Text as Dynamics: Quantum-Inspired Non-Markovian Multimodal Time-Series Forecasting

Anonymous submission

Abstract

Multimodal time-series forecasting requires capturing both long-range temporal dependencies and evolving interactions between numerical observations and textual context. Most existing approaches encode text as an auxiliary representation and inject it through static fusion or local cross-modal interactions, leaving delayed and history-dependent effects to be learned only implicitly. We introduce **QMUSE**, a **Q**uantum-inspired **Mu**ltimodal Time **S**eries **E**volution framework that instead treats text as a time-varying driver of quantum-state evolution. QMUSE encodes numerical observations as density operators, maps textual context to a time-dependent Hamiltonian, and evolves the latent states through projected stochastic dynamics inspired by the stochastic Liouville-von Neumann equation. Learnable quantum measurements decode the evolved states into multi-step forecasts, yielding an end-to-end differentiable framework for modeling persistent, dissipative, and history-dependent cross-modal interactions on classical hardware. Experiments on nine real-world datasets demonstrate aggregate performance across domains and improvements across the majority of backbone–dataset combinations. Ablation and diagnostic analyses further verify the complementary contributions of the text-driven Hamiltonian and non-Markovian evolution, demonstrating the effectiveness of QMUSE in capturing persistent cross-modal influences.

Introduction

Time-series forecasting plays a fundamental role in a wide range of critical applications, such as energy management (Liu et al. 2025c) and climate projection (Naisipour, Saeedpanah, and Adib 2025). However, temporal observations do not evolve autonomously. Instead, they are continuously influenced by external environments, such as policy announcements, breaking news, reports and social media streams. Textual information can substantially alter the future trajectory of a numerical system, often before its effects are fully reflected in historical observations (Jiang et al. 2025). This has motivated increasing interest in Multimodal Time-Series Forecasting (MTSF), which jointly exploits numerical sequences and time-aligned textual context to produce more accurate and context-aware predictions.

This is an anonymized submission for review purposes only. Distribution, citation, or public sharing of this manuscript is strictly prohibited. Copyright and publication details will appear in the final version if accepted.

Existing MTSF methods integrate textual and numerical information from four perspectives. First, structure-aware approaches construct temporal knowledge graphs or graph-augmented representations to organize events, entities, and spatiotemporal dependencies (Wang et al. 2024; Liu et al. 2025a). These methods provide structured priors for temporal reasoning, which are particularly effective when explicit relational information is available. Another line of work focuses on cross-modal alignment, mapping textual and numerical representations into a shared latent space through adaptive attention or alignment modules (Gao et al. 2025; Chowdhury, Akter, and Arib 2026). Third, semantic-guided methods further introduce language representations as contextual priors or auxiliary guidance signals for the forecasting network (Ding et al. 2025; Li et al. 2026). More recently, diffusion-based models have been explored to characterize predictive uncertainty and complex output distributions in time-series generation (Zhang et al. 2025; Liu et al. 2025b).

Nevertheless, most existing methods model cross-modal interaction through a sequence of discrete fusion operations, such as feature concatenation, cross-attention, or residual injection. In these designs, textual information mainly modifies intermediate representations at individual time steps, rather than explicitly governing how the latent temporal state evolves. Consequently, delayed and persistent textual effects must be captured implicitly through finite context windows, recurrent states, or increasingly deep fusion networks. This may be insufficient when the same textual event produces different responses depending on the system history, or when its influence emerges with a delay, persists over an extended horizon, and gradually decays or changes over time.

These characteristics are naturally analogous to non-Markovian open quantum-system dynamics. In quantum mechanics, the state of an open system is represented by a density operator whose evolution is jointly determined by its internal Hamiltonian and interactions with the surrounding environment (Chen, Pan, and Dong 2021). Under the Markovian assumption, the state transition depends only on the current density operator, whereas non-Markovian dynamics retain information from prior system–environment interactions through temporally correlated environmental effects. This provides a natural quantum-inspired formulation for multimodal forecasting: the numerical sequence is encoded as the evolving system state, while the dynamic text stream is

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

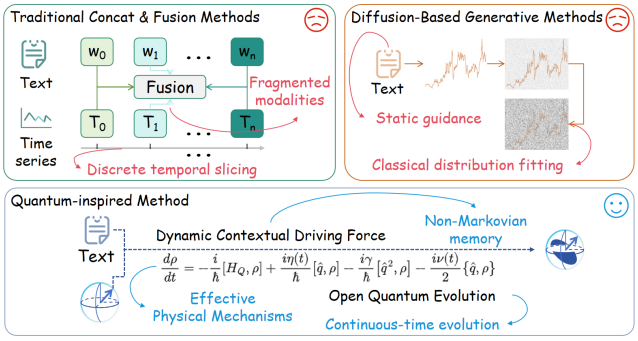


Figure 1: Conceptual comparison of multimodal time-series forecasting paradigms. Existing fusion and diffusion-based methods rely on discrete cross-modal integration or static textual guidance, whereas QMUSE models text as a dynamic contextual driving force governing continuous non-Markovian state evolution.

mapped to a time-dependent Hamiltonian that modulates its evolution. Stochastic and dissipative terms further describe how external influences may fluctuate, accumulate, decay or be redistributed over the forecasting horizon.

Motivated by this connection, we propose **QMUSE**, a **Q**uantum-inspired **Mu**ltimodal Time **S**eries **E**volution framework for text-conditioned forecasting. Instead of treating text as an auxiliary feature to be fused after temporal encoding, QMUSE incorporates textual information directly into the latent evolution law. Specifically, numerical observations are encoded as density operators in a high-dimensional Hilbert space, while the evolving textual context is transformed into a time-dependent Hamiltonian drive. The latent states are then propagated through a stochastic Liouville–von Neumann equation, which provides an explicit mechanism for modeling history-dependent and dissipative dynamics. Finally, a learnable quantum-measurement module decodes the evolved states into multi-step forecasting trajectories. The entire framework is differentiable, implemented on classical hardware, and trained end-to-end. As illustrated in Fig. 1, conventional approaches typically encode the numerical and textual modalities independently and exchange information through discrete fusion modules. In contrast, QMUSE treats the text stream as a time-varying external drive that continuously modulates the evolution of the numerical state. This formulation introduces an explicit inductive bias for delayed, persistent, and history-dependent cross-modal effects, while retaining the flexibility of end-to-end neural learning.

