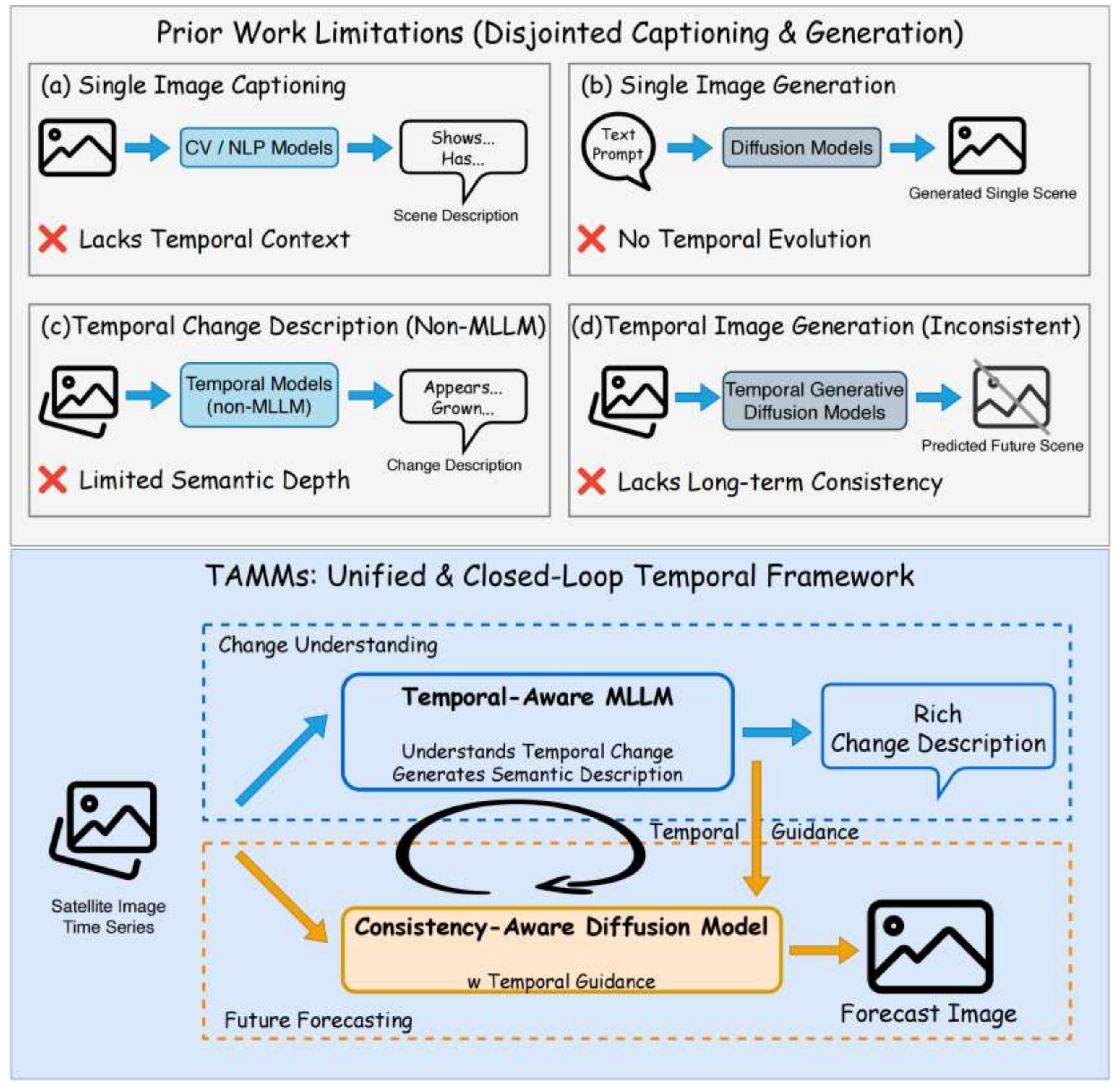
TAMMS: Change Understanding and Forecasting in Satellite Image Time Series with Temporal-Aware Multimodal Models

Zhongbin Guo[1], Yuhao Wang[1], Ping Jian[1*], Chengzhi Li[1], Xinyue Chen[1], Zhen Yang[1], Ertai E[2]
[1]School of Computer Science & Technology, Beijing Institute of Technology [2]School of Computing, National University of Singapore

