Diffusion-based Decoupled Deterministic and Uncertain Framework for Probabilistic Multivariate Time Series Forecasting

Qi Li[1*], Zhenyu Zhang[1*], Lei Yao[1], Zhaoxia Li[3], Tianyi Zhong[4], Yong Zhang[1,2]

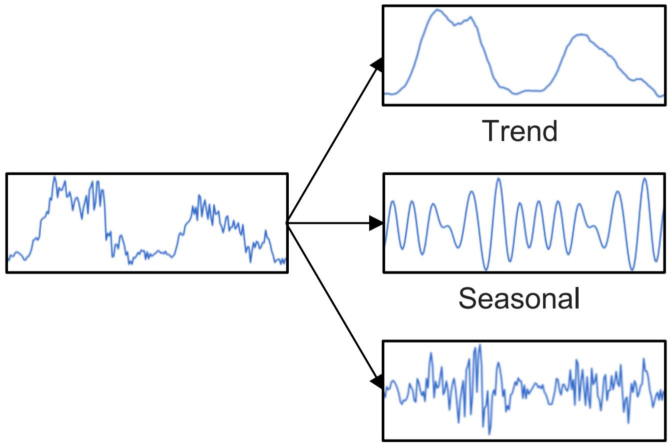
[1]Beijing University of Posts and Telecommunications
[2]Beijing Key Laboratory of Work Safety Intelligent Monitoring
[3]China Unicom Cloud Data Co., Ltd.
[4]University of Melbourne

Paper ID 7497

Background

Objective of **probabilistic multivariate time series (MTS) forecasting**:

Given history MTS data $x_{1:H} = \{x_1, x_2, \dots, x_H\} \in \mathbb{R}^{C \times H}$, probabilistic MTS forecasting tackles the problem of **estimating the distribution** of the subsequent future time series $y_{1:L} = \{p(y_1), p(y_2), \dots, p(y_L)\} \in \mathbb{R}^{C \times L}$.



Uncertainty in point forecasting:

$$X_{nf}: X = X_{nf} + \epsilon_X.\ X_{nf}$$

ϵ_X : the inherent noise part

X_{nf} : the ideal noise-free part contains components with clear temporal patterns, such as trend components X_T and seasonal components X_S. In an ideal scenario, the inherent noise ϵ_X corresponds to the residual component, X_R.