The main contributions are summarized as follows:

- **Quantum-inspired forecasting formulation.** We formulate multimodal time-series forecasting as a text-driven open-system evolution problem, in which numerical observations constitute the evolving system and dynamic textual context acts as an external driving signal. This formulation moves beyond treating text as a static auxiliary feature and provides an explicit representation of persistent cross-modal influence.
- **Text-driven non-Markovian evolution.** We develop QMUSE, which combines density-operator representation, a text-conditioned time-dependent Hamiltonian, stochastic Liouville–von Neumann evolution, and quantum-measurement decoding within a unified differentiable framework. The resulting model explicitly captures delayed responses, long-term memory, and dissipative temporal effects.
- **Extensive empirical evaluation.** We evaluate QMUSE on nine real-world multimodal time-series datasets. The results demonstrate consistent improvements over strong time-series and multimodal forecasting baselines in terms of predictive accuracy and robustness, while further analyses verify the complementary benefits of text-driven control and non-Markovian evolution.

Preliminaries

Quantum State

Pure State In quantum mechanics, the state of an isolated finite-dimensional system is represented by a unit vector in a complex Hilbert space $\mathcal{H} \simeq \mathbb{C}^n$. Using Dirac notation, the state vector is written as a ket $|\psi\rangle$, while its Hermitian adjoint is written as the corresponding bra, $\langle\psi| = |\psi\rangle^\dagger$. Given an orthonormal basis $\{|i\rangle\}_{i=0}^{n-1}$ of $\mathcal{H}$, a pure state can be expanded as

$$|\psi\rangle = \sum_{i=0}^{n-1} \alpha_i |i\rangle, \tag{1}$$

where $\alpha_i \in \mathbb{C}$ denotes the probability amplitude associated with the basis state $|i\rangle$. The normalization condition $\langle\psi|\psi\rangle = \sum_{i=0}^{n-1} |\alpha_i|^2 = 1$ ensures that the measurement probabilities are well defined. According to the Born rule, a projective measurement in the basis $\{|i\rangle\}_{i=0}^{n-1}$ yields the outcome i with probability $|\alpha_i|^2$. Equivalently, the pure state can be represented by the rank-one density operator $\rho_\psi = |\psi\rangle\langle\psi|$, which satisfies $\rho_\psi \succeq 0$, $\mathrm{Tr}(\rho_\psi) = 1$, and $\rho_\psi^2 = \rho_\psi$.

Mixed State A mixed state describes a statistical ensemble in which the system is prepared in one of several pure states with corresponding classical probabilities. Unlike a coherent superposition, a mixed state reflects classical uncertainty about the underlying state preparation. It is represented by a density operator

$$\rho = \sum_{i=1}^{m} p_i |\psi_i\rangle\langle\psi_i|, \tag{2}$$

where $|\psi_i\rangle \in \mathcal{H}$ is a normalized pure state and $p_i \geq 0$ is its associated probability, satisfying $\sum_{i=1}^{m} p_i = 1$. Consequently, a valid density operator is Hermitian, positive semidefinite, and trace-normalized:

$$\rho^\dagger = \rho, \qquad \rho \succeq 0, \qquad \mathrm{Tr}(\rho) = 1. \tag{3}$$

The purity of the state can be characterized by $\mathrm{Tr}(\rho^2)$. A state is pure if $\mathrm{Tr}(\rho^2) = 1$, whereas a genuinely mixed state satisfies $\mathrm{Tr}(\rho^2) < 1$.

Quantum Evolution

For an isolated quantum system, the density operator evolves unitarily according to the von Neumann equation. In contrast, an open quantum system exchanges information and energy with its environment, leading to reduced dynamics that may be non-unitary and history dependent. For a quantum system linearly coupled to a Gaussian reservoir, the environmental influence can be represented through an ensemble of time-local stochastic trajectories governed by the Stochastic Liouville–von Neumann (SLN) equation (Stockburger 2004).

Let $\rho_{\xi,\nu}(t)$ denote the trajectory-dependent system operator associated with a realization of the stochastic processes $\xi(t)$ and $\nu(t)$. The corresponding SLN equation is written as

$$\begin{aligned}\frac{d\rho_{\xi,\nu}(t)}{dt} = & -\frac{i}{\hbar}\left[H_{\mathrm{S}}(t), \rho_{\xi,\nu}(t)\right] + \frac{i\xi(t)}{\hbar}\left[\hat{q}, \rho_{\xi,\nu}(t)\right] \\ & -\frac{i\mu}{2\hbar}\left[\hat{q}^2, \rho_{\xi,\nu}(t)\right] + \frac{i\nu(t)}{2}\left\{\hat{q}, \rho_{\xi,\nu}(t)\right\},\end{aligned} \tag{4}$$

where i is the imaginary unit, $\hbar$ is the reduced Planck constant, and $H_{\mathrm{S}}(t)$ denotes the system Hamiltonian. The Hermitian operator $\hat{q}$ specifies the system degree of freedom coupled to the environment, while μ accounts for the static Hamiltonian renormalization induced by the system–environment interaction. The physical reduced density operator is recovered by averaging over stochastic trajectories:

$$\rho_{\mathrm{S}}(t) = \mathbb{E}_{\xi,\nu}\left[\rho_{\xi,\nu}(t)\right]. \tag{5}$$

Although each stochastic trajectory is governed by a time-local differential equation, the temporal correlations of $\xi(t)$ and $\nu(t)$ retain information from previous system–environment interactions. Consequently, the ensemble-averaged evolution can exhibit non-Markovian memory and dissipative behavior.