1.Motivation & Challenge



2.Main Contribution

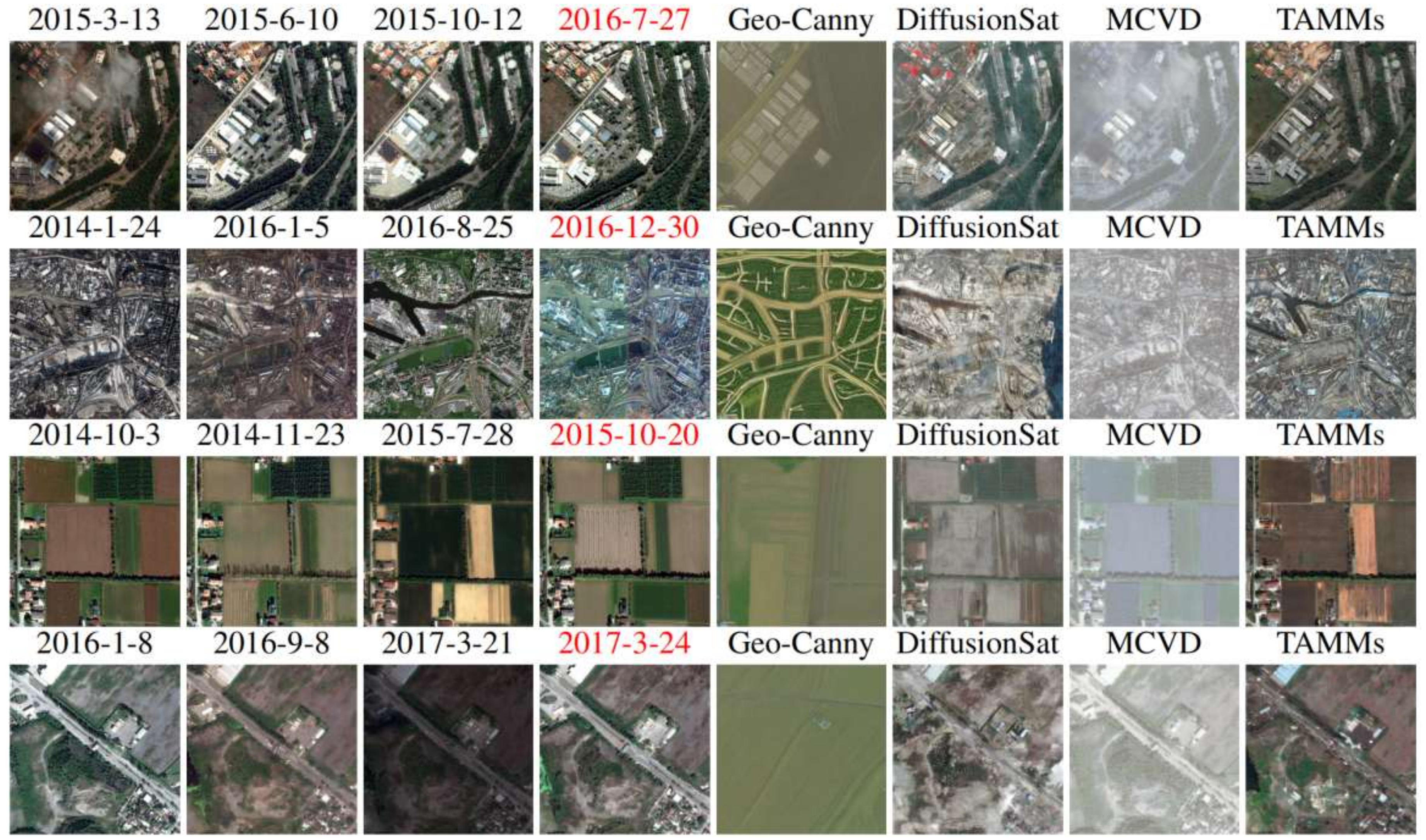
① We propose **TAMMs**, the first unified framework to synergistically couple temporal change understanding with future forecasting in a mutually reinforcing manner.

② We introduce **TAM** to awaken latent long-range temporal reasoning abilities of MLLM for SITS.