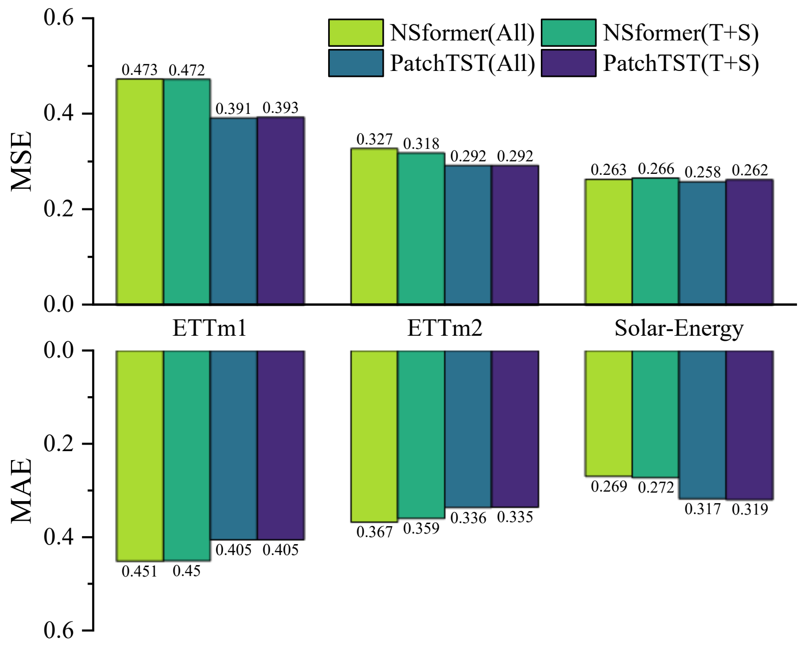
$$X_R = X_r + \epsilon'_X$$

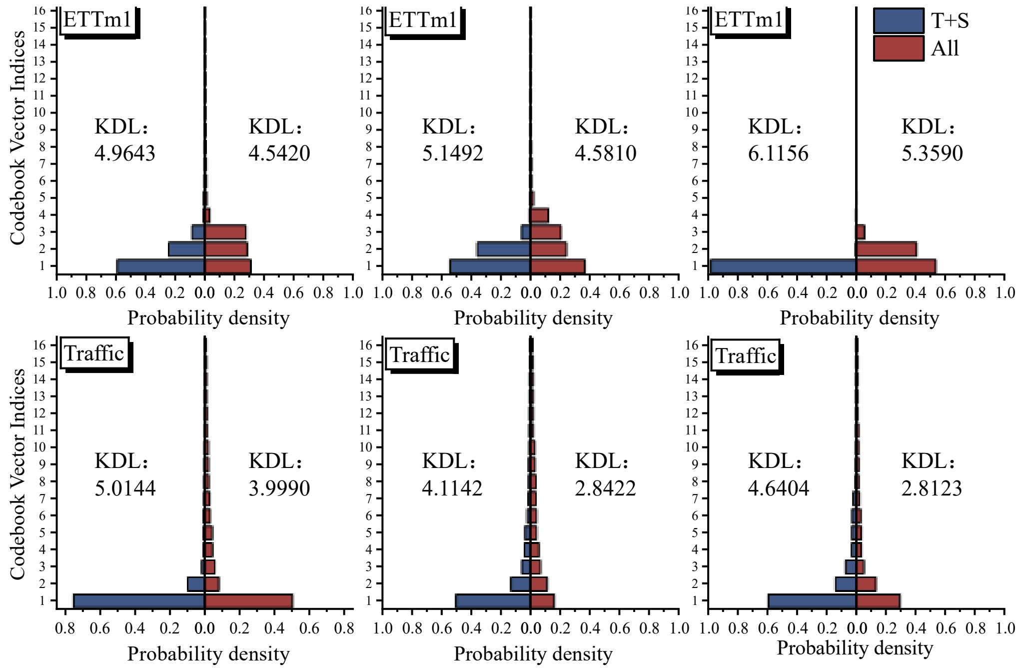
X_r : the non-noise component of X after removing the evident trend and seasonal information,

ϵ'_X : the noise component of ϵ_X, excluding the noise that remains in X_T and X_S.

Petropoulos, Fotios, et al. "Forecasting: theory and practice." *International Journal of forecasting* 38.3 (2022): 705-871.