Quantum Measurement

A projective quantum measurement is defined by a set of mutually orthogonal projectors $\{P_k\}_{k=1}^{M}$ satisfying $P_k P_j = \delta_{kj} P_k$ and $\sum_{k=1}^{M} P_k = I$. For a non-degenerate measurement associated with an orthonormal basis $\{|k\rangle\}_{k=1}^{M}$, each projector is given by $P_k = |k\rangle\langle k|$. Given a density operator ρ, the probability of obtaining outcome k is determined by the Born rule:

$$p_k = \mathrm{Tr}(\rho P_k) = \mathrm{Tr}(\rho|k\rangle\langle k|) = \langle k|\rho|k\rangle, \tag{6}$$

where $p_k \geq 0$ and $\sum_{k=1}^{M} p_k = 1$. Conditioned on observing outcome k, the post-measurement state is

$$\rho_k = \frac{P_k \rho P_k}{p_k}, \tag{7}$$

provided that $p_k > 0$. In the proposed quantum-inspired framework, the probabilities $\{p_k\}_{k=1}^{M}$ are interpreted as classical readouts of the latent density operator and are subsequently used to construct the forecasting outputs.

Hermitian Parameterization

Hamiltonians and quantum observables are represented by Hermitian operators, whose eigenvalues are real. To construct a valid Hermitian operator from an unconstrained complex-valued matrix A, we use the parameterization

$$A_{\mathrm{herm}} = \frac{A + A^{\dagger}}{2}, \tag{8}$$

where $A^{\dagger}$ denotes the conjugate transpose of the matrix A. This construction guarantees $A_{\mathrm{herm}}^{\dagger} = A_{\mathrm{herm}}$. When A_{herm} is used as the Hamiltonian in the von Neumann term $-\frac{i}{\hbar}[A_{\mathrm{herm}}, \rho]$, it generates trace-preserving unitary evolution for a closed system. In QMUSE, this parameterization is imposed on the learnable Hamiltonian-related operators to maintain their required algebraic structure throughout optimization.

Related Work

Existing MTSF studies mainly follow three directions: structure-aware alignment (Wang et al. 2024; Liu et al. 2025a; Zhao et al. 2026), semantic-guided forecasting (Ding et al. 2025; Li et al. 2026; Jia et al. 2024), and generative diffusion modeling (Zhang et al. 2025; Liu et al. 2025b). These methods integrate external text with numerical sequences through structured representations, semantic priors, or probabilistic generation. In parallel, quantum-inspired approaches have shown promise in modeling complex dependencies and dynamical interactions (Qiao, Zhao, and Zhang 2025; Tang et al. 2026), but remain largely confined to unimodal numerical data. Their potential for text-conditioned multimodal forecasting therefore remains underexplored. A detailed review is provided in the **Supplementary Material**.

Methodology

Problem Formulation

Given a historical numerical sequence $\boldsymbol{X} \in \mathbb{R}^{l \times d_{\mathrm{in}}}$ and its associated textual context $\boldsymbol{S} \in \mathbb{R}^{k \times d_{\mathrm{txt}}}$, MTSF aims to predict the future numerical sequence $\boldsymbol{Y} \in \mathbb{R}^{h \times d_{\mathrm{out}}}$. Here, l and k denote the lengths of the numerical and textual sequences, respectively; d_{in} and d_{txt} are their feature dimensions; and h and d_{out} denote the forecasting horizon and target dimension. A forecasting model parameterized by θ learns the mapping

$$\hat{\boldsymbol{Y}} = g_{\theta}(\boldsymbol{X}, \boldsymbol{S}), \tag{9}$$

where $\hat{\boldsymbol{Y}} \in \mathbb{R}^{h \times d_{\mathrm{out}}}$ is the predicted sequence. The model parameters are optimized by minimizing a forecasting loss $\mathcal{L}(\hat{\boldsymbol{Y}}, \boldsymbol{Y})$ between the prediction and the ground truth.

Motivation

Real-world MTSF can be viewed as open dynamical systems driven by external context (Jiang et al. 2026). Textual events may induce delayed, persistent, decaying, and stochastic effects that are difficult to capture through static or discrete fusion. Inspired by open quantum-system theory, we encode numerical states as density operators and textual semantics as a time-dependent Hamiltonian, while stochastic and dissipative terms model environmental fluctuations and non-Markovian memory. This yields a principled framework for dynamic cross-modal interaction.

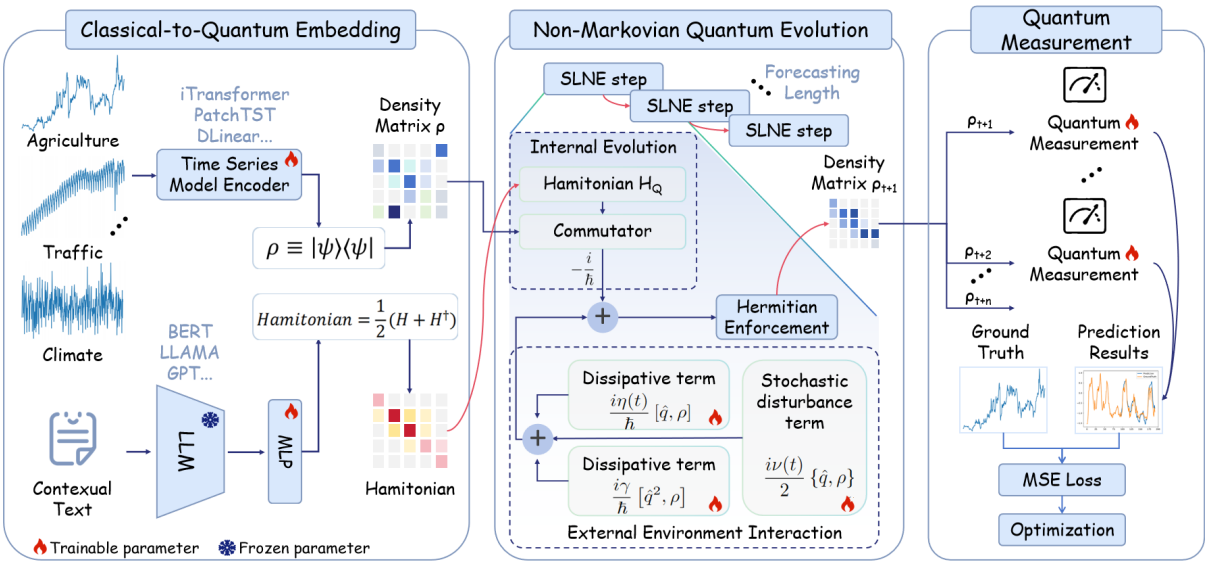


Figure 2: Overall architecture of the QMUSE framework. Numerical observations are encoded as latent density operators, while contextual text is transformed into a time-dependent Hamiltonian. The latent states are propagated through SLN-inspired non-Markovian evolution and decoded by learnable quantum measurements for multi-step forecasting.

