled and flexible alternative to frequentist methods in clinical and experimental research environments. Central to this paradigm is the utilization of posterior probabilities and Bayes factors for dynamic hypothesis evaluation—particularly pertinent for measures such as relative risk in two-arm clinical trials with binary endpoints. Recent advances have rigorously formalized these procedures by specifying Beta priors and explicit stopping rules, which yield not only clear acceptance or rejection criteria but also delineate ‘no-decision’ regions that faithfully reflect evidence ambiguity, following the approach of Berger and Sellke regarding evidential thresholds [75].

A defining strength of Bayesian sequential analysis is its formulation of stopping rules and evaluation of Bayes factors at every interim analysis point. In contrast to classical group-sequential and alpha-spending frequentist designs, Bayesian methods are inherently immune to issues arising from optional stopping; this enables adaptive allocation of sample size and flexible interim analyses without an inflated risk of error or systematic bias. Such intrinsic robustness is particularly advantageous in time-sensitive settings—such as rapidly evolving clinical trials—where swift reaction to interim data can be imperative [75]. Additionally, Bayesian inference affords direct computations of Type I and Type II error rates, reported both marginally and conditionally, thus supporting more nuanced, transparent reporting of evidential strength throughout the trial’s progression.

Recent methodological refinements have resolved several longstanding challenges within the Bayesian framework. One notable advancement is the explicit recognition of ‘no-decision’ regions, which appropriately capture situations where the accumulating data yield inconclusive evidence and preclude premature categorical decisions [75]. Marginal error reporting, meanwhile, confers greater granularity over the risks associated with possible actions at stopping boundaries. For example, in H1N1 vaccine safety trials, Bayesian monitoring frequently led to earlier or later termination than classical rules, contingent on the empirical evidence, and enabled comprehensive sensitivity analyses across a spectrum of prior choices—including uniform, informative, and Jeffreys priors [75]. Importantly, the interpretive richness of posterior distributions

and high posterior density (HPD) intervals enables investigators to directly report uncertainty in critical treatment effect parameters.

Nevertheless, the Bayesian approach entails nontrivial challenges, most notably in regard to prior selection and justification. While the inclusion of prior information is advantageous in leveraging historical or domain knowledge, it also introduces potential subjectivity—especially problematic in contexts of marginal or discordant evidence [75]. Ambiguous data commonly present as Bayes factors near unity or as wide HPD intervals, complicating interpretation and downstream decision-making. Further, the generalizability of Bayesian sequential designs to complex settings (e.g., multiple treatment arms, non-binary outcomes) remains an active area for methodological innovation. Recent developments, such as the introduction of Uniformly Most Powerful Bayesian Tests (UMPBTs), establish a rigorous theoretical correspondence with frequentist testing, thereby facilitating broader interpretability and potential regulatory acceptance [75].

### 3.2 Sequential and Recursive Statistical Tests in Broader Contexts

The increasing demand for robust, online adaptive testing schemes is not confined to clinical trials but spans a spectrum of high-dimensional, nonstationary, and partially observed data settings across scientific and industrial domains. In dynamic environments—including personalized healthcare, algorithmic trading, and automated monitoring systems—the prerequisites of rapid change detection, resilience to missing or irregular data, and robustness to outliers are paramount [1, 7, 11, 12, 34, 37, 43, 60, 81, 86].

A diverse set of sequential and recursive statistical tests has been devised to respond to these challenges and support real-time decision-making. Traditional change-point detection methods—such as the Cumulative Sum (CUSUM) and Generalized Likelihood Ratio (GLR) tests—have long provided foundational tools for sequential inference. However, these methods are often constrained by assumptions of stationarity or single-parameter shifts and may underperform in the presence of multiple or nonstationary changes. Advances such as Data-Adaptive Symmetric CUSUM (DAS-CUSUM) directly address these shortcomings by:

- Symmetrically detecting both increases and decreases in the mean or variance,
- Re-centering the detection process to enable equivalence in responding to diverse change types,
- Employing a fixed threshold that adapts to evolving distributions,
- Delivering improved theoretical guarantees for expected detection delay and false alarm rates,
- Consistently outperforming classical CUSUM in real-world, nonstationary time series settings [1].

Further breakthroughs have facilitated recursive and intersection-union testing (IUT) frameworks designed for highly adaptive experimentation. For example, online studies with covariate-adaptive randomization and temporally dependent outcomes increasingly leverage sequential updating of test statistics. In these settings, robust Type I error control at arbitrary stopping times is obtained through p-value-based aggregation across null components, supporting finite-sample error guarantees even under adaptation and complex

dependence structures. Recent applications in mobile health interventions—powered by reinforcement learning—demonstrate the practical viability and statistical robustness of these approaches [67].

Contemporary implementations are further bolstered by innovations in robust estimation, adaptation to missing data, and online normalization protocols. Notably, methods including Temporal Dynamic Embedding (TDE) are employed to effectively handle irregular sampling in clinical time series, while Review-based Instance Normalization (RevIN) and Structure-Adaptive Normalization (SAN) are utilized to address distributional drift and environmental evolution [7, 11, 34, 43, 81]. This arsenal is essential as modern data streams—spanning sectors such as energy, finance, and healthcare—are characterized by:

- Pronounced nonstationarity,
- Increasing dimensionality,
- Elevated sensitivity to noise and outliers,
- The imperative for automated adaptation under frequentist error guarantees.

As illustrated in Table 2, recent approaches bridge the gap between flexibility in handling complex, evolving data and maintaining rigorous error guarantees. The continual advancement of sequential and recursive inference methodologies embodies an ongoing trade-off between adaptive power and principled risk control. Notably, the convergence of Bayesian, frequentist, and hybrid strategies—each incorporating innovative stopping rules, robust normalization, and explicit error quantification—underscores a maturing field that is increasingly responsive to the challenges of modern data-centric science.

As these methodologies become more prevalent in both experimental and operational settings, ongoing assessment of their scalability, interpretability, and cross-domain robustness constitutes a central focus for future research [1, 7, 11, 12, 24, 34, 37, 43, 60, 67, 75, 81, 86].

## 4 Methodologies for Splitting, Preprocessing, and Representing Sequential Data

### 4.1 Data Splitting Strategies and Representation Protocols

Robust analysis of sequential data—spanning domains such as time series, sensor traces, and video streams—requires a deliberate and principled approach to both dataset partitioning and representation. The selection of data splitting strategies is fundamental, as it safeguards against information leakage, upholds temporal causality, and enables fair evaluation of model performance in tasks like object tracking, anomaly detection, forecasting, and event analysis. Inadequate splitting can lead to contaminating the evaluation process with future knowledge or yield unrepresentative partitions, thereby jeopardizing predictive validity and real-world generalizability [10].

Quality criteria for a splitting protocol must simultaneously uphold statistical rigor and respect domain-specific imperatives. In scenarios involving temporal analysis, chronological splitting is essential to preserve genuine causal relationships; this is particularly critical in applications such as anomaly or event detection, where preventing future data from influencing the modeling of past

**Table 2: Summary of Key Sequential Testing Methodologies and Features**

Method	Adaptivity	Handles Nonstationarity	Error Control
Classical CUSUM	Fixed	No	Asymptotic (stationary)
GLR Test	Fixed/ad-hoc	Limited	Asymptotic (simple shifts)
DAS-CUSUM	Highly adaptive	Yes	Finite-sample
Recursive IUT	Sequential/online	Yes (via robust aggregation)	Finite-sample
Bayesian Sequential	Fully adaptive, prior-driven	Yes (via model updating)	Direct (posterior/report)

behaviors is paramount. The choice of split ratio—such as the frequently used 70/30 or 80/20 configuration—should be tailored to the dataset’s size, variance, and primary application. Notably, smaller or highly variable datasets necessitate more conservative splits to ensure stable estimates of model reliability [10]. Furthermore, the context of data acquisition and intended deployment informs split design: for example, in the context of motor test benches or particle tracking, domain expertise concerning operational regimes, experimental phases, or cycle boundaries is often integral to the rationale underlying specific split protocols [10].

The assessment of split reliability encompasses both quantitative and qualitative considerations. Protocols must verify that essential statistical properties—such as the distribution of seasonality, infrequent events, and trend shifts—remain comparable in post-split subsets relative to the full dataset. Particularly for sequential recommender systems, maintaining the integrity of sequential structure is critical; neglecting to account for order may yield misleadingly optimistic evaluation metrics and compromised generalizability in practice. Empirical findings demonstrate that failure to respect authentic sequential dependencies can distort the apparent strengths and weaknesses of models, underscoring the need for standard, scenario-specific criteria in dataset selection and splitting [10].

To facilitate a structured comparison of commonly employed splitting strategies, Table 3 provides an overview of their defining characteristics and typical use cases.

The strategic choice among these methods, as summarized in Table 3, must be guided by the domain requirements, sequential dependencies, and the specific objectives of the research or application context.

## 4.2 Preprocessing, Feature Engineering, and Data Preparation

Rigorous preprocessing is foundational to the development of robust sequential models. The inherent challenges of sequential data—ranging from non-stationarity and seasonality to the presence of noise, outliers, and missing values—necessitate a sophisticated array of data transformation techniques [31]. The available methodological toolkit encompasses:

- **Normalization:** Standardizes feature scales for improved model convergence and comparability.
- **Spectral Decomposition:** Reveals underlying frequency components, enabling targeted modeling or removal of periodic behavior [86].
- **Deseasonalization:** Removes recurring seasonal structures to stabilize the learning process, particularly beneficial in

contexts with non-stationary or heterogeneous seasonalities [65, 77].

- **Imputation:** Addresses missing values through strategies such as mean/mode replacement, interpolation, or advanced temporally-aware algorithms [45, 73].
- **Outlier and Noise Handling:** Distinguishes between meaningful anomalies (crucial for detection tasks) and corrupt data, utilizing robust thresholds or adaptive denoising modules [31, 73].

While recent advances in neural architectures (notably RNNs and Transformers) have facilitated end-to-end learning of seasonal and trend dynamics, empirical evidence supports the explicit removal or modeling of seasonality for datasets characterized by irregular or non-stationary patterns [65, 77]. This pragmatic approach mitigates risks such as overfitting and ensures model generalizability, especially in small-scale or complex datasets.

The advent of automated feature engineering and preprocessing frameworks—exemplified by systems like Preptimize—has broadened access to effective time series modeling by algorithmically optimizing pipelines for imputation, noise reduction, and feature transformation [45]. Such systems leverage a combination of statistical profiling and meta-learning to recommend best practices tuned to dataset-specific properties, enhancing both model accuracy and usability in practitioner and non-expert settings.

Despite technological advances, core challenges persist. The fine-grained treatment of outliers and noise requires calibration to neither overlook critical anomalies nor inadvertently remove informative structures. Further, the choice of imputation strategy must balance the preservation of temporal dependencies against the risk of distorting sequential patterns [45, 73]. Thus, careful validation—including scenario-based testing and ablation studies—is imperative to ensure preprocessing pipelines neither compromise nor artificially enhance downstream model evaluation.

Fundamentally, preprocessing and feature engineering should not be viewed as isolated or linear steps. Instead, their design and implementation are inherently interactive with both data splitting and model selection protocols. Decisions at each stage reverberate through the modeling pipeline, directly influencing reliability, interpretability, and the ultimate real-world applicability of sequential learning systems. This underscores the necessity for transparent, empirically validated, and context-sensitive methodological standards in the research and deployment of sequential models [10, 31, 45, 65, 73, 77].

Table 3: Comparison of Common Sequential Data Splitting Strategies

Splitting Strategy	Description	Primary Use Cases
Chronological Split	Partitions data according to time order (e.g., first $N\%$ for training, remainder for testing); strictly forbids future data in training	Forecasting, anomaly detection, event detection, real-time control
Random Split	Samples records randomly without regard to sequence; may break temporal dependencies	Weakly sequential data, baseline comparisons, early-stage model development
Block/Holdout Split	Splits data by discrete blocks (operational cycles, experimental regimes), preserving structural boundaries	Laboratory experiments, particle tracking, multi-period process monitoring
Rolling/Expanding Window	Sequentially increment training window or sliding test window across time	Online learning, incremental model updating, validation of time-varying behavior

## 5 Statistical, Automated, Deep Learning, and Sequential Modeling Approaches

### 5.1 Classical, Bayesian, and Generative Models

The domain of time series modeling has progressed from established statistical methodologies to highly sophisticated high-dimensional and generative frameworks, each targeting specific challenges such as nonlinearity, overdispersion, multivariate dependencies, and compositional structures.