③ We design a novel **SFCI** mechanism to translate the MLLM's high-level temporal understanding into fine-grained generative control.

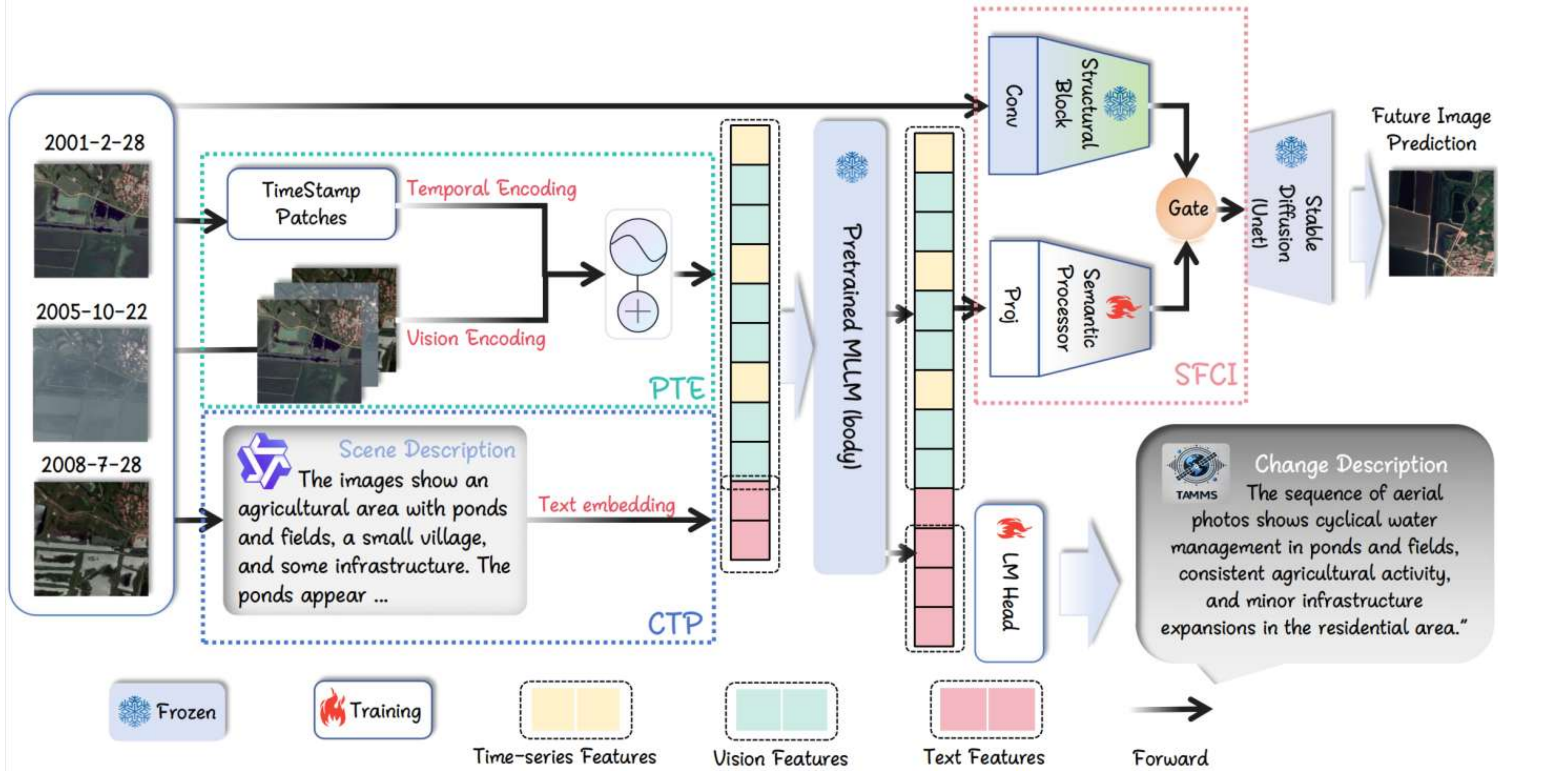
④ We design the **TCS**, the first metric designed to specifically quantify the consistency of predicted changes against historical dynamics.