Overall Framework

Figure 2 presents the overall framework of QMUSE, which consists of three modules:

- **Classical-to-Quantum Embedding.** The historical numerical sequence $\boldsymbol{X}$ is encoded as an initial latent density operator ρ_t, while the contextual text $\boldsymbol{S}$ is mapped to a text-conditioned Hamiltonian $H_Q(t)$ that governs the subsequent state evolution.
- **Non-Markovian Quantum Evolution.** Starting from ρ_t, the latent state is propagated over the forecasting horizon through a sequence of SLN-inspired evolution steps. The Hamiltonian term models internal system dynamics, whereas the stochastic and dissipative terms introduce history-dependent environmental effects.
- **Quantum Measurement.** The evolved density operators $\{\rho_{t+\tau}\}_{\tau=1}^{h}$ are projected through learnable measurement operators to produce the multi-step predictions $\hat{\boldsymbol{Y}}$. The entire framework is optimized end-to-end using the forecasting loss.

Classical-to-Quantum Embedding

Given the historical numerical sequence $\boldsymbol{X} \in \mathbb{R}^{l\times d_{\text{in}}}$, we first employ a time-series encoder, such as iTransformer (Liu et al. 2024b) or PatchTST (Nie et al. 2023), to extract a latent representation $\boldsymbol{z}_X$. A trainable complex-valued projection then maps $\boldsymbol{z}_X$ to an unnormalized state vector $\tilde{\psi} \in \mathbb{C}^D$, where D is the dimension of the latent Hilbert space. The normalized pure state and its corresponding density operator are constructed as

$$|\psi\rangle = \frac{|\tilde{\psi}\rangle}{\sqrt{\langle\tilde{\psi}|\tilde{\psi}\rangle}}, \qquad \rho_0 = |\psi\rangle\langle\psi|, \tag{10}$$

where $\rho_0 \in \mathbb{C}^{D\times D}$ satisfies $\rho_0^\dagger = \rho_0$, $\rho_0 \succeq 0$, and $\mathrm{Tr}(\rho_0) = 1$. The density operator ρ_0 serves as the initial latent state of the numerical system.

Meanwhile, the textual context $\boldsymbol{S} \in \mathbb{R}^{k\times d_{\text{txt}}}$ is treated as a time-dependent external driving signal. A frozen pretrained language model, such as BERT (Devlin et al. 2019) or LLaMA (Touvron et al. 2023), extracts its semantic representation $\boldsymbol{z}_S$, which is subsequently transformed by a trainable multilayer perceptron into a complex matrix $H_{\text{raw}}(t) \in \mathbb{C}^{D\times D}$. We define the text-conditioned Hamiltonian as

$$H_Q(t) = \frac{H_{\text{raw}}(t) + H_{\text{raw}}^\dagger(t)}{2}. \tag{11}$$

This construction maps the numerical history to an initial density operator and the textual context to a Hamiltonian drive, thereby placing both modalities within a unified quantum-inspired dynamical representation.

Non-Markovian Quantum Evolution

After embedding the multimodal inputs into the Hilbert space, QMUSE formulates forecasting as the evolution of the latent density operator over the prediction horizon. Starting from ρ_0, the state is propagated using an SLN-inspired dynamical equation that combines text-conditioned coherent evolution, environmental fluctuations, Hamiltonian renormalization, and history-dependent response. **Table** 1 summarizes the physical interpretation and modeling role of each term.

For a forecasting horizon of h, the evolution is discretized into h prediction steps. At the τ-th step, the state is updated as

$$\tilde{\rho}_{\tau+1} = \rho_\tau + \Delta t\, \mathcal{F}\left(\rho_\tau, H_Q(\tau), \hat{q}, \eta_\tau, \nu_\tau, \gamma\right), \tag{12}$$

where $\mathcal{F}(\cdot)$ denotes the right-hand side of Eq. 4 and Δt is the numerical integration step. This one-to-one correspondence

Mathematical Term	Physical Interpretation	Modeling Role
$-\frac{i}{\hbar}[H_Q(t), \rho]$	Hamiltonian-driven evolution	Text-conditioned state transition
$\frac{i\eta(t)}{\hbar}[\hat{q}, \rho]$	Environmental fluctuating force	Stochastic event perturbation
$-\frac{i\gamma}{\hbar}[\hat{q}^2, \rho]$	Hamiltonian renormalization	Nonlinear phase and trajectory drift
$-\frac{i\nu(t)}{2}\{\hat{q}, \rho\}$	Environmental response and back-action	History-dependent attenuation

Table 1: Physical terms in the SLN-inspired evolution and their roles in multimodal time-series forecasting.

between evolution steps and prediction steps is a modeling design rather than an identification of the forecasting interval with physical time.

The coupling operator $\hat{q}$ and the renormalization coefficient γ are learnable parameters optimized jointly with the forecasting network. To preserve their required algebraic structure, $\hat{q}$ is parameterized as a Hermitian operator. The stochastic variables η_τ and ν_τ represent environmental fluctuations and response effects. Importantly, non-Markovian memory is introduced through their temporal correlations rather than through independent Gaussian noise. In practice, their trajectories are generated from temporally correlated stochastic processes whose covariance structure is learnable or predefined.