Traditional univariate models—including autoregressive (AR), autoregressive integrated moving average (ARIMA), exponential smoothing, and state space models—have historically underpinned time series analysis by delivering interpretable parameters alongside a robust inferential theory. These models continue to demonstrate reliable performance for stationary or moderately complex series, owing to their proven efficacy, computational expediency, and transparent diagnostic structures [2, 3, 7, 11, 12, 15, 16, 22–24, 31, 32, 36, 39, 41, 43, 48, 55, 56, 61, 64, 75, 84]. Generalizations to vector autoregressive (VAR) and various index models, aggregated and disaggregated approaches, extend these merits to multivariate and hierarchical contexts, facilitating advanced applications in macroeconomics, finance, and hierarchical forecasting [12, 24, 43, 48, 55, 61, 84].

Nevertheless, the assumptions of linearity and stationarity inherent in these classical methods can be limiting when modeling phenomena with pronounced temporal variation, nonlinear dependencies, or intricate error structures. To address these limitations, Bayesian frameworks and time-varying parameter models have grown in prominence, offering mechanisms to incorporate prior knowledge, adapt to structural change, and improve forecast calibration. Approaches such as Beta-ARMA for bounded data, Dirichlet and zero-inflated models for compositional and sparse series, and techniques leveraging Markov Chain Monte Carlo (MCMC) inference, enable principled quantification of uncertainty and flexible accommodation of nonstationarity in both univariate and multivariate environments [7, 15, 25, 41, 75]. An exemplar in this space, the Bayesian Dirichlet autoregressive moving average model (BDARMA), demonstrates the advantages of Bayesian regularization and shrinkage; it consistently surpasses transformation-based and frequentist alternatives in flexibility and predictive accuracy, particularly for high-dimensional compositional time series [41].

Empirical benchmarking has played a pivotal role in reinforcing methodological advancement. Major competitions such as M3 and M4, along with the publication of open-access datasets, have established objective performance baselines and fostered rigorous diagnostic practices. These efforts reveal that while deep or hybrid models frequently yield superior predictive performance, traditional statistical models still offer considerable strengths in interpretability and computational efficiency—attributes that remain practically important in many domains [10, 15, 33, 36, 41, 48, 56, 57, 65, 77, 84, 86].

The emergence of advanced visualization tools and instance space analyses has facilitated finer-grained understanding of model behavior across varying time series characteristics, promoting targeted methodological improvements [65].

Moving beyond univariate settings, the modeling of rare events and tail dependencies, as well as the ability to capture higher-order and nonlinear dynamics, has garnered increasing attention. Nonlinear VARs and universal function approximators—such as neural networks, Volterra, and Weiner series—enable enhanced modeling flexibility, particularly for economic and financial series that exhibit regime shifts or abrupt nonlinearities [54, 70]. Empirical studies confirm that multioutput neural networks and nonlinear VARs frequently outperform their linear counterparts, especially in settings involving complex interactions or long memory effects [70]. For applications where risk quantification is paramount, such as modeling climate extremes or financial shocks, copula-based techniques and transformed-linear time series models afford interpretable and tractable tools for quantifying extremal dependence and tail risk, effectively bridging classical time series approaches with extreme value theory [35, 46, 58, 80]. These methods’ explicit attention to tail dependence addresses a critical limitation of Gaussian or ARMA-style models, though open challenges remain around model selection, interpretability, and scalability in ultra-high dimensional or sparse regimes [35, 80].

Matrix and tensor autoregressive (TAR) models, along with hypergraph-based approaches, further extend the capacity to model structured, high-dimensional, and multimodal time series. Techniques such as alternating least squares and fast Fourier transforms within TAR frameworks, as well as spectral and incidence matrix formulations for hypergraphs, promote computational efficiency and robust management of seasonality, nonlinearity, and nonstationarity across multivariate and spatiotemporal datasets [11, 22, 32, 35, 46, 57, 58, 60, 80, 84]. However, these advanced models confront specific challenges, including the absence of standardized procedures for hypergraph construction, limited explainability, and a need for scalable architectures that effectively aggregate heterogeneous data sources [57, 58]. Current research efforts are thus directed toward reconciling statistical rigor with scalable computational design, especially for applications in climate science, bioinformatics, and network surveillance.

### 5.2 Deep Learning and Neural Sequential Models

The last decade has witnessed a fundamental shift in time series modeling, driven by deep learning architectures—especially sequential neural networks—which have established dominance across both research and applied domains. Recurrent Neural Networks (RNNs), including long short-term memory (LSTM), bidirectional LSTM (BiLSTM), and spatiotemporal extensions, are widely adopted

owing to their capacity to capture long-range dependencies and intricate temporal dynamics. Their effectiveness spans a diverse range of applications, from healthcare and traffic forecasting to financial time series analysis [12, 13, 19, 26–28, 30, 31, 36, 37, 54, 60, 68, 70, 76–78, 86].

A significant leap in modeling capabilities has been achieved with the introduction of attention-based architectures, notably the Transformer family and its derivatives (PatchTST, iTransformer, Mixer, Mamba). These models excel at representing global and multiscale dependencies in temporal data, effectively overcoming RNN limitations such as vanishing gradients and computational inefficiencies in very long sequence regimes [19, 26, 30, 68, 76, 78]. Notably, PatchTST leverages a patching strategy that processes time series segments and channels independently, demonstrating that lightweight configurations can rival, and occasionally surpass, larger and more complex architectures by optimizing the trade-off between semantic retention and computational cost [19].

More recently, state-space-based models—such as HiPPO, S4, S7, and Mamba—have begun matching or exceeding the empirical performance of Transformer-based architectures in both efficiency and scalability. These approaches utilize novel parameterizations of state transitions, together with hardware-conscious computational schemes and dynamic input-dependent filtering, enabling the capture of dependencies over sequences that encompass up to a million time steps. Such advantages manifest most notably on benchmarks requiring expansive memory and adaptability [26, 68, 76, 78]. Nevertheless, these innovations bring new questions, particularly regarding their generalizability, requirements for input-dependent stability, and the integration of domain-specific inductive priors.

In parallel, deep generative models—including energy-based approaches, variational autoencoders (VAEs), generative adversarial networks (GANs), normalizing flows, and unified frameworks such as UCGM—have extended the reach of time series modeling into probabilistic, simulation, and decision-theoretic domains. These methods facilitate uncertainty quantification, data synthesis, scenario simulation, and Bayesian inference, yielding substantial benefits in high-dimensional, likelihood-intractable, or rare-event scenarios [8, 47, 59, 69]. Frameworks like UCGM further streamline the training and inference landscapes for generative models by offering versatile objectives that can accommodate various techniques (including diffusion, flow-matching, and VAE-based approaches), thereby improving both inference speed and sample fidelity [47]. Deep network-based simulation-augmented Bayesian inference methods (e.g., using normalizing flows) now render posterior estimation feasible in scenarios where conventional Markov Chain Monte Carlo approaches would be computationally impractical, as exemplified in the study of stochastic epidemic processes and extremely high-dimensional parameter spaces [69]. Despite these advances, substantive limitations persist—notably, posterior collapse in VAEs, instability in GAN training, and enduring challenges in producing interpretable and well-calibrated uncertainty estimates [59].

The capabilities of deep sequential models are further underscored by their application to complex, non-Euclidean, and high-dimensional time series, such as sequential point clouds, dynamic 3D scenes, and tracking in autonomous or clinical settings [74]. These contexts frequently entail data sparsity, irregular sampling,

and intricate spatiotemporal dependencies. Recent approaches leveraging dynamic embeddings—where model parameters evolve with the data context—show promise for managing missingness and irregularity in clinical and sensor data, often outperforming imputation-based or fixed-architecture alternatives [60].

Nevertheless, each domain retains its own modeling challenges. For example, in financial forecasting, hybrid models (which combine or ensemble predictions from deep and statistical approaches) routinely outperform single-model solutions, with rigorous error metrics and tests (e.g., Diebold-Mariano) supporting the need for synthesis between statistical robustness and deep feature extraction [12, 77]. Interpretability and explainability also remain significant obstacles in fields such as traffic and healthcare forecasting [30]. Moreover, recent empirical results indicate that simpler models—such as linear or multilayer perceptrons (MLPs)—can occasionally outperform more elaborate attention mechanisms, especially in long-horizon or homogeneous datasets, highlighting the importance of aligning model complexity with data characteristics rather than employing complexity as a default [19, 30, 54, 76].

The ongoing challenges in this landscape can be summarized as follows:

- Scaling to massive, multivariate, and nonstationary series
- Managing distributional shifts and data heterogeneity
- Achieving interpretable and actionable predictions
- Improving feature representation, self-supervised learning, and transferability
- Balancing model transparency and predictive accuracy

While recent advances in representation learning, self-supervision, and transfer learning have contributed to improved generalization, the inherent opacity of deep architectures continues to impede diagnostic transparency and actionable insight. Thus, developing models that are intrinsically interpretable or easily explainable post hoc—using, for instance, frameworks such as SHAP or LIME—remains an ongoing priority for both researchers and practitioners [30].

### 5.3 Kernel, Sequence Modeling, and Efficient Computation

Kernel and sequence kernelization methodologies have recently enlarged the algorithmic toolkit for time series analysis, providing flexible and theoretically sound mechanisms to model sequential patterns across diverse data modalities (numerical, textual, or graph-structured). Through the transformation of arbitrary input kernels into sequence-consistent forms, these methodologies guarantee positive definiteness and afford principled ways to capture complex temporal dependencies without reliance on explicit feature engineering [44]. Advances in dynamic programming, as well as the adoption of low-rank tensor techniques, have rendered these kernels tractable for moderately large sequences and facilitated their seamless integration into support vector or Gaussian process-based modeling frameworks.

As time series datasets have proliferated in scale and complexity, computational efficiency has become increasingly critical. Progress in mesh-free learning, stochastic gradient optimization, fast Fourier transform algorithms, and the co-design of neural algorithms with



underlying hardware has addressed numerous performance bottlenecks associated with high dimensionality and large-scale inference tasks [32, 35, 52, 84]. The advent of neural architecture search further automates the design and selection of deep models, ensuring optimal alignment with both data characteristics and available computational resources [52]. Stochastic gradient and dimension-wise decomposition strategies also facilitate efficient training of physics-informed or spatiotemporal neural architectures, which are pivotal for accurate prediction in high-dimensional dynamical systems and partial differential equation simulations [32, 35, 84].

The interplay between model complexity and computational tractability remains delicate. While efficient algorithms alleviate the curse of dimensionality and support scalable, low-latency modeling, challenges such as parallelization bottlenecks, inter-process communication, and memory constraints for extremely long sequences persist—especially as foundation models and real-time interactive systems set increasingly ambitious expectations for time series analysis. Sustained progress in this area critically depends on the continued synthesis of methodological innovation, scalability, interpretability, and computational efficiency.

## 6 Principal Analytical Tasks and Application Domains

### 6.1 Analytical and Predictive Tasks

The field of sequential data analysis has undergone a profound transformation, propelled by high-capacity models and advanced computational frameworks. This evolution has expanded the core analytical repertoire to encompass tasks such as forecasting, classification, imputation, anomaly detection, change or structural break identification, index modeling, model-order selection, sequential restoration, and complex trajectory generation. These tasks facilitate a deep understanding of temporal dependencies and underpin robust decision-making within dynamic, data-rich environments.

Forecasting represents a foundational application area, strengthened by the maturation of both classical and contemporary deep learning methodologies. Recent advances, such as nonparametric trend extraction through adaptive wavelet shrinkage, have demonstrated notable improvements in capturing both abrupt and gradual trends in noisy, nonstationary series. These innovations often surpass classical ARIMA and spline-based models in terms of short- and long-term predictive performance, while also reducing computational requirements and the necessity for extensive parameter tuning [64]. Parallel to these developments, dynamic network-based volatility models—including the network HEAVY model—have expanded the frontiers of financial forecasting by modeling evolving dependency structures among asset volatilities. Such approaches offer enhanced forecasting accuracy and enable real-time identification of economically relevant clusters, underscoring the significance of temporal network dynamics, particularly during periods of financial stress [11].

In the domain of classification, sequential analysis leverages both discriminative and generative paradigms. Notably, causal-based distance measures rooted in Granger causality have facilitated multivariate time series clustering, revealing latent structures that traditional methods may overlook [77]. Moreover, model architectures designed for spatiotemporal dependencies, as commonly found in

traffic, meteorology, and environmental datasets, employ graph neural networks and attention mechanisms. These advances have substantially improved anomaly and change point detection and enhanced sequence classification accuracy in dynamic, multivariate contexts [80].

Tasks such as imputation and anomaly/change detection have benefited from progress in autoregressive, energy-based, and generative modeling. For example, Data-Adaptive Symmetric CUSUM algorithms have been developed for detecting regime shifts involving simultaneous mean and variance changes, thereby increasing adaptability compared to conventional approaches [41]. Sequential restoration tasks—including deblurring, dehazing, and deraining—are now frequently addressed using generative adversarial frameworks that integrate explicit physical constraints. The result is solutions that are both data- and physics-consistent [62].