4. Generated samples from our temporal prediction task



Left columns are input sequences and target (with red dates indicating the target images to be predicted); right are predictions from different models. It can be clearly seen that TAMMs well captures the changing trends in the sequence and also ensures good image quality.

3. The TAMMs framework architecture



5. Experiments & Results

Model	Temporal Change Description				Future Satellite Image Forecasting			
	B-4↑	M↑	R-L↑	C-D↑	P↑	S↑	L↓	TCS↑
Baselines for Temporal Change Description								
RSICC-Former Liu et al. (2022)	0.1285	0.1930	0.3489	0.5344	-	-	-	-
SITSCC Peng et al. (2024)	0.2122	0.2961	0.4701	0.6244	-	-	-	-
Qwen2.5-VL-7B Bai et al. (2025)	0.2089	0.2845	0.4612	0.7156	-	-	-	-
DeepSeek-VL2 Wu et al. (2024)	0.1920	0.2382	0.4603	0.7362	-	-	-	-
TEOChat Irvin et al. (2025)	0.2398	0.3102	**0.4735**	0.8267	-	-	-	-
Baselines for Future Satellite Image Forecasting								
Geosynth-Canny Sastry et al. (2024)	-	-	-	-	11.3752	0.1757	0.7559	0.2170
MCVD Voleti et al. (2022)	-	-	-	-	9.2208	**0.2098**	0.4970	0.1930
DiffusionSat Khanna et al. (2024)	-	-	-	-	11.8878	0.1520	0.5225	0.7624
Our Unified Model								
TAMMs (Ours)	**0.2669**	**0.3312**	0.4690	**0.9030**	**12.0697**	0.1831	**0.4931**	**0.9690**

Table 1:**Comprehensive evaluation on Temporal Change Description and Future Satellite Image Forecasting tasks**. Metrics are abbreviated: B-4 (BLEU-4), M (METEOR), R-L (ROUGE-L), C-D (CIDEr-D), P (PSNR), S (SSIM), L (LPIPS). Models are grouped by their primary task. A dash (-) indicates the metric is not applicable for that model. Bold is best; underline is second best.

Model / Config	Change Description Metrics				Image Prediction Metrics			
	B-4↑	M↑	R-L↑	C-D↑	P↑	S↑	L↓	TCS↑
Ablation on Temporal Adaptation (for Change Description)								
SFT only	0.2134	0.2901	0.4634	0.7523	-	-	-	0.6842
w/o PTE	0.2387	0.3089	**0.4721**	0.8234	-	-	-	0.7456
w/o CTP	0.2445	0.3156	0.4678	0.8567	-	-	-	0.8234
Ablation on MLLM Integration (for Image Prediction)								
w/o Semantic Fusion	-	-	-	-	12.0642	0.1596	0.5198	0.7911
w/o Text Guidance	-	-	-	-	11.9655	0.1709	0.5065	0.9410
Base Control Block	-	-	-	-	11.8878	0.1520	0.5225	0.7624
Our Full Model								
TAMMs (Full)	**0.2669**	**0.3312**	0.4690	**0.9030**	**12.0697**	**0.1831**	**0.4931**	**0.9690**

Table 2:Ablation studies for both tasks

6. Conclusion and Future Work

In this paper, we addressed the fragmented nature of SITS, where temporal change description (TCD) and future forecasting (FSIF) are treated as disjointed tasks limited by a common bottleneck in long-range temporal understanding. We proposed TAMMs, the first unified framework designed to synergistically solve both, built on core hypothesis that empowering MLLM to deeply comprehend historical dynamics would enable more consistent future forecasting. Our comprehensive experiments validate that TAMMs achieves state-of-the-art performance on both tasks, significantly outperforming specialist baselines. Critically, its substantial gains on our proposed TCS metric provide strong evidence that unified approach leads to forecasts are demonstrably more consistent with historical evolution, establishing a new paradigm for reasoning-based SITS analysis.

Building on this work, several promising research avenues are identified. To enhance forecasting over very long-term horizons and for rare, abrupt events, future work will explore hierarchical temporal modeling and the integration of external data sources. To improve the framework's efficiency and reliability for real-world applications, we will also investigate model compression techniques and incorporate uncertainty quantification (e.g., through Bayesian approaches) to move beyond deterministic predictions and provide richer forecasting outputs.