Because Hermitian symmetrization and trace normalization alone do not guarantee a valid density operator, we project each updated state onto the set of positive semidefinite, trace-one matrices. Specifically, let

$$\rho_{\tau+1}^{\mathrm{H}} = \frac{\tilde{\rho}_{\tau+1} + \tilde{\rho}_{\tau+1}^{\dagger}}{2} = U \,\mathrm{diag}(\boldsymbol{\lambda}) U^{\dagger}. \tag{13}$$

The stabilized state is obtained as

$$\rho_{\tau+1} = \frac{U \,\mathrm{diag}(\max(\boldsymbol{\lambda}, \epsilon))\, U^{\dagger}}{\mathrm{Tr}[U \,\mathrm{diag}(\max(\boldsymbol{\lambda}, \epsilon))\, U^{\dagger}]}, \tag{14}$$

where $\epsilon > 0$ is a small numerical constant. This operation maintains Hermiticity, positive semidefiniteness, and unit trace throughout the recursive evolution. The module ultimately produces the latent state sequence $\{\rho_1, \rho_2, \ldots, \rho_h\}$ for subsequent prediction decoding.

Quantum Measurement

The evolved density operators $\{\rho_1, \rho_2, \ldots, \rho_h\}$ reside in a complex Hilbert space and therefore require a measurement-inspired readout to produce real-valued forecasts. We parameterize a set of learnable complex matrices and impose Hermitian parameterization to obtain the observables $\{O_i\}_{i=1}^{d_{\mathrm{out}}}$:

$$O_i = \frac{A_i + A_i^{\dagger}}{2}. \tag{15}$$

For the τ-th future step and the i-th target variable, the prediction is given by the expectation value of O_i with respect to ρ_τ:

$$\hat{Y}_{\tau,i} = \mathrm{Tr}(\rho_\tau O_i), \qquad \tau = 1, \ldots, h, \quad i = 1, \ldots, d_{\mathrm{out}}. \tag{16}$$

Since both ρ_τ and O_i are Hermitian, $\hat{Y}_{\tau,i}$ is real-valued. Unlike a physical projective measurement, this deterministic expectation-value readout does not sample or collapse the latent state; rather, it serves as a quantum-measurement-inspired decoder that maps the evolved density operators back to the classical forecasting space. The resulting prediction matrix $\hat{\boldsymbol{Y}} \in \mathbb{R}^{h \times d_{\mathrm{out}}}$ is optimized end-to-end against the ground truth using the forecasting loss.

Experiments

Datasets

We evaluate the proposed framework on Time-MMD datasets (including nine real-world datasets) (Liu et al. 2024a) across diverse domains. Following the prior works (Li et al. 2026; Du et al. 2026), we adopt consistent dataset settings, reporting the average performance. **Table** 2 presents detailed specifications for each domain.

Domain	Forecast Length	Dimension	Frequency	Samples
Agriculture	{6, 8, 10, 12}	1	Monthly	496
Climate	{6, 8, 10, 12}	5	Monthly	496
Economy	{6, 8, 10, 12}	3	Monthly	423
Energy	{12, 24, 36, 48}	9	Weekly	1479
Environment	{48, 96, 192, 336}	4	Daily	11102
Health	{12, 24, 36, 48}	11	Weekly	1389
Security	{6, 8, 10, 12}	1	Monthly	297
Social Good	{6, 8, 10, 12}	1	Monthly	900
Traffic	{6, 8, 10, 12}	1	Monthly	531

Table 2: Overview of the numerical data in the Time-MMD datasets. The tasks focus on univariate time series forecasting.

Baselines

To comprehensively assess the proposed model, we select a diverse array of baseline models for comparison, including LCMF (Du et al. 2026), Cart (Gu et al. 2026), TaTS (Li et al. 2026), DDATG (Lei et al. 2025), DuaSG (Ding et al. 2025), UniDiff (Zhang et al. 2025), ChatTime (Wang et al. 2025), and GPT4MTS (Jia et al. 2024).

In addition, to evaluate the compatibility of the proposed model, we integrate it with 5 widely used models from different categories of time series, including Transformer-based models: iTransformer (Liu et al. 2024b), PatchTST (Nie et al. 2023), Crossformer (Zhang and Yan 2023); Linear-based models: DLinear (Zeng et al. 2023); Frequency-based models: FEDformer (Zhou et al. 2022).

Hyperparameter	Description	Value
batch_size	The batch size for training	32
learning_rate	The learning rate for the optimizer	1e-4
seq_len	Input sequence length	24
label_len	Start token length for prediction	12
train_epochs	Number of training epochs	30
patience	Early stopping patience	5
dropout	Dropout	0.1
hilbert_dim	Dimension of the quantum Hilbert space	64
text_dim	Dimension of raw text embeddings	1024
dt	Moderate discretization step size for SLNE	0.05

Table 3: Hyperparameter settings.

Methods	QMUSE(Proposed)		LCMF(2026)		Cart(2026)		TaTS (2026)		DDATG(2026)		DuaSG(2025)		UniDiff(2025)		ChatTime (2025)		GPT4MTS (2024)	
Datasets	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Agriculture	**0.085**	**0.195**	0.102	0.22	0.109	0.226	0.109	0.229	0.087	0.198	0.315	0.369	0.213	0.294	0.508	0.447	0.327	0.393
Climate	0.976	**0.759**	1.01	0.802	**0.975**	0.764	1.028	0.804	1.198	0.892	1.031	0.825	1.063	0.747	1.568	1.019	1.127	0.873
Economy	**0.007**	**0.073**	0.025	0.124	0.009	0.078	0.008	0.077	0.017	0.105	0.014	0.094	0.018	0.101	0.049	0.166	0.014	0.096
Energy	0.234	0.351	**0.195**	**0.319**	0.245	0.357	0.265	0.376	/	/	0.258	0.37	0.235	0.338	0.305	0.417	0.269	0.378
Environment	**0.255**	**0.362**	/	/	0.27	0.373	0.267	0.369	0.415	0.479	0.343	0.409	0.279	0.382	0.58	0.594	0.348	0.422
Health	1.251	0.738	1.355	0.771	**1.199**	**0.727**	1.315	0.744	/	/	1.63	0.851	1.353	0.778	1.668	0.909	1.766	0.875
Security	108.72	4.853	111.21	4.964	77.49	4.321	112.05	5.151	115.64	5.434	**72.52**	**4.139**	/	/	133.11	6.887	119.41	5.198
Social Good	0.883	0.412	**0.749**	**0.37**	0.919	0.443	0.987	0.452	1.174	0.454	1.052	0.455	0.854	0.452	1.264	0.652	1.414	0.559
Traffic	**0.156**	**0.198**	0.156	0.233	0.169	0.205	0.187	0.217	0.199	0.251	0.166	0.224	0.187	0.288	0.363	0.429	0.195	0.246
Best Count	4	5	2	2	2	1	0	0	0	0	1	1	0	0	0	0	0	0