A key dimension of recent progress lies in the deployment of advanced generative and reinforcement learning (RL) models for sequential control and decision-making. These generative frameworks—encompassing variational autoencoders (VAEs), generative adversarial networks (GANs), energy-based models, diffusion flows, autoregressive systems, and Bayesian adaptive methods—offer flexible tools for modeling complex data manifolds and supporting both distribution learning and scenario or trajectory synthesis [8, 33, 47, 62, 75]. Within dynamic control and adaptive interaction, generative models play dual roles as simulators of plausible futures and as optimizers that direct policy selection toward optimal or robust results under uncertainty [47]. Furthermore, RL-augmented sequence-to-sequence models have proven effective in mitigating exposure bias and measurement mismatch, achieving improved outcomes in demanding sequence prediction and summarization applications [47].

Index modeling and model-order selection, especially in multivariate or high-dimensional settings, necessitate not only predictive accuracy but also information-theoretic rigor. Techniques that quantify log ratios of entropy powers, for instance, have enriched the methodological toolkit for autoregressive model order selection by connecting mean squared prediction error with mutual information gains. This perspective is particularly valuable in applications characterized by intricate interdependencies, such as speech and physiological signal processing [25]. Recent efforts in standardizing benchmarks and creating open-source libraries, along with well-curated curation pipelines, have facilitated reproducible evaluation and equitable comparison across indexing, forecasting, and scenario generation tasks [45].

Scenario and trajectory generation—critical for simulation, planning, and synthetic benchmarking—has witnessed significant advances via autoregressive generative models and diffusion-based architectures. These model classes enable temporally coherent trajectory synthesis under complex constraints, as exemplified by applications ranging from financial risk prediction to autonomous robotic planning [33, 47, 62].

### 6.2 Application Areas

The theoretical advancements described above have yielded broad, high-impact applications across scientific, industrial, and societal

domains. The diversity and depth of these applications are summarized in Table 4, which provides an at-a-glance mapping between analytical tasks, relevant modeling methods, and representative domains.

The following itemized overview offers additional detail on domain-specific analytical needs and the impact of methodological innovation:

- **Healthcare, Clinical Trials, and Epidemiology:** Time series and sequential modeling approaches are essential in digital health, enabling adaptive clinical trials, personalized interventions, and real-time epidemiological surveillance. Methods for online/adaptive experimentation and hierarchical testing provide rigorous control over type I error rates and enable adaptive sampling in mobile health and digital interventions [14, 21, 67, 75]. Techniques for dynamic resource allocation and patient stratification further benefit from multivariate clustering and causal inference strategies [77].
- **Finance, Economics, and Asset/Resource Management:** Rising complexity in trading, asset dependencies, and economic volatility has catalyzed breakthroughs in volatility modeling, hybrid forecasting, and regime detection. Deep learning and hybrid approaches—including transformer-based models and neural-statistical combinations—have elevated predictive precision, especially when rigorously benchmarked [11, 36, 63, 64, 82]. Copula-based models have proven effective for rare or extreme event analysis in risk-sensitive sectors [50, 58].
- **Meteorology and Environmental Science:** Accurate modeling of environmental phenomena requires methods attuned to both local and global spatiotemporal dependencies. Innovations in spatial/temporal stationarity testing, deep learning for weather forecasting, and hypergraph models for rare event prediction have all pushed analytical capabilities forward [42, 58, 66, 74, 80].
- **Robotics, 3D Perception, Vision, and Natural Language Processing:** Sequence modeling enables advanced scene understanding, trajectory prediction, and restoration tasks in robotics and vision. Models such as PredRNN and unified sequence architectures facilitate accurate forecasting and robust restoration in dynamic, visually rich scenarios, while sequential generative models provide flexible frameworks for learning from multimodal spatiotemporal data [19, 33, 53, 62, 76].
- **Industrial Applications: Traffic, Recommender Systems, Social Networks:** The analysis and management of interdependent, nonstationary phenomena—such as urban traffic or user behaviors—benefit from deep sequence-aware architectures, graph-based models, and real-time benchmarking protocols. Methods for efficient restoration, automated partitioning, and adaptive experimentation drive both operational excellence and platform innovation [5, 10, 38, 45, 80, 82].
- **Advanced Science: Higher-Order and Compositional Data:** Modern challenges in molecular biology, chemistry, and network science necessitate frameworks capable of capturing high-order and compositional structures. Approaches

utilizing hypergraph neural networks, tensor autoregressions, and Bayesian compositional models enable rigorous inference and prediction amidst complex multirelational data and rare phenomena [32, 50, 58, 62, 80, 84].

- **Restoration, Adaptive Experimentation, and Benchmarking:** Restoration tasks in image and signal domains—such as deblurring, dehazing, and deraining—are now frequently addressed via physics-informed deep generative frameworks. Advances in adaptive experimentation, including contextual bandits and sequential A/B testing, underscore the importance of online analysis. Standardized protocols and dataset curation ensure replicable, fair assessments in sequential analysis and recommendation systems [10, 44, 45, 62, 67, 75].

In summary, the synergy between methodological innovation and application-driven demands has fueled a vibrant and rigorous field of sequential data analysis. Looking forward, deeper integration of generative, reinforcement learning, and graph/hypergraph architectures is expected, alongside a continued focus on interpretability, benchmarking, and domain-specific problem solving within rapidly evolving application areas.

## 7 Graph-Based, High-Order, and Recommender Architectures

### 7.1 Graph Neural and Hypergraph Models

The evolution of graph neural network (GNN) architectures has fundamentally transformed approaches to modeling complex relational and high-order interactions across diverse domains, including social analysis, bioinformatics, financial forecasting, and recommender systems. At the forefront of this paradigm shift are models such as Graph Convolutional Networks (GCN), GraphSAGE, and GATv2. These models leverage the structural properties of graphs to effectively capture both local and global dependencies among entities, offering substantial advantages over traditional shallow embedding or vector-based models [23]. GCNs, for instance, excel in high-homophily graphs—where neighboring node features exhibit substantial similarity—while architectures like GraphSAGE and GATv2 employ robust aggregation and attention mechanisms, enabling flexibility and adaptability in heterogeneous or low-homophily settings. These innovations mitigate noise and alleviate the over-smoothing problem frequently observed in deeper GNN layers [23].

Nevertheless, several persistent challenges remain. Notably, limitations in model generalization, computational scalability, and robustness to both feature and structural noise become particularly pronounced when addressing real-world datasets characterized by varying degrees of homophily and signal-to-noise ratio [23, 61].

A significant advancement in this field is the application of GNNs to dynamic and high-dimensional data, such as those encountered in financial volatility networks and evolving social systems. For example, dynamic network models—like the network HEAVY framework—capture temporal evolution by integrating time-varying, factor-driven adjacency structures. This approach enables the extraction of meaningful communities and the identification of volatility clusters in multifaceted environments such as multi-asset financial markets [82]. Community detection and evolutionary clustering—facilitated by GNNs and non-negative matrix factorization approaches—reveal shifts in relational patterns

**Table 4: Representative Application Domains and Associated Analytical Frameworks**

Domain	Analytical Tasks	Representative Methods/Frameworks
Healthcare, Clinical Trials, Epidemiology Finance, Economics, Asset/Resource Management Meteorology, Environmental Science Robotics, 3D Perception, Vision, NLP Industrial, Traffic, Recommender Systems, Social Networks Advanced Science: Higher-Order/Compositional Data	Adaptive experimentation, dynamic clustering, causal inference Volatility modeling, hybrid forecasting, regime detection, rare event analysis Local/global dependency modeling, rare event prediction, compositional analysis Sequence prediction, restoration, scenario generation Streaming analysis, anomaly detection, sequential recommendation Higher-order inference, compositional sequence analysis, rare/extreme event modeling	Sequential testing, online inference, dynamic / multivariate clustering, causal graphs [14, 21, 67, 75, 77] Network-based volatility models, transformer time series models, copula models, hybrid statistical-deep learning predictors [11, 36, 50, 58, 63, 64, 82] Spatiotemporal deep models, copula/hypergraph frameworks, stationarity testing [42, 58, 66, 74, 80] PredRNN, unified neural architectures, sequential generative models [19, 33, 53, 62, 76] Sequence-aware deep learning, graph models, A/B testing, split/partitioning protocols [5, 10, 38, 45, 80, 82] Hypergraph neural networks, tensor autoregressions, Bayesian compositional models [32, 50, 58, 62, 80, 84]

and group memberships over time, offering insight into underlying mechanisms of financial shocks, biological processes, or social behaviors [5, 6, 83, 85]. Further extending the GNN paradigm, causality-aware distance measures, grounded in Granger causality, permit nuanced hierarchical clustering and group-based analysis within domains marked by strong directional dependencies, such as macroeconomics and epidemiology [5].

To overcome the inherent limitations of pairwise graph models, hypergraph neural networks (HGNNs) have emerged as a powerful alternative. HGNNs are uniquely capable of capturing high-order relationships that permeate complex systems in arenas such as bioinformatics, traffic forecasting, and recommender systems [38, 71, 80, 85]. By encoding multi-way interactions via hyperedges, HGNNs extend the representational capacity of standard GNNs. Architectural varieties include spectral and spatial hypergraph convolutions, hypergraph attention mechanisms, and generative models [71]. In the bioinformatics domain, for example, hypergraph convolutions and attention modules have facilitated the discovery of intricate biomolecular and functional group associations that remain elusive to pairwise analysis alone [80]. However, several significant bottlenecks persist:

- Absence of canonical protocols for hypergraph construction,
- Scalability limitations when applied to large-scale real-world graphs,
- Challenging interpretability in heterogeneous and dynamic settings [71].

Recent advancements in dynamic graph learning and differential equations, particularly in the context of traffic forecasting, highlight further progress. Models such as the SpatioTemporal Dynamic Graph Differential Equation (ST-DGDE) framework adaptively capture evolving connectivity within road networks. By integratively modeling both dynamic and static graph structures through unified temporal causal convolutional neural networks, these approaches demonstrate improved performance by accommodating both invariant, long-term associations and emergent, global spatiotemporal patterns [38]. Complementary innovations in forecast reconciliation and hierarchical modeling—anchored by graph-based and factor-driven methodologies—have yielded state-of-the-art error reduction and enhanced interpretability for hierarchical and multi-scale prediction tasks in finance and electricity load forecasting [55, 61, 83].

Despite notable methodological progress, ongoing debates shape the GNN and HGNN research landscape concerning best practices for explainability, benchmark standardization, and principled paradigm selection across diverse data structures [23, 71, 80]. The field continues to seek robust theoretical foundations—particularly concerning the adaptation of spectral theory, causal inference, and manifold learning—to adequately address the challenges posed by large-scale, noisy, and heterogeneous graphs. This is crucial as

applications become more impactful and demand deeper understanding and interpretability of learned relational representations [23, 71].

To clarify and compare the distinctive features and challenges of key graph-based and high-order models, see Table 5.

## 7.2 Recommender Systems and Sequential Benchmarking

The fusion of graph-based and sequential modeling paradigms has catalyzed significant advances in modern recommender systems, enabling the modeling of intricate user-item interactions, temporal dependencies, and high-order behavioral dynamics [18]. When collaborative filtering and pattern-based approaches are integrated with sequential information and graph-relational structures, they consistently achieve superior predictive accuracy and address core challenges such as data sparsity and reduced novelty in recommendations [18]. Importantly, the incorporation of sequential patterns empowers models to learn complex user trajectories—spanning purchases, clicks, or media consumption—while inherently offering scalability and robustness by exploiting structural redundancies within user-item graphs.

Despite these advancements, the efficacy and generalizability of sequential recommender systems (SRSs) remain critically dependent on the selection and preparation of benchmarking datasets. Recent comprehensive analyses underscore a major concern: many datasets widely adopted for sequential recommendation research lack sufficiently rich sequential structure. This is evidenced by only marginal declines in performance metrics and rule counts when user histories are randomly shuffled [45]. In contrast, datasets imbued with substantive sequential patterns exhibit pronounced performance degradation upon history perturbation, thereby revealing the limitations of current benchmarking practices and raising significant concerns regarding the external validity of SRS architectures trained or assessed under inadequately structured data regimes [45].

Consequently, these findings highlight the necessity for rigorous dataset curation. Essential steps include:

- Systematic filtering and deduplication of interaction records,
- Assessment and quantification of sequential rule prevalence prior to experimentation,
- Development and adoption of standardized community protocols to ensure reproducibility and meaningful evaluation in SRS research.