Motivation



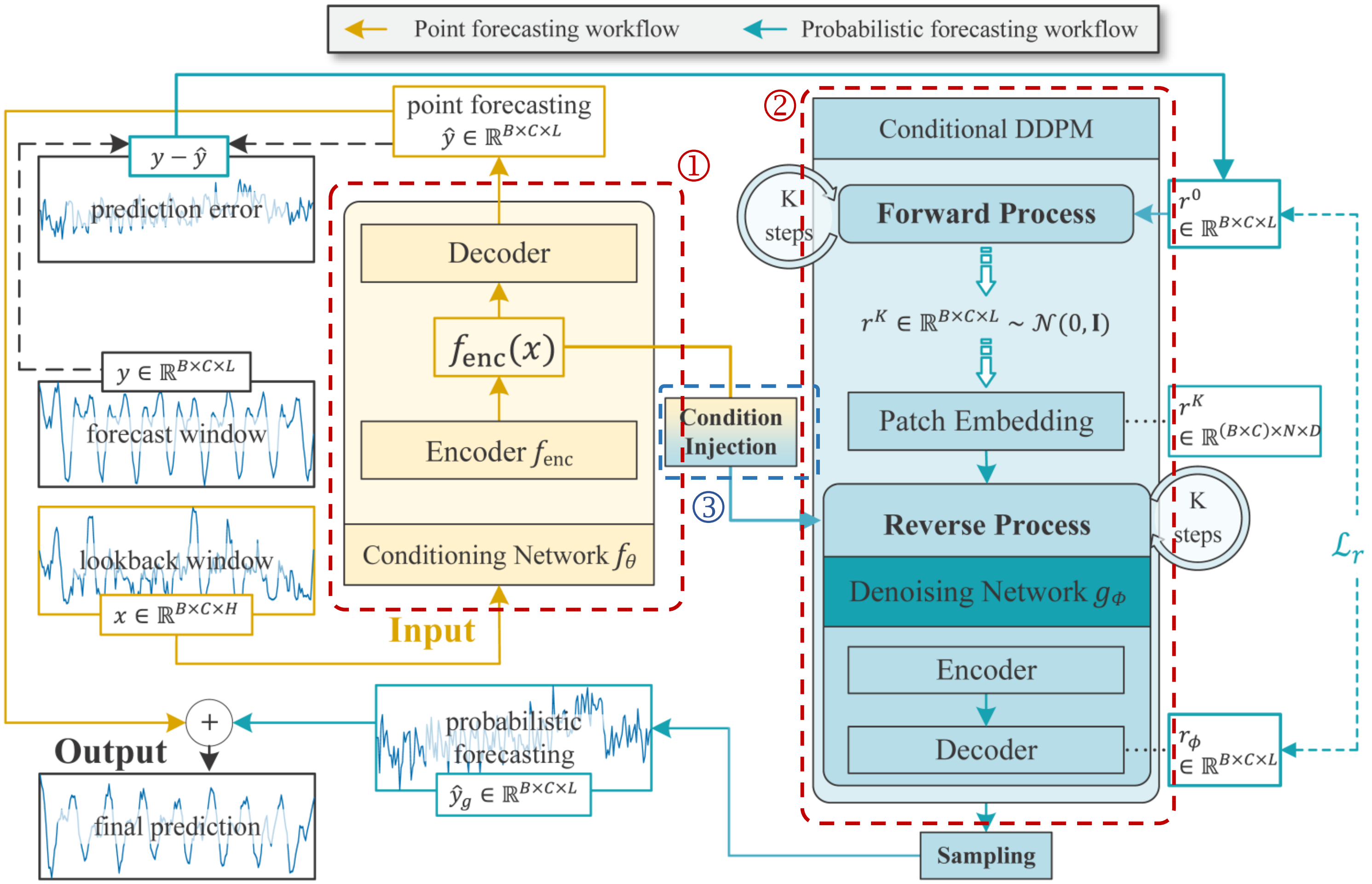


- **Point forecasting** models exhibit **varying** capabilities in modeling $X_T + X_S$ and X_r.
- **Residual** components X_r have **higher uncertainty**.

Divide-and-conquer

Decoupling deterministic and uncertain components

Framework



① Point forecasting model ② Conditional DDPM ③ Condition injection

Diffusion-based Decoupled Deterministic and Uncertain (D3U) framework:

High deterministic components modeled by pretrained point forecasting model + Components with **high-uncertainty** modeled by the conditional denoising diffusion probabilistic model (DDPM) =

Better probabilistic forecasts

Framework

Conditional network f_θ:

Extracting useful information from input series x_h.

Conditional DDPM:

Prediction error $r^0_{1:L}$

- Modeling the distribution of residual components in the prediction target
- Prediction error of the conditioning network is used as the residual:

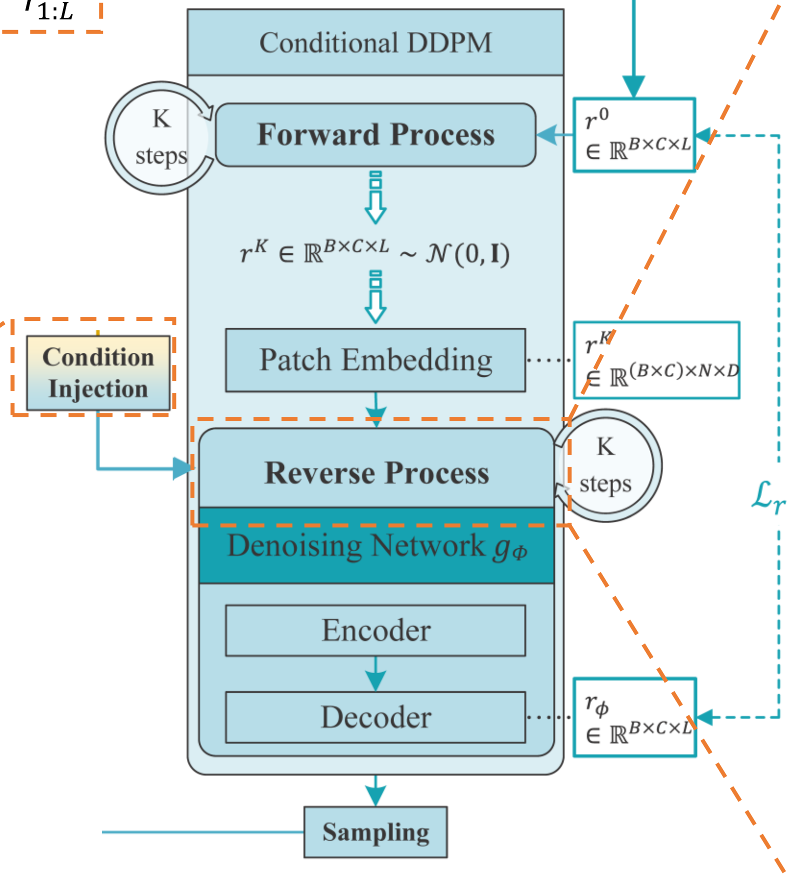
$$r^0_{1:L} \coloneqq y - \hat{y} = y - f_\theta(x_h)$$