Table 4: Overall performance comparison of QMUSE against state-of-the-art baselines across nine multimodal time-series datasets. The best and second-best results are highlighted in **red (bold)** and blue (underlined), respectively.

Experimental Settings

The experiments were conducted using PyTorch with Python 3.11.4 on an NVIDIA 4090 GPU. To keep consistent with previous works, we adopt Mean Squared Error (MSE) and Mean Absolute Error (MAE) as evaluation metrics and Adam as the optimizer. The default choices of hyperparameters are shown in **Table** 3. A detailed sensitivity analysis of hyperparameters is provided in the **Supplementary Material**.

Mainstream Baseline Comparison

As illustrated in **Table** 4, we compare the proposed QMUSE against several mainstream independent baselines. QMUSE achieves the best MSE on four datasets and the best MAE on four datasets, while remaining competitive across the broader benchmark. The improvements are particularly pronounced on Agriculture, Economy, and Traffic, whereas the results on Energy, Health, Security, and Social Good indicate that the benefit of text-driven evolution varies across domains. Overall, the results demonstrate a strong cross-domain performance profile rather than dominance on every individual dataset. Statistical significance and stability results are provided in the **Supplementary Material**.

Backbone Enhancement Comparison

We further integrate QMUSE with five representative forecasting backbones, including Transformer-, linear-, and frequency-based architectures. Across the evaluated backbone–dataset combinations, QMUSE improves or matches the corresponding TaTS-enhanced models in the majority of settings, with particularly consistent gains for iTransformer, PatchTST, and DLinear. The complete results are reported in **Supplementary Material**.

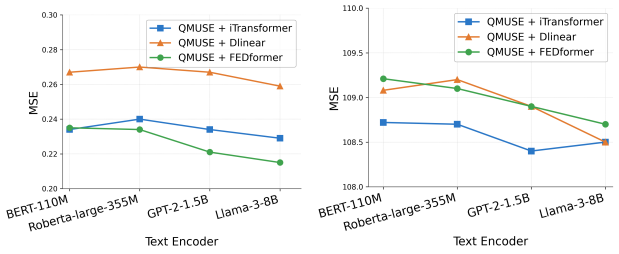
Discussion

Ablation Analysis

Ablation on Core Components Ablation results in **Table** 5 indicate that all components contribute to the overall performance. Specifically, omitting the Hermitian enforcement or text-driven Hamiltonian causes severe performance degradation on the less regular Security and Climate datasets, confirming their decisive roles in ensuring numerical stability and transmitting cross-modal driving forces. Conversely, removing the SLNE stochastic term exerts a negligible impact on the highly regular Traffic dataset but significantly degrades Security, strongly validating the efficacy of this stochastic mechanism in capturing sudden environmental perturbations.

Dataset	Climate		Security		Traffic	
Methods/Metric	MSE	MAE	MSE	MAE	MSE	MAE
QMUSE w/o Text-driven Hamiltonian	1.113	0.829	117.64	5.543	0.192	0.236
QMUSE w/o SLNE dissipative term	1.065	0.801	115.4	5.199	0.175	0.211
QMUSE w/o SLNE stochastic term	0.998	0.764	112.37	5.186	0.157	0.198
QMUSE w/o Hermitian enforcement	1.125	0.831	120.96	5.735	0.184	0.216
QMUSE w/o Quantum measurement	1.053	0.794	110.91	4.927	0.165	0.201
QMUSE (full)	**0.976**	**0.759**	**108.72**	**4.853**	**0.156**	**0.198**

Table 5: Core components ablation test. The best and second-best results are highlighted in **red (bold)** and blue (underlined), respectively.



(a) Energy. (b) Security.

Figure 3: Performance comparison of different text encoders.

Ablation on Different Text Encoders To evaluate the impact of different text encoders, we select three temporal backbone networks with different technical frameworks and conduct validation on two datasets with distinct data scales, as shown in **Figure** 3. The minimal fluctuation in forecasting errors across model scales from BERT-110M to Llama3-8B demonstrates that our framework does not rely on massive representational capacity. Instead, it effectively captures continuous cross-modal dynamical interactions using only lightweight semantic embeddings. The complete results are reported in **Supplementary Material**.