Taken collectively, the integration of graph-based, high-order, and sequential modeling strategies constitutes both a methodological consolidation and an ongoing challenge for the development of effective, scalable, and interpretable AI architectures. Continued research is essential to bridge extant gaps in scalability, fairness,

**Table 5: Comparison of Key Graph Neural and Hypergraph Neural Architectures**

Model	Interaction Type	Key Strengths	Principal Bottlenecks
GCN	Pairwise (high-homophily)	Efficient for homophilous graphs; simple implementation	Prone to over-smoothing in deep layers; less effective in low-homophily settings
GraphSAGE	Pairwise (arbitrary homophily)	Flexible aggregation; generalizes across unseen nodes	Sensitive to feature noise; scalability challenges with massive graphs
GATv2	Pairwise (attention-based)	Handles heterogeneous neighborhoods; mitigates noise	Complexity of attention mechanism; computationally intensive for large graphs
HGNN (Spectral/Spatial)	High-order (hyperedges)	Captures multi-way relationships; suited for bioinformatics, traffic, recommendations	Absence of standardized construction protocols; scalability and interpretability issues

explainability, and robust benchmarking—ensuring that these sophisticated models not only achieve empirical excellence but also adhere to foundational scientific standards for reproducibility and transparency across diverse application domains.

## 8 Robustness, Evaluation, and Benchmarking

### 8.1 Robustness and Ensemble Methods

Robustness remains a foundational imperative in sequential modeling and forecasting, continually challenged by adversarial scenarios, the necessity for rare event prediction, and the complexities of real-world applications. Single-model approaches often falter in the face of these demands, leading to the rise of ensemble methods—such as bagging and diversified regional or probabilistic aggregation—which deliver marked improvements in predictive stability and generalizability, particularly under conditions of nonstationarity or chaotic dynamics. For example, ensemble approaches incorporating techniques like bagging with competitive associative neural networks (CAN2) succeed in both extending predictive horizons and enhancing the reliability of probabilistic forecasts in chaotic systems. This efficacy arises from model diversity: mechanisms such as attractor similarity-based selection and leave-one-out cross-validation facilitate the identification of member forecasts with optimal expected predictability. However, these selection strategies are not without limitations; negative cross-validation correlations can undermine their effectiveness within complex regimes, emphasizing the need for more refined evaluation metrics and advanced learning frameworks [46].

In the context of climate modeling, the use of ensembles—ranging from straightforward averaging and regression-based superensembles to nonlinear architectures like artificial neural networks—consistently produces outputs more closely aligned with observed statistics relative to single-model baselines. Nonlinear ensembling, in particular, is well-suited for addressing intricate, multimodal data distributions, yet these methods may still underreport the incidence of rare or extreme events. This reveals a fundamental tension between minimizing average predictive bias and maintaining high sensitivity to distributional tails [57].

Adversarial robustness has become a core concern in the era of modern machine learning, as high-capacity models, especially deep neural networks (DNNs), have proven susceptible to targeted attacks and marked deterioration under data perturbations. Structural innovations, such as hash-compression, present lightweight and easily deployable defenses by constraining decision boundaries and suppressing exploitable gradients, thereby reducing attack success rates within DNNs [51]. Modular model architectures—drawing from graph neural networks and transformer-based designs—introduce redundancy and flexibility, offering further protection against both structured adversarial manipulations and non-stationary environmental shifts [49]. Systems-oriented advances,

including frameworks for energy-efficient deep learning, simultaneously optimize computational resource consumption and can enhance robustness, as they automate defense strategy adoption and refine compositional model architectures. These innovations often enable a meaningful reduction in computational overhead without sacrificing, and sometimes even improving, baseline resilience to adversarial or out-of-domain data [38, 40]. Nevertheless, as security research evolves in tandem with increasingly sophisticated attack methodologies, it has become clear that effective adversarial defenses must be dynamic, adaptive, and multi-layered—capable of countering both conventional input perturbations and emergent threats such as model jailbreaking or overzealous filtering in large language models [1, 20].

Regionally adaptive and probabilistic aggregation strategies have gained traction for their capacity to model spatial-temporal variability and capture local rare events, a critical asset in domains marked by data heterogeneity and distributional uncertainty. By aligning ensembles or subnetworks adaptively with distinct context regions—whether temporal, spatial, or otherwise—these methods reliably improve both average performance and robustness against outlier scenarios when compared to monolithic architectures [57]. However, this increased mean accuracy often coincides with diminished sensitivity to extremes, as ensemble aggregation can homogenize outputs and underrepresent distributional tails. Effectively mitigating this risk demands architectural advancements combined with a critical perspective on prevailing evaluation and benchmarking standards.

### 8.2 Evaluation and Benchmarking Methodologies

Rigorous evaluation and benchmarking are indispensable for determining both the empirical validity and generalizability of sequential models, particularly as model architectures grow in complexity and application domains diversify. The landscape of performance metrics is extensive and contested: conventional point-wise measures such as mean absolute error (MAE), root mean squared error (RMSE), and their percentage-based derivatives (MAPE, SMAPE) remain foundational, offering critical insights into average model quality. However, reliance solely on these measures may obscure essential aspects, including mismatches in predictive distributions and insensitivity to rare events. To address these shortcomings, statistical tests—most notably the Diebold-Mariano test—enable direct, quantitative comparisons of competing models’ predictive accuracies under realistic cross-validation and partitioning schemes [6, 7, 12, 13, 19, 32–34, 41, 42, 53, 62, 63, 65, 66, 70, 74, 76, 78, 81, 84].

In domains involving sequential imagery or spatio-temporal signals, perceptual and structural metrics such as the structural similarity index (SSIM), peak signal-to-noise ratio (PSNR), and learned perceptual image patch similarity (LPIPS) have grown increasingly

important. These measures align closely with human perceptual judgments and effectively capture higher-order spatial or temporal structures overlooked by standard numerical error metrics. Their relevance is most pronounced in meteorology, video prediction, and restoration tasks, where qualitative fidelity is as important as, if not more than, quantitative accuracy [19, 39, 76, 86]. For multi-sequence or high-dimensional analyses, the use of temporal and sequence-level metrics—including custom-designed consistency measures and context-specific prediction intervals—ensures that both instantaneous and accumulative aspects of model performance are systematically assessed.

Benchmarking datasets serve as critical touchstones for progress assessment and reproducibility. Foundational datasets—such as those from the M3 and M4 competitions, and corpora like TSLib—have substantially influenced time series forecasting, standardizing evaluation protocols and stimulating methodological innovation across both statistical and deep learning paradigms [6, 7, 10, 12, 13, 19, 32–34, 39, 41, 42, 45, 53, 56, 62, 63, 65, 66, 70, 76, 78, 81, 84, 86]. However, their adequacy for certain contemporary evaluation requirements is sometimes questioned, particularly with respect to diversity, sequential structure, and representativeness. Several high-profile benchmarks have been shown to exhibit weak sequential dependencies, thereby limiting their suitability for sequential recommender or forecasting system evaluation. To address these issues, assessment techniques such as rule mining, model-based sequence shuffling, and various forms of sensitivity analysis are employed to characterize dataset structure and ensure it aligns with real-world application complexity. Without such alignment, benchmarking can yield artificially inflated or deflated performance estimates, ultimately diverting research efforts and hampering practical deployment [10].

Data curation—including preprocessing, filtering, normalization, and accessibility—substantially shapes both evaluation outcomes and the utility of research for practical applications. Standardization of dataset split protocols is crucial: optimal approaches vary by application, with temporal splits preferred for forecasting, activity-based splits for behavioral analysis, and particle-tracking splits for physical system modeling. Inadequate splitting strategies risk data leakage and biased results, particularly in settings where violation of i.i.d. assumptions is common due to the presence of autocorrelated and heteroscedastic data-generating processes [45].

The intrinsic properties of datasets—such as structure, scale, heterogeneity, and the mechanisms applied for handling missing data, sequence length variation, and class imbalance—are determinant factors in the statistical power of benchmarking studies and the external validity of reported findings. As the field advances, greater emphasis is placed on:

- Transparent reporting and justification of data curation decisions,
- Adoption of split and preprocessing protocols tailored to both domain and task requirements,
- Continual refinement of benchmarking datasets to capture emerging application scenarios, including rare event modeling and multimodal, sequential decision-making [10, 45].

Taken together, the synergy between robust modeling, adversarial and energy-aware architectural innovations, comprehensive

evaluation strategies, and rigorous benchmarking standards forms the backbone of sustained progress in sequential data modeling. It is imperative for the research community to critically assess both modeling approaches and evaluation infrastructures, thereby fostering reliability, robustness, and practical value in increasingly complex, high-stakes environments.

Table 6 provides a concise overview of the main metric families used in evaluating sequential modeling performance. Selecting and justifying appropriate metrics is critical to aligning evaluation with domain-specific requirements and application objectives.

## 9 Interpretability, Explainability, Security, and Fairness

### 9.1 Interpretability and Explainability

Interpretability within time series and sequential models is now indispensable across domains such as healthcare, finance, intelligent transportation, and climate science. The widespread utilization of deep neural and Bayesian models has delivered significant predictive performance gains. However, these advancements have exacerbated the challenge of articulating the internal mechanisms and decision-making processes of increasingly complex models, particularly as model architecture and data heterogeneity advance.

The interpretability toolkit for time series and sequence models has matured considerably. Both post-hoc and model-intrinsic methods are widely used. Prominent tools include SHAP (SHapley Additive exPlanations), LIME (Local Interpretable Model-agnostic Explanations), saliency maps, and Layer-wise Relevance Propagation (LRP), all of which facilitate attribution and feature importance analysis for both standard and sequence models [2][86][77][81][4]. Temporal attention mechanisms, notably exemplified by the Temporal Fusion Transformer (TFT), further enhance interpretability by enabling dynamic insight into instance-specific and aggregated temporal patterns within deep architectures [2][77][81]. The advent of open-source interpretability frameworks such as iNNvestigate has strengthened reproducibility and comparability in interpretability research, providing unified interfaces for a wide spectrum of qualitative and quantitative analysis techniques—such as perturbation-based evaluation—across varied models and tasks [4].

Interpretability is relevant across multiple modeling paradigms, each with distinctive challenges:

- **Classical linear and autoregressive models:** These models provide transparent parameterization and facilitate direct hypothesis testing; however, their limited expressivity often leads to poor performance when encountering nonlinearities or complex temporal dependencies [24][85][83][43].
- **Deep neural networks:** While highly expressive and adept at modeling complex structures, these models often function as “black boxes,” necessitating supplementary interpretability techniques or the incorporation of attention mechanisms and explicit decomposition structures [81][38][57][29][17][41][15][70][74][75].
- **Bayesian frameworks:** Bayesian models offer inherent uncertainty quantification and transparent posterior analysis. Nonetheless, their probabilistic interpretations can be challenging to navigate, especially in high-dimensional or non-conjugate regimes, and may still require post-hoc explanations for practitioner use [83][43][29][17][15][75].

**Table 6: Overview of Key Evaluation Metrics for Sequential Data Modeling**

Metric	Type	Calibration Purpose / Applicability
MAE, RMSE, MAPE, SMAPE	Point-wise/Scalar	Assess average predictive accuracy; foundational but insensitive to distributional and rare event fidelity
Diebold-Mariano Test	Statistical Test	Rigorous model comparison for predictive accuracy over realistic partitions
SSIM, PSNR, LPIPS	Perceptual/Structural	Capture human-aligned perceptual and structural fidelity in sequential imagery or spatio-temporal signals
Custom Consistency Metrics / Prediction Intervals	Sequence/Interval-based	Address sequence-level or temporal consistency and provide uncertainty quantification; tailored to context

Sustained research has highlighted ongoing trade-offs between interpretability, model complexity, scalability, and the reliability of explanations. In particular, while TFT enables actionable and instance-level explanations with competitive forecasting accuracy, its scalability can be hindered when the number of variables or sequence lengths increases [81][41][70]. Conversely, post-hoc explainer methods such as SHAP and LIME offer broad model-agnosticism and are more scalable, yet may not yield explanations faithful to the intricate nonlinear and temporal interactions present in high-capacity models [81][70][49][74]. Moreover, the faithfulness of explanations—that is, the degree to which attributions accurately reflect model reasoning—remains a subtle issue. Saliency and attribution-based approaches may sometimes produce misleading or inconsistent outputs, particularly in temporal contexts where causal or sequential structure is critical [4][75]. This challenge is particularly pronounced in high-stakes or industrial applications, where both scalability and explanation reliability are imperative.

To clarify the comparative strengths and limitations of major interpretability methods in the context of sequential modeling, Table 7 summarizes key characteristics for select techniques.

As depicted in Table 7, no single approach offers universal strengths; methodological choice involves practical trade-offs among model agnosticism, handling of nonlinearity, temporal awareness, and computational scalability.