- Condition provided by the encoder of f_θ:

$$c = f_{\text{enc}}(x_h)$$

- Residual prediction:

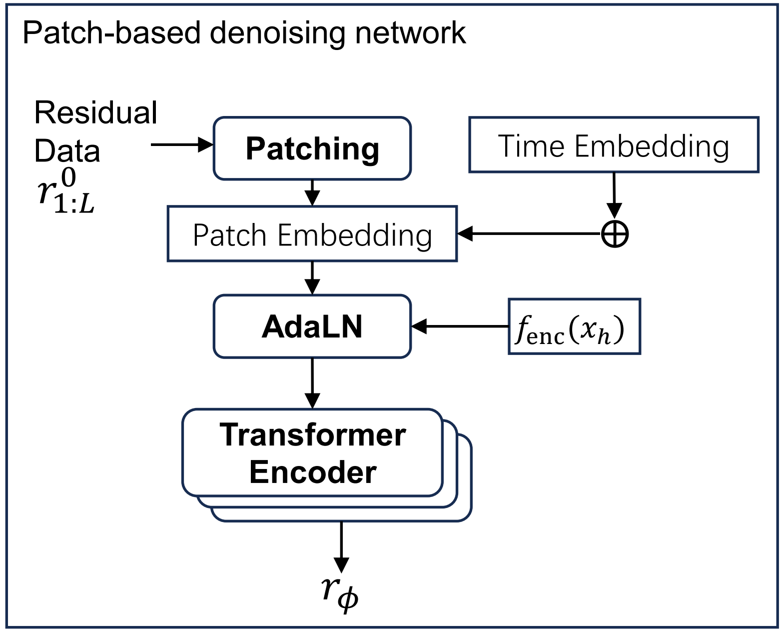
$$p_\phi(r^{0:K}_{1:L} | c) = p_\phi(r^K_{1:L}) \prod_{k=1}^{K} p_\phi(r^{k-1}_{1:L} | r^k_{1:L}, c)$$



Reverse process at step k:

$$p_\phi(r^{k-1}_{1:L} | r^k_{1:L}, c) = \mathcal{N}(r^{k-1}_{1:L}; \boxed{\mu_\phi(r^k_{1:L}, k | c)}, \sigma_k^2 \mathbf{I})$$

$$\mu_\phi(r^k_{1:L}, k|c) = \frac{\sqrt{\alpha_k}(1-\bar{\alpha}_{k-1})}{1-\bar{\alpha}_k} r^k_{1:L} + \frac{\sqrt{\bar{\alpha}_{k-1}}\beta_k}{1-\bar{\alpha}_k} \boxed{r_\phi(r^k_{1:L}, k|c)}$$



Main Results

Table 1: Performance comparison on six real-world datasets based on MSE and MAE. The **best**/<u>second</u> results are highlighted in **bold**/<u>underline</u>, respectively. Lower MSE and MAE values indicate better performance. *SparveVQ is used as the point forecasting model in the* $\mathrm{D^3U}$ *(ours).*