Robustness Analysis

To evaluate model robustness against multimodal data missingness or noise perturbations, we conduct random text

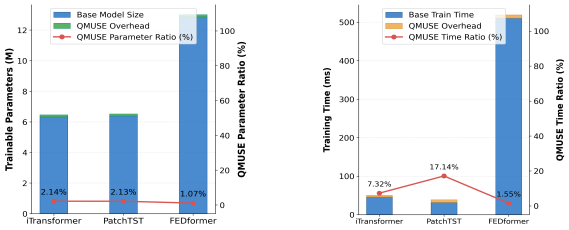
Dataset	Climate		Security		Traffic	
Methods/Metric	MSE	MAE	MSE	MAE	MSE	MAE
MM-TSFLib+ original data	1.044	0.81	116.34	5.532	0.199	0.347
TaTS + original data	1.028	0.804	112.05	5.151	0.187	0.217
QMUSE + original data	0.976	0.759	108.72	4.853	0.156	0.198
MM-TSFLib + 10% text dropped	1.058	0.814	117.03	5.585	0.207	0.366
TaTS + 10% text dropped	1.032	0.809	112.94	5.372	0.194	0.239
QMUSE + 10% text dropped	0.997	0.764	109.53	4.935	0.162	0.201
MM-TSFLib + 25% text dropped	1.082	0.822	119.67	5.64	0.225	0.381
TaTS + 25% text dropped	1.056	0.818	114.59	5.231	0.203	0.254
QMUSE + 25% text dropped	1.031	0.793	111.62	5.017	0.168	0.213

Table 6: Robustness evaluation under different dropout rates.

dropout tests on three datasets, as shown in **Table** 6. Although increasing text dropout rates degrade the performance of all models, our method consistently outperforms or matches baselines operating on completely intact data, which further validates the trajectory reconstruction capability of the non-Markovian quantum evolution mechanism.

Computational Analysis

Empirical evaluations on efficiency highlight the lightweight nature of the QMUSE, shown in **Figure** 4. Across diverse architectures, it maintains a bounded parameter footprint and invariant latency, and its relative resource consumption is thus rapidly marginalized as the backbone scales, which validates that the module can seamlessly endow diverse mainstream models with quantum-enhanced features at near-zero computational and memory overhead.



(a) Trainable Parameters. (b) Training Time.

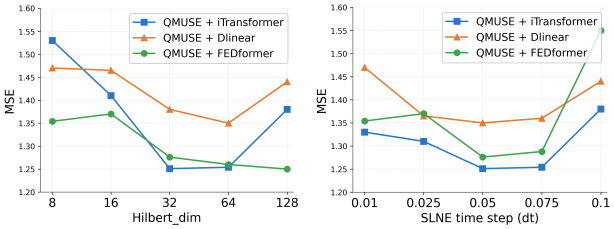
Figure 4: Computational overhead of QMUSE across different forecasting backbones.

Hyper-parameter Sensitivity Analysis

As shown in **Figure** 5, we evaluate two key hyper-parameters on the Health dataset. The Hilbert dimension of 64 best balances representation capacity against overfitting, while an SLNE step size of 0.05 prevents discretization errors to ensure numerical stability during evolution. The complete results are reported in **Supplementary Material**.

Case Study

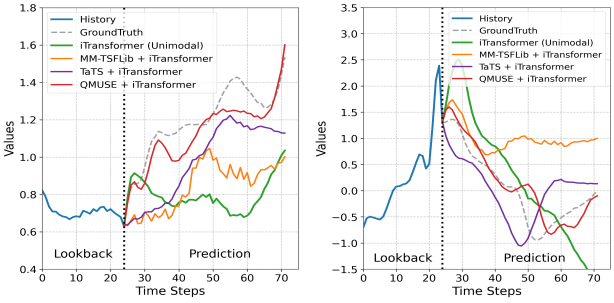
As shown in **Figure** 6, QMUSE better preserves the global target trajectory than competing baselines. In the left example, it captures the upward transition and final sharp increase, while other methods underestimate the long-term trend or accumulate drift. In the right example, QMUSE more closely



(a) (b)

Figure 5: Parameter study about $hilbert_dim$ and dt.

follows the rise-decline-recovery pattern, whereas several baselines overshoot, miss the negative regime, or progressively diverge. These results indicate that QMUSE is effective in modeling nonlinear turning points and long-horizon cross-modal dynamics.



(a) Energy. (b) Health.

Figure 6: Qualitative forecasting comparison on representative samples from the Energy and Health datasets.

Conclusion

We presented QMUSE, a quantum-inspired framework that formulates multimodal time-series forecasting as latent-state evolution under open-system dynamics. QMUSE represents numerical observations as density operators, maps contextual text to a time-dependent Hamiltonian drive, and propagates the latent states through SLN-inspired dynamics to capture persistent, stochastic, and history-dependent cross-modal effects. Experiments on nine real-world datasets show that QMUSE consistently outperforms strong unimodal and multimodal baselines in forecasting accuracy and robustness. These results further indicate that incorporating text into the state-transition dynamics provides a promising alternative to conventional feature-level multimodal fusion.

The current framework models a unidirectional interaction in which textual context drives the evolution of the numerical state. However, many real-world systems like market movements exhibit reciprocal cross-modal dependencies. Extending QMUSE to bidirectionally coupled open-system dynamics, in which numerical and textual states evolve jointly, is therefore an important direction for future work.

References

Chen, Y.; Pan, Y.; and Dong, D. 2021. Quantum language model with entanglement embedding for question answering. *IEEE Transactions on Cybernetics*, 53(6): 3467–3478.

Chowdhury, A. M.; Akter, R.; and Arib, S. H. 2026. T3time: Tri-modal time series forecasting via adaptive multi-head alignment and residual fusion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, 20597–20605.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies*, 4171–4186.

Ding, K.; Fan, F.; Wang, Y.; Jian, R.; Wang, X.; Gong, L.; Jiang, Y.; Luo, C.; and Zhan, J. 2025. Dualsg: A dual-stream explicit semantic-guided multivariate time series forecasting framework. In *Proceedings of the 33rd ACM International Conference on Multimedia*, 508–517.

Du, S.; Jiang, J.; Shi, H.; Yang, Y.; Zhang, J.; Li, T.; and Zheng, Y. 2026. LCMF: A LLM-enhanced cross-modal fusion framework for universal temporal forecasting. *Knowledge-Based Systems*, 115804.

Gao, M.; Zhou, S.; Gu, W.; Wu, Z.; Liu, H.; Zhou, A.; and Wang, X. 2025. MMGPT4LF: Leveraging an optimized pre-trained GPT-2 model with multi-modal cross-attention for load forecasting. *Applied Energy*, 392: 125965.