## 9.2 Security, Adversarial Robustness, and Fairness

The deployment of deep generative and sequential models in mission-critical environments has foregrounded manifold vulnerabilities, particularly regarding adversarial manipulation and system exploitation. Recent studies have empirically demonstrated the susceptibility of sequence models—including transformers, recurrent neural networks (RNNs), and graph-based architectures—to adversarial attacks capable of provoking incorrect predictions or undermining intended system objectives [51][40][49][20][1]. Such vulnerabilities may be especially acute in real-time and adaptive systems, where resulting disruptions can propagate and trigger cascading failures.

Red teaming, defined as systematic adversarial probing of generative and sequential models, has become essential in contemporary security evaluations. These assessments expose vulnerabilities not only to standard input perturbations but also to advanced threats such as multimodal adversariality, reward hacking within reinforcement learning from human feedback (RLHF), and the circumvention of alignment mechanisms in large language models [51][49][1]. Advances in model compression—such as hash compression-based defenses—demonstrate that shrinking adversarial gradients and fortifying decision boundaries can substantially raise the cost of

successful attacks [40]. Nonetheless, the fundamental trade-off between robustness and efficiency persists: adversarial defense measures may introduce additional overhead, constrain model flexibility, or inadvertently affect fairness and transparency.

Fairness and bias mitigation pose intertwined technical and societal questions. Real-world sequential models are increasingly deployed in heterogeneous and sensitive domains, including clinical decision support, personalized recommendations, and adaptive traffic management. Notably, fairness concerns have been articulated not solely at the inference stage but throughout the machine learning lifecycle—involving data selection, algorithm design, and ongoing model adaptation [38][72][20][79][74]. In dynamic, online environments, sequential feedback mechanisms may compound or perpetuate biases—for example, through the amplification of minority under-representation or reinforcement of existing selection imbalances. As such, algorithmic interventions must account for both temporal data drift and cross-domain policy adaptation [38][72][20]. RLHF-based alignment, although enhancing model capabilities and alignment with human intent, introduces new risks, such as feedback loop biases and reward hacking, underscoring the necessity of robust, causality-driven fairness diagnostics [38][20][74].

The interaction between interpretability, robustness, security, and fairness defines a critical and evolving research frontier. Approaches that intentionally integrate interpretable explanation (for auditability and trust), robust optimization (for adversarial resilience), and algorithmic fairness (for equity) are rapidly becoming requirements in sensitive and dynamically shifting deployment scenarios [81][41][70][49][74][75]. Continued progress in both model-intrinsic and post-hoc interpretability, adversarial defense design, and domain-sensitive bias mitigation will be decisive in determining the trustworthiness and societally beneficial impact of next-generation sequential learning systems.

## 10 Synthesis, Integrative Insights, and Interdisciplinary Opportunities

Advances in time series modeling are characterized by a proliferation of frameworks—including classical statistical models, Bayesian approaches, nonlinear dynamics, deep learning architectures, reinforcement learning (RL), hybrid paradigms, matrix/tensor representations, compositional methods, and graph/hypergraph-based algorithms. Each paradigm contributes unique strengths to the field, while also presenting distinct challenges to effective integration.

A comparative perspective reveals that classical statistical and Bayesian methodologies retain strong advantages in terms of interpretability, principled uncertainty quantification, and statistical rigor. These methods are indispensable in high-stakes settings such as clinical experimentation and sequential decision processes under uncertainty. For example, Bayesian sequential monitoring

**Table 7: Comparative overview of interpretability approaches for time series and sequential models.**

Method	Model-Agnostic	Handles Nonlinearity	Temporal/Seq. Awareness	Scalability
SHAP	Yes	Yes	Limited	High
LIME	Yes	Yes	Limited	High
Saliency Maps	No	Yes	Yes	Medium
LRP	No	Yes	Yes	Medium
Temporal Attention (TFT)	No	Yes	Yes	Medium-Low
Linear/AR Models	N/A	No	Yes	High

affords real-time, interpretable adaptive stopping rules without incurring the penalties associated with frequentist optional stopping—though its efficacy depends on judicious prior selection and can lead to ambiguity in borderline decision cases [10]. In contrast, nonlinear and deep learning models—especially those employing advanced architectures, including Transformers, Graph Neural Networks (GNNs), and hybrid generative frameworks—have realized marked improvements on heterogeneous, high-dimensional, and temporally irregular datasets. Their adoption has been particularly transformative in domains such as traffic forecasting, medical diagnostics, financial risk evaluation, and large-scale recommendation systems [6, 34, 36, 64, 75, 81, 82].

The limitations of isolated modeling approaches become especially evident within interdisciplinary application domains, yet these arenas also provide fertile ground for methodological cross-pollination. Consider, for example, the application of the network HEAVY model in finance, which seamlessly fuses classical volatility modeling with high-frequency network analytics to produce interpretable asset clusters and dynamic responses to external shocks [82]. Similarly, hybrid frameworks that synthesize sentiment analysis features with ensemble and deep learning methods have shown that multimodal contextual integration is as crucial to performance as the underlying algorithmic sophistication [11, 65]. The recent ascendance of foundation models in healthcare and computer vision attests to their state-of-the-art predictive capabilities but also highlights persistent vulnerabilities, including susceptibility to adversarial inputs and diminished interpretability [32, 73, 75, 84].

A sharp rise in multimodal and compositional modeling, spanning from tensor-based temporal analysis to adaptive Bayesian frameworks, further underscores the growing need for unified approaches equipped to reason over diverse data types and domain constraints [20, 40, 49, 51]. This imperative for unification is motivated by several critical desiderata:

- **Scalability:** Methods must accommodate growing data volumes and complexity without steep computational trade-offs.
- **Robustness and Adaptivity:** Models should reliably handle outliers, distributional shifts, and nonstationary environments.
- **Interpretability:** The ability to produce human-understandable outputs remains essential, particularly for regulated and safety-critical domains.
- **Societal Factors:** Security, fairness, and alignment with AI safety principles are increasingly urgent as models are deployed in sensitive contexts [45, 74].

Efforts to address these often-competing requirements have spurred both methodological innovation and practical reengineering. For instance, energy-efficient deep neural network optimization has become vital, ensuring that high-accuracy modeling aligns with resource constraints and sustainability goals, particularly in applications such as cyber-physical threat detection and edge-domain healthcare analytics [84]. The adoption of rigorous and context-aware benchmarking methodologies is now foundational; the limitations of standard datasets for sequential recommender systems and the call for customized evaluation protocols highlight the risk of spurious generalization and reinforce the necessity of thoughtful experimental design [10, 33].

To provide a structured summary of paradigm-specific strengths and challenges, the following comparison is presented in Table 8.

Looking forward, emerging research is driven by the convergence of automated scientific discovery, hardware-aware model learning, and the seamless integration of classical with generative neural methodologies. Large language models now provide unified interfaces for explainable financial forecasting, cross-modal inference, and logical reasoning—consistently outperforming statistical baselines in both accuracy and human interpretability, yet raising persistent concerns regarding alignment, security, and fairness [6, 77, 86]. The melding of RL and generative modeling facilitates optimization for complex, non-differentiable objectives across diverse fields, including natural language, molecular synthesis, and autonomous strategic planning. However, these advances amplify open challenges in reward modeling, bias mitigation, and defense against adversarial manipulation [74, 81, 86].

In the domain of explainability, the pragmatic deployment of transparency-enhancing methods—including post hoc analyses such as SHAP and LIME, alongside model-intrinsic interpretability in temporal fusion architectures—strengthens model acceptance in regulated sectors, while simultaneously revealing trade-offs between transparency and predictive power [34, 65, 75].

Underlying all progress is the imperative of data quality, curation, and evaluation. The trend toward standardized, automated frameworks for preprocessing, quality assessment, and reproducible model optimization—demonstrated by recent systems that democratize access to advanced forecasting workflows—provides crucial infrastructure for accountability and generalizability as models traverse boundaries from finance to healthcare, engineering to e-commerce [10, 45].

In synthesis, the integration of theoretical advances, algorithmic innovation, and empirical deployment is most fruitfully realized within interdisciplinary paradigms—where rigorous evaluation,

**Table 8: Comparative Strengths and Challenges of Major Time Series Modeling Paradigms**

Paradigm	Key Strengths	Principal Challenges
Classical Statistical	Interpretability; Statistically grounded uncertainty quantification; Maturity in regulated domains	Limited scalability; Difficulty with high-dimensional or irregular data
Bayesian	Principled probabilistic reasoning; Real-time adaptation; Comprehensive uncertainty measures	Sensitivity to prior specification; Computational intensity; Ambiguity in borderline inferences
Deep Learning (incl. Transformers, GNNs)	State-of-the-art performance on complex, heterogeneous data; Scalability; End-to-end learning	Loss of interpretability; Adversarial vulnerability; Resource consumption
Hybrid/Ensemble	Context-aware multimodal integration; Enhanced robustness; Improved predictiveness	Increased model complexity; Design and maintenance overhead; Explainability trade-off
Graph/Hypergraph-based	Captures relational structure; Adaptation to networked data; Discovery of latent communities	Scalability in large graphs; Model interpretability; Data curation challenges
Matrix/Tensor-based	Efficient modeling of seasonality and geometric trends; Suitability for compositional analysis	Complexity management; Limited adoption in certain domains

conscientious data stewardship, and context-sensitive adaptation act as enduring bridges across disparate modeling traditions. The ongoing exchange between classical inference and modern computational strategies continues to shape the evolution of robust, interpretable, and socially responsible time series modeling for the next generation of scientific and real-world applications.

## 11 Future Directions and Open Research Challenges

The accelerating advancement of large-scale deep generative and sequential models—particularly those utilizing structured state space (SSM), matrix, and tensor encodings—has substantially transformed the science of learning in high-dimensional, structured, and constrained domains. Nevertheless, the design and deployment of robust, interpretable, and accessible models at scale present several unresolved scientific and engineering challenges that must be overcome to catalyze the next wave of innovation.

### 11.1 Scaling Pretrained Structured Sequential Models

Recent progress in large-scale pretraining, especially involving SSMs and matrix- or tensor-based representations, has markedly enhanced the modeling of long-range dependencies, complex multimodal structures, and irregular sequences across diverse applications including genomics, finance, and climate science [1, 8, 19, 20, 26, 27, 32, 33, 37, 44, 47, 49, 53, 59, 62, 68, 69, 74, 76, 77, 81, 84, 86]. Transformer-derived architectures (e.g., PatchTST, iTransformer, diffusion models) offer considerable expressiveness but frequently face inefficiencies on lengthy or irregular sequences, in addition to limited interpretability in many scientific contexts. SSM-based models—including S4, Mamba, and S7—address these issues through input-adaptive and structured recurrences, showing promise in hardware efficiency and scalability for very long-range temporal modeling [33, 62, 76]. Despite their empirical successes, obstacles such as increased architectural complexity, intricate pretraining protocols, and calibration difficulties—especially under distribution shifts or out-of-distribution scenarios—persist. Multi-dimensional (e.g., matrix or tensor) autoregressive and generative models (e.g., TAR, AddARM) expand modeling capabilities but introduce unique difficulties regarding identifiability, nonstationarity, and parameter estimation, particularly for data exhibiting seasonality or nonlinearity [26, 27, 44, 68].

### 11.2 Advances in Bayesian and Sequential Inference

Bayesian sequential testing and adaptive inference form the methodological backbone of robust decision-making in environments with

multiple arms and structures, such as clinical trials, online experimentation, and dynamic treatment regimes [18, 20, 41, 67, 75, 79]. Modern advances facilitate precise control of error rates and Type I/II errors in adaptive frameworks, which increases transparency and supports optional stopping without bias. Nevertheless, significant barriers remain:

- **Computational Complexity:** High-dimensional and multimodal settings cause pronounced computational strain.
- **Prior Specification:** Defining suitable priors becomes increasingly challenging with model complexity and as the number of arms increases.
- **Design Generality:** Extending inference frameworks to support temporally dependent, contextual, or non-i.i.d. designs is an open issue.

There is also a pressing need for principled methodologies that bridge Bayesian and frequentist paradigms—such as Uniformly Most Powerful Bayesian Tests (UMPBT) and hybrid inferential strategies—to increase interpretability and regulatory acceptance in applied domains [20, 67].