Model	Dataset	ETTm1		ETTm2		Weather		Solar-Energy		Electricity		Traffic	
	Method	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Point forecasting	NSformer(2022b)	0.440	0.430	0.277	0.343	0.226	0.270	0.266	<u>0.270</u>	0.191	0.295	0.653	0.360
	TimesNet(2023)	0.374	0.387	0.249	0.309	**0.219**	0.261	0.296	0.318	0.184	0.289	0.617	0.336
	DLinear(2023)	0.380	0.389	0.284	0.362	0.237	0.296	0.320	0.398	0.196	0.285	0.598	0.370
	PatchTST(2023)	<u>0.370</u>	0.390	0.251	0.312	0.223	<u>0.258</u>	0.259	0.321	0.205	0.307	<u>0.463</u>	0.311
	SparseVQ(2024)	**0.363**	**0.380**	<u>0.242</u>	**0.302**	0.225	<u>0.258</u>	0.256	0.286	0.182	<u>0.267</u>	0.480	0.300
	iTransformer(2024)	0.377	0.391	0.250	0.309	<u>0.221</u>	**0.254**	**0.233**	**0.261**	**0.164**	**0.255**	**0.418**	**0.284**
Probabilistic forecasting	TimeGrad(2021)	1.716	1.057	1.385	0.732	0.885	0.551	1.211	1.004	0.645	0.723	0.932	0.807
	CSDI(2021)	0.867	0.690	1.291	0.576	0.842	0.523	0.848	0.818	0.553	0.795	0.921	0.678
	TimeDiff(2023)	0.796	0.577	0.284	0.342	0.277	0.331	1.169	0.936	0.730	0.690	1.465	0.851
	TMDM(2024)	0.607	0.558	0.524	0.493	0.244	0.286	0.295	0.317	0.222	0.329	0.721	0.411
	ours	**0.363**	<u>0.386</u>	**0.241**	**0.302**	0.222	0.264	<u>0.237</u>	<u>0.270</u>	<u>0.179</u>	<u>0.267</u>	0.468	<u>0.299</u>

- With *SparseVQ* as the conditional network, $\mathrm{D^3U}$ retains the point forecasting performance of the conditional network employed, and even performs better on high-dimensional datasets (e.g. Solar-Energy)
- On point forecasting task, $\mathrm{D^3U}$ achieves a 28% improvement in MSE and a 21% improvement in MAE compared to the current state-of-the-art (SOTA) probabilistic MTS long-term forecasting baseline

Main Results

Table 2: Performance comparisons on six real-world datasets regarding CRPS and CRPS_{sum}. The **best**/second results are highlighted in **bold**/underline. Lower CRPS and CRPS_{sum} values indicate better performance. *SparveVQ is used as the point forecasting model in the* $\mathrm{D^3U}$ *(ours)*.

Model	Dataset	ETTm1		ETTm2		Weather		Solar-Energy		Electricity		Traffic	
	Method	CRPS	CRPS_{sum}	CRPS	CRPS_{sum}	CRPS	CRPS_{sum}	CRPS	CRPS_{sum}	CRPS	CRPS_{sum}	CRPS	CRPS_{sum}
Probabilistic Forecasting	**TimeGrad(2021)**	0.665	0.996	0.785	1.051	0.482	0.503	0.783	1.167	0.503	1.452	0.657	1.683
	CSDI(2021)	0.773	0.852	0.625	0.782	0.508	0.465	0.649	0.681	0.465	0.823	0.612	1.275
	TimeDiff(2023)	0.454	0.846	0.316	0.180	0.293	0.400	0.900	1.164	0.475	0.594	0.671	0.823
	TMDM(2024)	0.429	0.633	0.380	0.226	0.226	0.292	0.375	0.267	0.446	**0.137**	0.552	**0.179**
	ours	**0.285**	**0.574**	**0.243**	**0.141**	**0.207**	**0.283**	**0.186**	**0.266**	**0.202**	0.160	**0.232**	0.186

- $\mathrm{D^3U}$ presents a superior probabilistic forecasting performance, showing a 40% improvement in CRPS and a 5% improvement in $\mathrm{CRPS_{sum}}$ compared to the SOTA baseline.

Main Results

Table 3: Performance promotion by applying $\mathrm{D}^3\mathrm{U}$ to TMDM.

Mode	**TMDM**				**TMDM ($\mathrm{D}^3\mathrm{U}$)**			
Datasets	MSE	MAE	CRPS	$\mathrm{CRPS}_{\mathrm{sum}}$	MSE	MAE	CRPS	$\mathrm{CRPS}_{\mathrm{sum}}$
ETTm1	0.607	0.558	0.429	0.633	**0.441**	**0.432**	**0.324**	**0.616**
ETTm2	0.524	0.493	0.380	0.226	**0.317**	**0.399**	**0.302**	**0.147**
Weather	0.244	0.286	0.226	0.292	**0.215**	**0.267**	**0.196**	**0.273**
Solar-Energy	0.295	0.317	0.375	0.267	**0.269**	**0.299**	**0.328**	**0.260**
Electricity	0.222	0.329	0.446	**0.137**	**0.216**	**0.328**	**0.381**	0.157
Traffic	0.721	0.411	0.552	**0.179**	**0.678**	**0.402**	**0.472**	0.207

- To demonstrate the effectiveness of the $\mathrm{D}^3\mathrm{U}$ framework, a same setting as TMDM is applied to $\mathrm{D}^3\mathrm{U}$: using NSformer and MLP as point prediction models and denoising diffusion networks, respectively
- Compared to the original TMDM, the use of the $\mathrm{D}^3\mathrm{U}$ framework has resulted in significant improvements in both point and probabilistic MTS long-term forecasting tasks

Table 4: MSE, MAE and CRPS scores for different variants of the proposed method.