Gu, H.; Xiong, H.; Huang, K.; Zhu, W.; Dai, Y.; He, Z.; Jing, Y.; Zhang, K.; and Wang, X. S. 2026. Cart: A Future-Oriented Text Impacts Alignment Framework for Time Series Forecasting. In *International Conference on Database Systems for Advanced Applications*, 237–252. Springer.

Jia, F.; Wang, K.; Zheng, Y.; Cao, D.; and Liu, Y. 2024. Gpt4mts: Prompt-based large language model for multimodal time-series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, 23343–23351.

Jiang, Y.; Ning, K.; Pan, Z.; Shen, X.; Ni, J.; Yu, W.; Schneider, A.; Chen, H.; Nevmyvaka, Y.; and Song, D. 2025. Multimodal time series analysis: A tutorial and survey. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*, 6043–6053.

Jiang, Y.; Yu, W.; Lee, G.; Song, D.; Shin, K.; Cheng, W.; Liu, Y.; and Chen, H. 2026. TimeXL: Explainable multi-modal time series prediction with LLM-in-the-loop. *Advances in Neural Information Processing Systems*, 38: 135431–135467.

Lei, J.; Zhang, W.; Yang, Q.; Zhang, X.; and Tang, H. 2025. Transformer-PLM Enhanced Multimodal Time Series Forecasting via Decoupled Dual-Temporal Graph Adaptation. *IEEE Signal Processing Letters*, 33: 11–15.

Li, Z.; Lin, X.; Liu, Z.; Zou, J.; Wu, Z.; Zheng, L.; Fu, D.; Zhu, Y.; Hamann, H.; Tong, H.; and He, J. 2026. Language in the Flow of Time: Time-Series-Paired Texts Weaved into a Unified Temporal Narrative. In *The Fourteenth International Conference on Learning Representations*.

Liu, C.; Hettige, K. H.; Xu, Q.; Long, C.; Xiang, S.; Cong, G.; Li, Z.; and Zhao, R. 2025a. St-llm+: Graph enhanced spatio-temporal large language models for traffic prediction. *IEEE Transactions on Knowledge and Data Engineering*.

Liu, H.; Xu, S.; Zhao, Z.; Kong, L.; Kamarthi, H.; Sasanur, A. B.; Sharma, M.; Cui, J.; Wen, Q.; Zhang, C.; et al. 2024a. Time-mmd: Multi-domain multimodal dataset for time series analysis. *Advances in Neural Information Processing Systems*, 37: 77888–77933.

Liu, Y.; Hu, T.; Zhang, H.; Wu, H.; Wang, S.; Ma, L.; and Long, M. 2024b. itransformer: Inverted transformers are effective for time series forecasting. In *International conference on learning representations*, volume 2024, 11116–11140.

Liu, Y.; Wijewickrema, S.; Hu, D.; Bester, C.; O'Leary, S.; and Bailey, J. 2025b. Stochastic diffusion: A diffusion based model for stochastic time series forecasting. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*, 1939–1950.

Liu, Y.; Zhong, S.; Luo, G.; Ruan, W.; and Liang, Y. 2025c. Towards Multi-Scenario Forecasting of Building Electricity Loads with Multimodal Data. In *Proceedings of the 33rd ACM International Conference on Multimedia*, 2188–2196.

Naisipour, M.; Saeedpanah, I.; and Adib, A. 2025. Multimodal deep learning for two-year ENSO forecast. *Water Resources Management*, 39(8): 3745–3775.

Nie, Y.; Nguyen, N. H.; Sinthong, P.; and Kalagnanam, J. 2023. A Time Series is Worth 64 Words: Long-term Forecasting with Transformers. In *The Eleventh International Conference on Learning Representations*.

Qiao, W.; Zhao, J.; and Zhang, P. 2025. Quantum time-index models with reservoir for time series forecasting. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 1*, 1173–1184.

Stockburger, J. T. 2004. Simulating spin-boson dynamics with stochastic Liouville–von Neumann equations. *Chemical physics*, 296(2-3): 159–169.

Tang, Y.; Cai, Z.; Zhang, Y.; Gao, Z.; and Yu, J. 2026. Qnnformer: quantum-enhanced neural networks with multi-attention mechanisms for long-term time series forecasting: Y. Tang et al. *Quantum Information Processing*, 25(2): 66.

Touvron, H.; Martin, L.; Stone, K.; Albert, P.; Almahairi, A.; Babaei, Y.; Bashlykov, N.; Batra, S.; Bhargava, P.; Bhosale, S.; et al. 2023. Llama 2: Open foundation and fine-tuned chat models. arXiv:2307.09288.

Wang, C.; Qi, Q.; Wang, J.; Sun, H.; Zhuang, Z.; Wu, J.; Zhang, L.; and Liao, J. 2025. Chattime: A unified multimodal time series foundation model bridging numerical and textual data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 12694–12702.

Wang, J.; Sun, K.; Luo, L.; Wei, W.; Hu, Y.; Liew, A. W.; Pan, S.; and Yin, B. 2024. Large language models-guided dynamic adaptation for temporal knowledge graph reasoning. *Advances in Neural Information Processing Systems*, 37: 8384–8410.

Zeng, A.; Chen, M.; Zhang, L.; and Xu, Q. 2023. Are transformers effective for time series forecasting? In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, 11121–11128.

Zhang, D.; Li, B.; Zhao, Z.; Gao, J.; Nie, F.; and Li, X. 2025. UniDiff: A Unified Diffusion Framework for Multimodal Time Series Forecasting. arXiv:2512.07184.

Zhang, Y.; and Yan, J. 2023. Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting. In *The eleventh international conference on learning representations*.

Zhao, F.; Liu, K.; Peng, T.; Yang, Y.; and Xu, G. 2026. DyMRL: Dynamic Multispace Representation Learning for Multimodal Event Forecasting in Knowledge Graph. In *Proceedings of the ACM Web Conference 2026*, 4232–4243.

Zhou, T.; Ma, Z.; Wen, Q.; Wang, X.; Sun, L.; and Jin, R. 2022. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *International conference on machine learning*, 27268–27286.