### 11.3 Automated and Adaptive Data Splitting for Complex Sequences

The preprocessing and partitioning of sequential, high-dimensional, and non-i.i.d. data are critical for fair benchmarking and for fostering robust generalization in applications such as forecasting, anomaly detection, and recommendation systems [10, 44, 47, 62, 73]. Recent developments—including the introduction of automated pipelines like Preptimize—leverage statistical and machine learning insights to optimize split ratios and criteria, and automate the construction of reproducible, reliable evaluation datasets [73]. However, key obstacles continue to impede progress:

- **Heterogeneity and Non-Gaussianity:** Standard splitting techniques often falter with highly variable, non-Gaussian, sparse, or zero-inflated datasets.
- **Fairness and Representativeness:** Ensuring that splits adequately represent rare or extreme events, such as those encountered in climate analytics, remains unsolved [10, 44].
- **Guideline Fragmentation:** The field lacks consensus on best practices for splitting, affecting interpretability, reproducibility, and regulatory compliance.

### 11.4 Fusion of Classical and Modern Generative Approaches

Hybrid frameworks that integrate classical time series models (e.g., ARMA, Markov chains) with neural generative and tensorial innovations are emerging as promising solutions for achieving both interpretability and generalization [2, 4, 10, 17, 29, 44, 49, 57, 58, 74,



75, 77, 81, 86]. For example, blending GARCH-style volatility models with neural learners or engineering deep generative models to inherit explicit classical structures enhances both uncertainty quantification and causal interpretability, as demonstrated in financial forecasting and risk modeling [4, 17, 81]. The technical challenges in this synthesis include:

- **Balancing Expressiveness and Interpretability:** Preserving scientific transparency while leveraging the flexibility of deep nets.
- **Identifiability:** Ensuring models remain uniquely interpretable, particularly in the presence of complex non-linearities.
- **Computational Efficiency:** Achieving tractable training and inference, especially for high-dimensional generative processes.

Robustness and generalization outside of distribution—especially for applications in healthcare or the modeling of climate extremes—remain critical open challenges for this fusion paradigm.

### 11.5 Real-Time, Efficient, and Accessible Model Development

Enabling real-time adaptation and broad accessibility demands the concurrent optimization of training pipelines, model architectures, and inference systems [10, 20, 32, 44, 47, 49, 62, 73, 84]. State-of-the-art deep models frequently impose substantial computational and memory costs, compounded as sequence lengths and problem complexity escalate. Hardware-aware algorithm design—epitomized by SSM-based models such as S4, S7, and Mamba—facilitates scalable, parallelizable computation while preserving model fidelity [33, 47, 62]. However, further innovation is required to:

- **Democratize Model Development:** Lower technical barriers for non-experts by developing user-friendly, reproducible tools and automated model selection protocols.
- **Optimize for Edge/Resource-Constrained Environments:** Enhance adaptability without sacrificing rigor, reliability, or interpretability [10, 73].
- **Streamline Deployment:** Integrate fine-tuning, parameter tuning, and deployment into robust, accessible pipelines.

### 11.6 Security, Robustness, and Adaptive Defenses

Deep generative and sequential models are increasingly susceptible to adversarial threats, reward exploitation, and unpredictable out-of-distribution (OOD) or extreme-value events [1, 19, 20, 22, 33, 49, 58, 60, 62, 72, 74, 76, 86]. Recent red-teaming initiatives have categorized a growing taxonomy of vulnerabilities, spanning multi-modal attack vectors, reinforcement learning instabilities, and sophisticated reward hacking strategies [22, 60]. To address these issues, prioritized research directions include:

- **Context-sensitive and Scalable Defenses:** Develop efficient, adaptive protections that remain effective as attack surfaces evolve.
- **Explicit OOD and Extreme-Value Adaptation:** Integrate approaches from extreme-value theory and Bayesian learning to accommodate rare events.

- **Robustness Evaluation Protocols:** Establish standardized, reproducible benchmarks for adversarial and OOD robustness.

Such advancements are essential for the trustworthy deployment of sequential models in critical infrastructure and regulated domains.

### 11.7 Interpretable, Causal, and Fair Prediction

Interpretability and causality in prediction—for tasks such as event forecasting and symbolic modeling—remain central to regulatory approval and public trust [45, 75]. State-of-the-art explanation methods (e.g., Layer-wise Relevance Propagation (LRP), SHAP, iNNvestigate), causal frameworks, and transparent evaluation protocols are increasingly deployed in sectors such as healthcare and autonomous systems [4, 17, 75]. Persistent limitations include:

- **Traceable Explanations:** Difficulty in both explaining and tracing model decisions end-to-end.
- **Causal Inference:** Challenges in actionable cause-effect interpretation within multi-modal and temporally structured data.
- **Fairness Guarantees:** Quantifying and standardizing fairness remains particularly problematic in non-i.i.d. or temporally evolving environments.

Addressing these gaps will require expanded benchmark tasks and diagnostic toolkits focused on transparency, fairness, and generalizability, as well as the integration of domain knowledge and causal reasoning directly into model architectures.

### 11.8 Standardization, Benchmarking, and Reproducibility

The standardization of empirical practices, benchmarking tasks, and reporting protocols is increasingly regarded as essential for ensuring the rigor, interpretability, and comparability of machine learning research [10, 45, 75]. Open-source libraries and cross-domain testbeds (including TSLib, Preptimize, iNNvestigate) are helping to address fragmentation, yet several deficiencies persist:

- **Dataset Structure and Representativeness:** Many benchmarks lack adequate sequential context or fail to represent realistic distributions, reducing empirical validity [10].
- **Fairness and Reproducibility Metrics:** There is insufficient support for measuring interpretability, fairness, and domain relevance.
- **Community-driven Data Efforts:** Enhanced coordination is needed to curate rich, annotated, and contextually relevant datasets.

Developing evaluation protocols that emphasize interpretability, fairness, and application alignment is of critical importance for real-world impact.

### 11.9 Bridging Scalability, Reliability, and Domain-Specific Needs

A fundamental challenge is to reconcile the theoretical scalability and empirical performance of generative and sequential models with the often nuanced and high-stakes requirements of specific

scientific and industrial domains [10, 18, 30, 75]. Key research imperatives include:

- Enhancing model calibration and robust uncertainty quantification,
- Developing domain-sensitive interpretability frameworks,
- Fostering interdisciplinary collaboration among machine learning, statistics, domain sciences, and regulatory bodies.

Meeting these challenges will be essential for unlocking the transformative promise of data-driven decision-making in complex real-world systems.

Table 9 provides a high-level overview of the major open challenges that persist across each of the principal research directions discussed. Addressing these interconnected issues will require sustained interdisciplinary collaboration and methodological innovation to ensure that future generative and sequential modeling advances are not only scientifically rigorous, but also practically reliable, interpretable, and fair.

## 12 Conclusion

In recent years, the landscape of sequential and temporal modeling has been profoundly transformed by advances spanning statistics, signal processing, deep learning, and the seamless integration of these paradigms within complex application domains. This transformation is reflected in the proliferation of diverse model classes, including classical autoregressive and moving average models, Bayesian and autoregressive formulations, nonlinear approaches utilizing global approximators, and deep neural architectures such as RNNs, CNNs, and Transformers. Further, state space, ensemble, matrix/tensor, compositional, graph, hypergraph, and kernel-based frameworks have each contributed to broadening the scope and utility of time series and sequence modeling in both academic research and practical deployments [4, 6, 7, 9–16, 18, 19, 21, 33, 34, 36, 40–42, 45, 47, 53, 63–66, 70, 73–76, 78, 81, 82, 86].

A prominent trajectory in this evolution has been the increasing convergence of generative modeling and reinforcement learning (RL), which underpin sequential decision-making and dynamic inference. Generative models—including variational autoencoders, diffusion models, and attention-based architectures—have become instrumental not only for simulation-based inference but also as critical components within RL agents and scientific decision systems. This synergy facilitates unification of learning-from-data with downstream optimization and control objectives [7, 10, 15, 16, 41, 47, 70, 74, 75]. Simultaneously, Bayesian and sequential testing methodologies—particularly those leveraging adaptive and online strategies—enable rigorous uncertainty quantification and immediate decision-making in applications such as clinical trials and experimental sciences [15, 18, 33, 40, 47, 63]. The advent of matrix, tensor, and compositional models has further enabled efficient modeling of high-dimensional, inhomogeneous, and structurally complex sequential data by capturing domain-specific dependencies and symmetries [19, 41, 53, 65, 76].

It is essential to recognize that the real-world impact and reliability of these approaches are determined not solely by predictive performance or model expressiveness, but by a spectrum of critical properties, including interpretability, scalability, robustness, actionability, security, and fairness. These dimensions are foundational for

responsible and effective deployment of sequential models in high-impact domains such as healthcare, clinical sciences, and experimental settings [4, 7, 9, 11, 12, 14–16, 33, 40–42, 45, 64, 66, 78, 82].

- **Interpretability:** Essential for both scientific insight and accountable decision-making. Although advances such as explainability libraries and compositional architectures—e.g., Kolmogorov-Arnold Networks—offer promising avenues, achieving domain-specific transparency remains an ongoing challenge, particularly for black-box systems [42, 73].
- **Scalability:** Recent progress in hardware-adaptive architectures, distributed optimization, and memory-efficient state space models now enables practical large-scale sequence modeling, making it feasible to analyze massive datasets and long temporal sequences [4, 10, 11, 41, 81, 82].
- **Robustness:** Addressing challenges of covariate shift, adversarial perturbations, missing data, and domain adaptation is critical. Contemporary approaches have increasingly adopted adversarial defenses, pipeline automation, and uncertainty-aware ensemble techniques [4, 18, 41, 45].
- **Actionability:** The capacity to translate model outputs into stakeholder-relevant insights and interventions is critical, particularly in domains where model-driven recommendations can have significant human impact [12, 14, 33, 45, 47, 74, 78].
- **Security and Fairness:** With the deployment of algorithms in sensitive and societal contexts, ensuring model security and algorithmic fairness has become paramount. Developments in bias detection, counterfactual evaluation, and harm mitigation pipelines now accompany practical innovations in secure initialization, compression-based defenses, and red teaming [4, 9, 10, 14, 33, 40, 42, 45, 73–75, 78].

Integrative initiatives—such as the development of standardized benchmark suites, co-optimization frameworks, automated reproducibility workflows, and formalized fairness and security auditing protocols—are transforming not only the scientific methodology but also the broader framework of regulatory and deployment practices in the field [4, 41–43, 45, 47, 66, 73, 81]. The proliferation of open-source benchmarks and ecosystem libraries, while democratizing access to cutting-edge sequential models, underscores the necessity of rigorous dataset curation and thoughtful evaluation protocol design to ensure translational validity and guard against benchmarking artifacts [41, 66, 74, 81].

Looking to the future, several key directions are poised to shape the frontier of sequence modeling for science and society:

- **AI Safety and Human Alignment:** Approaches rooted in reinforcement learning from human feedback, preference calibration, and multimodal alignment are now foundational for scientific AI systems, particularly as generative models increase in creative and autonomous capacity [7, 10, 33, 47, 70, 75].
- **Interdisciplinary Integration:** There is substantial opportunity in bridging scientific, clinical, and sociotechnical data sources, alongside the design of generalizable and transparent benchmarking and adaptive evaluation protocols—enabling learning systems that not only predict, but

Table 9: Representative Open Challenges Across Major Research Directions in Structured Sequential Modeling

Research Direction	Open Challenges
Scaling Pretrained Structured Sequential Models	Architectural complexity and pretraining protocol optimization Calibration under distribution shift/OOD Identifiability and parameter estimation for multi-dimensional models
Bayesian and Sequential Inference	Scalability in high-dimensional, multi-arm settings Robust prior specification Bridging Bayesian and frequentist interpretability
Automated Data Splitting	Split strategies for heterogeneous, non-Gaussian, or sparse data Fairness and representativeness, especially for rare/extreme events Standardization and reproducibility guidelines
Classical–Modern Model Fusion	Balancing interpretability and expressiveness Identifiability in nonlinear/high-dimensional regimes Efficient training and inference mechanisms
Efficient and Accessible Model Development	Scalable and hardware-efficient architectures Automated toolchains for deployment Adaptability for non-experts and edge environments
Security and Robustness	Defenses against adversarial/OOD threats Robustness evaluation benchmarks Context-adaptive security mechanisms
Interpretability and Fairness	Explanation traceability Causal inference in structured/sequential data Robust fairness and bias diagnostics
Standardization and Reproducibility	Standardized, annotated benchmark datasets Protocols for interpretability/fairness evaluation Cross-domain comparability

iteratively refine themselves based on real-world feedback [6, 14, 18, 33, 42, 45, 66, 70, 74, 81].

- **Automation and Accessibility:** As automation in model selection, preprocessing, and decision pipelines becomes pervasive, new challenges arise for ensuring accessibility and reproducibility. Rigorous strategies for interpretability, debiasing, validation, and safety are particularly vital in high-stakes clinical and experimental contexts [4, 9, 12, 14, 15, 18, 41, 42, 45, 66].