Ablation Study	**Mode** ($f_\theta + g_\phi$)	**ETTm1**			**Solar-Energy**			**Traffic**		
		MSE	**MAE**	**CRPS**	**MSE**	**MAE**	**CRPS**	**MSE**	**MAE**	**CRPS**
Denoise Network	SVQ + MLP[a]	0.372	0.396	0.294	0.330	0.313	0.242	0.525	0.331	0.297
	SVQ + UNet[b]	0.385	0.410	0.301	0.267	**0.266**	0.219	0.469	0.301	0.289
Structure Design	SVQ + PatchDN(CAttn)[c]	0.370	0.390	0.295	0.237	0.281	0.193	0.483	0.307	0.240
	SVQ + PatchDN(InC)[d]	0.366	**0.385**	0.290	0.238	0.271	0.187	0.486	0.315	0.244
Framework Design	SVQ + PatchDN(All)[e]	0.859	0.699	0.516	0.348	0.302	0.251	0.687	0.396	0.302
	SVQ + PatchDN($\hat{y}$)[f]	0.408	0.421	0.312	0.259	0.301	0.241	0.479	0.308	0.251
	Ours	**0.361**	**0.385**	**0.284**	**0.236**	0.270	**0.186**	**0.468**	**0.299**	**0.232**

1 SVQ is the abbreviation for SparseVQ.
2 a means the MLP serves as the denoising network in the TMDM model and consists of four linear layers; b means the UNet, used as the denoising network in the TimeDiff model, is built using a convolutional neural network-based UNet architecture.
3 c marks PatchDN based on the cross-attention method; d marks PatchDN based on the in-context method.
4 e marks the framework variant that models the entire data distribution; f marks the framework variant that employs $\hat{y}$ as the guidance.

- **Decoupling scheme**: Whether using the original input series or the output of the point forecasting model as the learning objective of the diffusion model, the performance has significantly decreased (Framework Design). Decoupling the deterministic and uncertain components is important.
- **Patch-based denoising design**: Patch-based denoising network (Denoise Network) employing AdaLN (Structure Design) exhibits a stronger capability in modeling the high-uncertainty component.

Main Results

Table 5: Performance promotion by applying the proposed framework to point forecasting models.

Dataset	ETTm1			Solar-Energy			Traffic		
Method	MSE	MAE	CRPS	MSE	MAE	CRPS	MSE	MAE	CRPS
NSformer	0.440	0.430	–	**0.266**	**0.270**	–	**0.653**	**0.360**	–
NSformer (D^3U)	**0.436**	**0.427**	**0.317**	0.268	0.272	**0.202**	0.657	0.367	**0.284**
PatchTST	**0.370**	**0.390**	–	0.259	0.321	–	0.463	0.311	–
PatchTST (D^3U)	0.387	0.405	**0.299**	**0.233**	**0.281**	**0.221**	**0.452**	**0.297**	**0.234**
SparseVQ	0.363	**0.380**	–	0.256	0.286	–	0.480	**0.300**	–
SparseVQ (D^3U)	**0.361**	0.385	**0.284**	**0.237**	**0.270**	**0.185**	**0.475**	0.309	**0.232**

[1] – means that point forecasting models do not have probabilistic forecasting abilities. The CRPS value degrades to the Normalized Mean Square Error (NMAE), which is omitted here.

- **Framework generality**: The D^3U framework can be used as a plug-and-play solution for point forecasting models and provide them with probabilistic forecasting capabilities while retaining their original point forecasting ability

Summary

1. Novel Complementary Modeling Paradigm

- We propose a novel complementary modeling approach that **combines point forecasting models and probabilistic forecasting models** from the perspective of decoupling the deterministic and uncertain components of time series data.

2. Conditional Diffusion Framework

- Within the D3U framework, we design a **patch-based denoising network, PatchDN**, to enhance the diffusion model's ability to represent the high-uncertainty components in time series data.

- **Fewer diffusion steps** required by focusing only on uncertain components

3. Plug-and-Play Architecture

- **Seamless integration** with existing point forecasting and diffusion models

- Simultaneously improves **both probabilistic and point forecasting performance**

Thank you

Paper ID 7497