The ongoing convergence of scientific discovery, clinical translation, and algorithmic recommendation further blurs traditional boundaries, making the pursuit of robust, fair, and actionable sequential modeling not only a technical objective but also a significant societal mandate. Continued interdisciplinary innovation, ethical vigilance, and methodological rigor will be crucial in realizing the full potential of sequential models for science, medicine, and beyond.

References

[1] N. Ahad, M. A. Davenport, and Y. Xie. 2024. Data-adaptive symmetric CUSUM for sequential change detection. *Sequential Analysis* 43, 1 (2024), 1–27. doi:10.1080/07474946.2023.2272908

[2] A. Aknouche and S. Dimitrakopoulos. 2023. Autoregressive conditional proportion: a multiplicative-error model for (0, 1)-valued time series. *Journal of Time Series Analysis* 44, 4 (2023), 393–417. doi:10.1111/jtsa.12663

[3] A. Aknouche and M. G. Scotto. 2024. A multiplicative thinning-based integer-valued GARCH model. *Journal of Time Series Analysis* 45, 1 (2024), 4–26. doi:10.1111/jtsa.12690

[4] Maximilian Alber, Sebastian Lapuschkin, Philipp Seegerer, Miriam Hägele, Kristof T. Schütt, Grégoire Montavon, Wojciech Samek, Klaus-Robert Müller, Sven Dähne, and Pieter-Jan Kindermans. 2019. iNNvestigate Neural Networks! *Journal of Machine Learning Research* 20, 93 (2019), 1–8. <https://www.jmlr.org/papers/v20/18-540.html>

[5] A. Anastasiou, P. Hatzopoulos, A. Karagrigoriou, and G. Tzagkarakis. 2021. Causality Distance Measures for Multivariate Time Series with Applications. *Mathematics* 9, 21 (2021), 2708. <https://www.mdpi.com/2227-7390/9/21/2708>

[6] S. Bandyopadhyay, C. Jentsch, and S. Subba Rao. 2017. A Spectral Domain Test for Stationarity of Spatio-Temporal Data. *Journal of Time Series Analysis* 38, 2 (2017), 326–351. doi:10.1111/jtsa.12212

[7] C. Berninger, M. Stips, and C. Dewes. 2022. A Bayesian time-varying autoregressive model for improved short-term and long-term prediction. *Journal of Forecasting* 41, 1 (2022), 181–200. doi:10.1002/for.2802

[8] S. Bond-Taylor, A. Leach, Y. Long, C. G. Willcocks, T. Al-Maadeed, N. Moutari, and T. Breckon. 2023. Deep Generative Modelling: A Comparative Review of VAEs, GANs, Normalizing Flows, Energy-Based and Autoregressive Models. *IEEE*

- Transactions on Pattern Analysis and Machine Intelligence* 45, 11 (2023), 12764–12793. doi:10.1109/TPAMI.2021.3116668
- [9] Jannis Born and Matteo Manica. 2023. Regression Transformer enables concurrent sequence regression and generation for molecular language modelling. *Nature Machine Intelligence* 5 (2023), 432–444. doi:10.1038/s42256-023-00639-z
  - [10] D. Botache, K. Dingel, R. Huhnstock, A. Ehresmann, and B. Sick. 2023. Unraveling the Complexity of Splitting Sequential Data: Tackling Challenges in Video and Time Series Analysis. *arXiv preprint arXiv:2307.14294* (2023). <https://arxiv.org/abs/2307.14294>
  - [11] V. Cardiff, M. Ritchie, D. Hoy, and K. Esmail. 2019. Wavelet shrinkage for trend estimation and forecasting in time series. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 68, 2 (2019), 295–316. doi:10.1111/rssc.12296
  - [12] Si-An Chen, Chun-Liang Li, Nate Yoder, Sercan O. Arik, and Tomas Pfister. 2023. TSMixer: An All-MLP Architecture for Time Series Forecasting. *Transactions on Machine Learning Research* (Sept. 2023). Available online: <https://arxiv.org/abs/2303.06053>.
  - [13] J.-T. Chien and C.-W. Wang. 2022. Hierarchical and Self-Attended Sequence Autoencoder. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, 9 (2022), 4975–4986. doi:10.1109/TPAMI.2021.3068187
  - [14] J. Cordes and M. C. Castro. 2020. Spatial analysis of COVID-19 clusters and contextual factors in New York City. *Spatial and Spatio-temporal Epidemiology* 34 (2020), Article 100355. <https://www.sciencedirect.com/science/article/abs/pii/S1877584520300332>
  - [15] Francisco Cribari-Neto and Vinicius F. de Almeida. 2023. Beta autoregressive moving average model selection with application to modeling and forecasting stored hydroelectric energy. *International Journal of Forecasting* 39, 1 (2023), 98–109. doi:10.1016/j.ijforecast.2021.07.003
  - [16] Gianluca Cubadda, Stefano Grassi, and Barbara Guardabascio. 2025. The time-varying Multivariate Autoregressive Index model. *International Journal of Forecasting* 41, 1 (2025), 175–190. <https://www.sciencedirect.com/science/article/pii/S0169207024000384>
  - [17] Sjoerd Dirksen, Martin Genzel, Laurent Jacques, and Alexander Stollenwerk. 2022. The Separation Capacity of Random Neural Networks. *Journal of Machine Learning Research* 23, 309 (2022), 1–47. <https://www.jmlr.org/papers/v23/21-1079.html>
  - [18] Christie I. Ezeife and Hemmi Karlapalepu. 2023. A Survey of Sequential Pattern Based E-Commerce Recommendation Systems. *Algorithms* 16, 10 (2023), 467. doi:10.3390/a16100467
  - [19] X. Fang, F. Wang, L. Liu, J. He, D. Lin, Y. Xiang, K. Zhu, X. Zhang, H. Wu, H. Li, and L. Song. 2023. A method for multiple-sequence-alignment-free protein structure prediction using a protein language model. *Nature Machine Intelligence* 5 (2023), 1087–1096. <https://www.nature.com/articles/s42256-023-00721-6>
  - [20] Giorgio Franceschelli and Mirco Musolesi. 2024. Reinforcement Learning for Generative AI: State of the Art, Opportunities and Open Research Challenges. *Journal of Artificial Intelligence Research* 79 (2024), 1351–1412. <https://jair.org/index.php/jair/article/view/15278>
  - [21] Emma Fransson, Maria Karalexi, Mary Kimmel, Emma Bränn, Natasa Kollija, Auke Tas, Vera van Zoest, Eira Nording, Alkistis Skalkidou, and Fotios C. Papadopoulos. 2022. Differentiated mental health patterns in pregnancy during COVID-19 first two waves in Sweden: a mixed methods study using digital phenotyping. *Scientific Reports* 12, 1 (2022), 21253. doi:10.1038/s41598-022-25107-3
  - [22] Yuan Gao and Han Lin Shang. 2018. Forecasting high-dimensional functional time series: Application to sub-national age-specific mortality. *Journal of Time Series Analysis* 39, 4 (2018), 528–546. doi:10.1111/jtsa.12312
  - [23] Z. Gao and R. S. Tsay. 2019. A Structural-Factor Approach to Modeling High-Dimensional Time Series and Space-Time Data. *Journal of Time Series Analysis* 40, 3 (2019), 343–362. doi:10.1111/jtsa.12481
  - [24] A. García-Pérez. 2022. On Robustness for Spatio-Temporal Data. *Mathematics* 10, 10 (2022), 1–17. <https://www.mdpi.com/2227-7390/10/10/1785>
  - [25] J. Gibson. 2018. Entropy Power, Autoregressive Models, and Mutual Information. *Entropy* 20, 10 (2018), 750. <https://www.mdpi.com/1099-4300/20/10/750>
  - [26] A. Gu and T. Dao. 2024. Mamba: Linear-Time Sequence Modeling with Selective State Spaces. *arXiv preprint arXiv:2312.00752* abs/2312.00752 (2024). <https://arxiv.org/abs/2312.00752>
  - [27] A. Gu, K. Goel, and C. Ré. 2022. Efficiently Modeling Long Sequences with Structured State Spaces. *arXiv preprint arXiv:2111.00396* abs/2111.00396 (2022). <https://arxiv.org/abs/2111.00396>
  - [28] X. Guan, Y. Yang, J. Li, X. Xu, and H. T. Shen. 2022. Mind the remainder: Taylor's theorem view on recurrent neural networks. *IEEE Transactions on Neural Networks and Learning Systems* 33, 4 (2022), 1507–1519. doi:10.1109/TNNLS.2021.3057294
  - [29] Boris Hanin. 2024. Random Fully Connected Neural Networks as Perturbatively Solvable Hierarchies. *Journal of Machine Learning Research* 25, 267 (2024), 1–58. <https://www.jmlr.org/papers/v25/23-0643.html>
  - [30] Yuxin He, Ping Huang, Weihang Hong, Qin Luo, Lishuai Li, and Kwok-Leung Tsui. 2024. In-Depth Insights into the Application of Recurrent Neural Networks (RNNs) in Traffic Prediction: A Comprehensive Review. *Algorithms* 17, 9 (2024), 398. doi:10.3390/a17090398
  - [31] H. Hewamalage, C. Bergmeir, and K. Bandara. 2021. Recurrent Neural Networks for Time Series Forecasting: Current Status and Future Directions. *International Journal of Forecasting* 37, 1 (2021), 388–427. doi:10.1016/j.ijforecast.2020.06.008
  - [32] C. Hill, A. A. Nielsen, and A. K. Sahu. 2021. The tensor auto-regressive model. *Journal of Forecasting* 40, 4 (2021), 636–652. doi:10.1002/for.2735
  - [33] F. Hong, J. Zhang, L. Chen, S. Yang, and Z. Xu. 2024. DaGAN++: Depth-Aware Generative Adversarial Network for Talking Head Video Generation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 46, 5 (2024), 2997–3012. <https://ieeexplore.ieee.org/document/10345691>
  - [34] M. Housni. 2023. Multidimensional Forecasting and Analysis: Exploring Moroccan Learning Data Analytics in Business, Environment, and Sustainability. *IETI Transactions on Data Analysis and Forecasting (iTDAF)* 1, 3 (2023), 75–82. <https://online-journals.org/index.php/iTDAF/article/view/40353>
  - [35] Z. Hu, K. K. Shukla, G. E. Karniadakis, and K. Kawaguchi. 2024. Tackling the curse of dimensionality with physics-informed neural networks. *Neural Networks* 175 (May 2024), 155–176. doi:10.1016/j.neunet.2024.106369
  - [36] R. R. Jha, T. S. Teh, and A. R. Hassan. 2024. Statistical Modeling to Improve Time Series Forecasting Using Machine Learning, Time Series, and Hybrid Models: A Case Study of Bitcoin Price Forecasting. *Mathematics* 12, 23 (2024), 3666. <https://www.mdpi.com/2227-7390/12/23/3666>
  - [37] Ming Jin, Shiyu Wang, Lintao Ma, Zhixuan Chu, James Y. Zhang, Xiaoming Shi, Pin-Yu Chen, Yuxuan Liang, Yuan-Fang Li, Shirui Pan, and Qingsong Wen. 2024. Time-LLM: Time Series Forecasting by Reprogramming Large Language Models. In *Proceedings of the 12th International Conference on Learning Representations (ICLR 2024)*. <https://arxiv.org/abs/2310.01728>
  - [38] W. Ju, F. Zheng, Y. Gu, Z. Liu, Q. Long, Z. Qiao, Y. Qin, J. Shen, F. Sun, Z. Xiao, J. Yang, J. Yuan, Y. Zhao, Y. Wang, X. Luo, and M. Zhang. 2024. A comprehensive survey on deep graph representation learning. *Neural Networks* 173 (Feb. 2024), 1–36. doi:10.1016/j.neunet.2024.106207
  - [39] Y. Kang, R. J. Hyndman, and K. Smith-Miles. 2017. Visualising forecasting algorithm performance using time series instance spaces. *International Journal of Forecasting* 33, 2 (2017), 345–358. <https://www.sciencedirect.com/science/article/pii/S0169207016301030>
  - [40] A. Karamchandani, A. Mozo, S. Gómez-Canaval, and A. Pastor. 2024. A methodological framework for optimizing the energy consumption of deep neural networks: a case study of a cyber threat detector. *Neural Computing and Applications* 36 (2024), 10297–10338. doi:10.1007/s00521-024-09588-z
  - [41] Harrison Katz, Kai Thomas Brusch, and Robert E. Weiss. 2024. A Bayesian Dirichlet auto-regressive moving average model for forecasting lead times. *International Journal of Forecasting* 40, 4 (2024), 1556–1567. doi:10.1016/j.ijforecast.2024.01.004
  - [42] Y. Keneshloo, T. Shi, N. Ramakrishnan, and C. K. Reddy. 2020. Deep reinforcement learning for sequence-to-sequence models. *IEEE Transactions on Neural Networks and Learning Systems* 31, 7 (2020), 2469–2489. doi:10.1109/TNNLS.2019.2929141
  - [43] M. Kim and S.-Y. Shin. 2025. Temporal Dynamic Embedding for Irregularly Sampled Time Series. *arXiv preprint arXiv:2504.05768* (2025). <https://arxiv.org/abs/2504.05768>
  - [44] F. J. Király and H. Oberhauser. 2019. Kernels for Sequentially Ordered Data. *Journal of Machine Learning Research* 20, 31 (2019), 1–45. <https://jmlr.org/papers/volume20/16-314/16-314.pdf>
  - [45] A. Klenitskiy, A. Volodkevich, A. Pembek, and A. Vasilev. 2024. Does It Look Sequential? An Analysis of Datasets for Evaluation of Sequential Recommendations. *arXiv preprint arXiv:2408.12008* (Aug 2024). <https://arxiv.org/abs/2408.12008>
  - [46] S. Kurogi, M. Toidani, R. Shigematsu, and K. Matsuo. 2018. Performance improvement via bagging in probabilistic prediction of chaotic time series using similarity of attractors and LOOCV predictable horizon. *Neural Computing and Applications* 29, 9 (2018), 341–349. doi:10.1007/s00521-017-3149-7
  - [47] Yinchuan Li, Xinyu Shao, Jianping Zhang, Haozhi Wang, Leo Maxime Brunswic, Kaiwen Zhou, Jiqian Dong, Kaiyang Guo, Xiu Li, Zitang Chen, Jun Wang, and Jianye Hao. 2025. Generative Models in Decision Making: A Survey. *arXiv preprint arXiv:2502.17100* (2025). <https://arxiv.org/abs/2502.17100>
  - [48] Thoman Liboschik, Nadja Fokianos, and Roland Fried. 2017. tscount: An R Package for Analysis of Count Time Series Following Generalized Linear Models. *Journal of Time Series Analysis* 38, 3 (2017), 451–478. doi:10.1111/jtsa.12214
  - [49] Lizhi Lin, Honglin Mu, Zenan Zhai, Minghan Wang, Yuxia Wang, Renxi Wang, Junjie Gao, Yixuan Zhang, Wanxiang Che, Timothy Baldwin, Xudong Han, and Haonan Li. 2025. Against The Achilles' Heel: A Survey on Red Teaming for Generative Models. *Journal of Artificial Intelligence Research* 82 (2025), 1–63. <https://www.jair.org/index.php/jair/article/view/17654>
  - [50] Y. Lin and Q. Zhu. 2024. On Vector Linear Double Autoregression. *Journal of Time Series Analysis* 45, 3 (2024), 376–397. <https://onlinelibrary.wiley.com/doi/10.1111/jtsa.12551>
  - [51] Qi Liu and Wujie Wen. 2023. Model Compression Hardens Deep Neural Networks: A New Perspective to Prevent Adversarial Attacks. *IEEE Transactions on Neural Networks and Learning Systems* 34, 1 (2023), 3–14. <https://ieeexplore.ieee.org/document/9742065>
  - [52] Yuqiao Liu, Yanan Sun, Bing Xue, Mengjie Zhang, Gary G. Yen, and Kay Chen Tan. 2023. A Survey on Evolutionary Neural Architecture Search. *IEEE Transactions on Neural Networks and Learning Systems* 34, 2 (2023), 550–570. doi:10.1109/

- TNNLS.2022.3202272
- [53] Z. Liu, Y. Liu, X. Xie, Y. Zhang, and X. Lu. 2025. A novel sequence-to-sequence-based deep learning model for satellite cloud image time series prediction. *IEEE Transactions on Neural Networks and Learning Systems* 36, 1 (2025), 638–651. doi:10.1109/TNNLS.2023.3322146
  - [54] Z. Liu, Y. Wang, S. Vaidya, F. Ruehle, J. Halverson, M. Soljačić, T. Y. Hou, and M. Tegmark. 2025. KAN: Kolmogorov-Arnold Networks. *arXiv preprint arXiv:2404.19756* (2025). <https://arxiv.org/abs/2404.19756>
  - [55] J. K. Møller, P. Nystrup, and H. Madsen. 2024. Likelihood-based inference in temporal hierarchies. *International Journal of Forecasting* 40, 2 (2024), 515–531. doi:10.1016/j.ijforecast.2022.12.005
  - [56] S. Makridakis, E. Spiliotis, and V. Assimakopoulos. 2020. The M4 Competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting* 36, 1 (2020), 54–74. <https://www.sciencedirect.com/science/article/pii/S0169207019301128>
  - [57] B. Mesta, O. B. Akgun, and E. Kentel. 2024. Improving precipitation estimates for Turkey with multimodel ensemble: a comparison of nonlinear artificial neural network method with linear methods. *Neural Computing and Applications* 36 (2024), 10219–10238. doi:10.1007/s00521-024-09598-x
  - [58] N. Mhatre and D. Cooley. 2024. Transformed-Linear Models for Time Series Extremes. *Journal of Time Series Analysis* 45, 5 (2024), 671–690. doi:10.1111/jtsa.12570
  - [59] Maria Nareklashvili, Nick Polson, and Vadim Sokolov. 2025. Generative Modeling: A Review. *arXiv preprint arXiv:2501.05458* (2025). <https://arxiv.org/abs/2501.05458>
  - [60] Yuqi Nie, Nam H. Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. 2023. A Time Series is Worth 64 Words: Long-term Forecasting with Transformers. In *Proceedings of the International Conference on Learning Representations (ICLR 2023)*. <https://arxiv.org/abs/2211.14730>
  - [61] P. Nystrup, J. K. Møller, and H. Madsen. 2021. Dimensionality reduction in forecasting with temporal hierarchies. *International Journal of Forecasting* 37, 3 (2021), 1127–1146. doi:10.1016/j.ijforecast.2020.12.012
  - [62] J. Pan, J. Dong, Y. Liu, X. Zhang, M. Wang, Y. Tai, Y. Tang, K. Lu, Y. Zuo, and L. Zhang. 2021. Physics-Based Generative Adversarial Models for Image Restoration and Beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43, 7 (2021), 2449–2462. doi:10.1109/TPAMI.2020.2973646
  - [63] E. Rusu and O. Manta. 2023. Forecasting and Directions Regarding Sustainable Public Procurement. *IETI Transactions on Data Analysis and Forecasting (iTDAF)* 1, 2 (2023), 98–107. <https://online-journals.org/index.php/iTDAF/citationstylelanguage/get/apa?submissionId=39987&publicationId=77089>
  - [64] H. L. Shang, J. Cao, and P. Sang. 2022. Stopping time detection of wood panel compression: A functional time-series approach. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 71, 5 (2022), 1205–1224. doi:10.1111/rssc.12572
  - [65] J. Shang, T. Nakamura, and S. Hamori. 2023. Does the Sentiment Index Help Predict Crude Oil Prices? *IETI Transactions on Data Analysis and Forecasting (iTDAF)* 1, 1 (2023), 54–65. <https://online-journals.org/index.php/iTDAF/article/view/32683>
  - [66] B. Shi, X. Bai, and C. Yao. 2017. An End-to-End Trainable Neural Network for Image-Based Sequence Recognition and Its Application to Scene Text Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39, 11 (2017), 2298–2304. doi:10.1109/TPAMI.2016.2646371
  - [67] C. Shi and Y. Chen. 2021. An Online Sequential Test for Qualitative Treatment Effects. *Journal of Machine Learning Research* 22, 75 (2021), 1–43. <https://www.jmlr.org/papers/volume22/21-0383/21-0383.pdf>
  - [68] T. Soydan, N. Zubić, N. Messikommer, S. Mishra, and D. Scaramuzza. 2024. S7: Selective and Simplified State Space Layers for Sequence Modeling. *arXiv preprint arXiv:2410.03464* abs/2410.03464 (2024). <https://arxiv.org/abs/2410.03464>
  - [69] Peng Sun, Yi Jiang, and Tao Lin. 2025. Unified Continuous Generative Models. *arXiv preprint arXiv:2505.07447* (2025). <https://arxiv.org/abs/2505.07447>
  - [70] A. Taiebnia and S. Mohammadi. 2023. Forecast accuracy of the linear and nonlinear autoregressive models in macroeconomic modeling. *Journal of Forecasting* 42, 8 (2023), 2045–2062. doi:10.1002/for.3002
  - [71] J. H. Tanis, C. Giannella, and A. V. Mariano. 2024. Introduction to Graph Neural Networks: A Starting Point for Machine Learning Engineers. *arXiv preprint arXiv:2412.19419* (2024). <https://arxiv.org/abs/2412.19419>
  - [72] S. Tian, L. Li, W. Li, H. Ran, X. Ning, and P. Tiwari. 2024. A survey on few-shot class-incremental learning. *Neural Networks* 173 (2024), 49–77. doi:10.1016/j.neunet.2023.10.039
  - [73] Mehak Usmani, Zulfiqar Ali Memon, Adil Zulfiqar, and Rizwan Qureshi. 2024. Preptimize: Automation of Time Series Data Preprocessing and Forecasting. *Algorithms* 17, 8 (2024), 332. doi:10.3390/a17080332
  - [74] H. Wang and Y. Tian. 2024. Sequential Point Clouds: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 46, 8 (2024), 9276–9297. doi:10.1109/TPAMI.2024.3365970
  - [75] J. Wang and B. Boukai. 2025. Bayesian sequential analysis of adverse events with binary data. *arXiv preprint arXiv:2504.02959* (apr 2025). <https://arxiv.org/abs/2504.02959>
  - [76] Ning Wang, Jiang Bian, Yuchen Li, Xuhong Li, Shahid Mumtaz, Linghe Kong, and Haoyi Xiong. 2024. Multi-purpose RNA language modelling with motif-aware pretraining and type-guided fine-tuning. *Nature Machine Intelligence* 6, 5 (2024), 548–557. doi:10.1038/s42256-024-00836-4
  - [77] Y. Wang, H. Wu, J. Dong, Y. Liu, M. Long, and J. Wang. 2024. Deep Time Series Models: A Comprehensive Survey and Benchmark. *arXiv preprint arXiv:2407.13278* (2024). <https://arxiv.org/abs/2407.13278>
  - [78] Y. Wang, H. Wu, J. Zhang, Z. Gao, J. Wang, P. S. Yu, and M. Long. 2023. PredRNN: A Recurrent Neural Network for Spatiotemporal Predictive Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45, 2 (2023), 2208–2225. doi:10.1109/TPAMI.2022.3165153
  - [79] Genta Indra Winata, Hanyang Zhao, Anirban Das, Wenpin Tang, David D. Yao, Shi-Xiong Zhang, and Sambit Sahu. 2025. Preference Tuning with Human Feedback on Language, Speech, and Vision Tasks: A Survey. *Journal of Artificial Intelligence Research* 82 (2025), 522–601. doi:10.1613/jair.1.17541
  - [80] M. Yang and X.-J. Xu. 2025. Recent Advances in Hypergraph Neural Networks. *arXiv preprint arXiv:2503.07959* (2025). <https://arxiv.org/abs/2503.07959>
  - [81] X. Yu, Z. Chen, Y. Ling, S. Dong, Z. Liu, and Y. Lu. 2023. Temporal Data Meets LLM – Explainable Financial Time Series Forecasting. *arXiv preprint arXiv:2306.11025* (2023). <https://arxiv.org/abs/2306.11025>
  - [82] H. Yuan, K. Lu, G. Li, and J. Wang. 2024. High-Frequency-Based Volatility Model with Network Structure. *Journal of Time Series Analysis* 45, 4 (2024), 533–557. <https://onlinelibrary.wiley.com/doi/10.1111/jtsa.12751>
  - [83] L. Yuan, M. Yin, and C. Chen. 2025. Temporal Community Detection and Analysis with Network Representation Learning. *Mathematics* 13, 5 (2025), 698. <https://www.mdpi.com/2227-7390/13/5/698>
  - [84] H.-F. Zhang. 2024. Additive Autoregressive Models for Matrix Valued Time Series. *Journal of Time Series Analysis* 45, 3 (May 2024), 398–420. doi:10.1111/jtsa.12552
  - [85] J. Zhou, R. Gao, J. Yang, and P. Zhang. 2023. Spatial-Temporal Dynamic Graph Differential Equation Network for Traffic Prediction. *Mathematics* 11, 13 (2023), 2867. <https://www.mdpi.com/2227-7390/11/13/2867>
  - [86] H. Zhu, Y. Chen, X. Wei, and W. Wu. 2024. Deep Time Series Forecasting Models: A Comprehensive Survey. *Mathematics* 12, 10 (2024), 1504. <https://www.mdpi.com/2227-7390/12/10/1504>