

NeurIPS 2026

TimeES: Probabilistic and Deterministic Time Series Forecasting via Evolutionary Spectra



Weiwei Ye Renhe Jiang* Hangchen Liu Dongyuan Li Yoshihide Sekimoto

The University of Tokyo *Corresponding author

Paper: arxiv.org/abs/2609.32384

Code: github.com/www155/TimeES



Paper



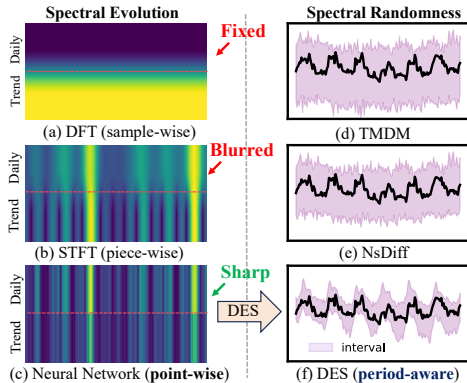
Code

Real-world time series are non-stationary

Trends, periodicity and uncertainty **evolve over time**, yet two common choices ignore this:

1. **Fourier features assume a fixed spectrum.** The FFT has no time resolution; the STFT blurs evolving components.
2. **Generative forecasters assume i.i.d. data.** Diffusion models from vision treat periodic structure as noise.

Goal: model how the spectrum evolves *and* its randomness, in one framework.



ETTm1 case study. (a–c) Spectra from the DFT, the STFT and TimeES. (d–f) Forecast intervals of TMDM, NsDiff and TimeES.

Evolutionary spectra: a time-varying spectrum for random processes

Stationary process (spectral representation)

$$X_t = \int_{-\infty}^{\infty} e^{i\omega t} dZ(\omega)$$

$dZ(\omega)$ is a random spectral measure; drawing it from a standard complex Gaussian gives white noise.

Non-stationary process (Priestley, 1965)

$$X_t = \int_{-\infty}^{\infty} A(t, \omega) e^{i\omega t} dZ(\omega)$$

The amplitude $A(t, \omega)$ evolves over time; $|A(t, \omega)|^2$ is the evolutionary power spectrum.

The gap. The theory is principled, but the evolutionary spectrum is not identifiable and has no reliable estimator, so it is rarely used in practice. **TimeES turns its estimation into representation learning.**

Discrete evolutionary spectra (DES)

Theorem. For samples $X_1, \dots, X_N$ and frequencies $\omega_k = 2\pi k/M, k = 0, \dots, M-1$,

$$X_n = \frac{1}{\sqrt{M}} \sum_{k=0}^{M-1} A(n, \omega_k) W_k e^{i2\pi kn/M}, \quad W_k \sim \mathcal{CN}(\mu_{W_k}, \sigma_{W_k}) \text{ independent.}$$

In matrix form, $\mathbf{x} = \frac{1}{\sqrt{M}} (\mathbf{A} \odot \mathbf{F}) \mathbf{w}$, which separates three roles:

Fourier basis $\mathbf{F}$

$\mathbf{F}_{n,k} = e^{i2\pi kn/M}$: fixed physical frequencies.

Amplitudes $\mathbf{A}$

$\mathbf{A}_{n,k} = A(n, \omega_k)$: evolving amplitudes, estimated by a network.

Realization $\mathbf{w}$

$\mathbf{w} = [W_0, \dots, W_{M-1}]^\top$: random excitation; one draw is one sample path.

One formula, two forecasting modes

Deterministic

$$W_k = 1$$

$$\mathbf{y} = \frac{1}{\sqrt{M}} (\hat{\mathbf{A}}_{\mathbf{y}} \odot \mathbf{F}) \mathbf{1}$$

Time-frequency decomposition.

Probabilistic

$$W_k \sim \mathcal{CN}(\mu_{W_k}, \sigma_{W_k})$$

$$\mathbf{y} = \frac{1}{\sqrt{M}} (\hat{\mathbf{A}}_{\mathbf{y}} \odot \mathbf{F}) \mathbf{w}$$

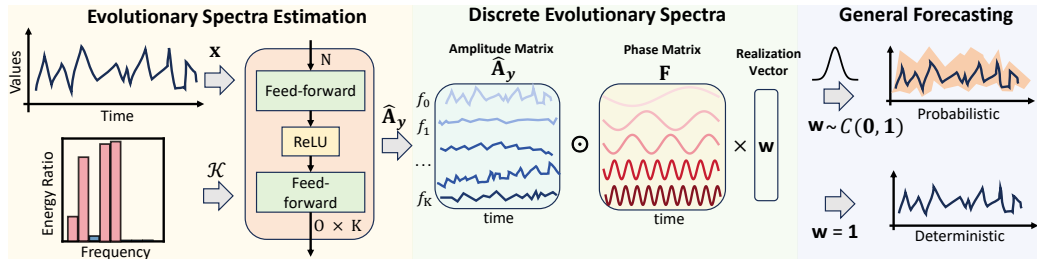
Generative model of the process.

μ_{W_k}, σ_{W_k} are learnable constants per frequency and channel, shared over the horizon; all time dependence comes from $\hat{\mathbf{A}}_{\mathbf{y}}$.

Classical transforms are special cases of the deterministic mode:

Method	Amplitude $A(n, \omega_k)$	W_k
Inverse DFT	$(\mathbf{F}\mathbf{x})_k$, constant in time	1
Inverse STFT	$(\mathbf{F}\mathbf{x}^{(m)})_k$ with $m = \lfloor n/L \rfloor$, piecewise constant	1
DES (ours)	$\text{MLP}_{n,k}(\mathbf{x})$, learned for every time step	$\mathcal{CN}(\mu_{W_k}, \sigma_{W_k})$

TimeES: estimate the spectrum, then synthesize



1 Estimate. An MLP maps the lookback x to the forecast amplitudes $\hat{\mathbf{A}}_y$ on the selected frequencies $\mathcal{K}$.

2 Synthesize. DES combines $\hat{\mathbf{A}}_y$ with the Fourier basis $\mathbf{F}$ and a realization $\mathbf{w}$.

3 Forecast. $\mathbf{w} = 1$ gives a point forecast; draws $\mathbf{w} \sim \mathcal{CN}$ give a distribution.

Making estimation cheap: from $\mathcal{O}(NM)$ to $\mathcal{O}(NK)$

Hermitian symmetry. A real series is fixed by its $\lfloor M/2 \rfloor + 1$ non-redundant frequencies: estimate half the spectrum.

Energy-based selection. Keep the smallest set $\mathcal{K}$ of frequencies whose average training periodogram satisfies

$$I(k) = \mathbb{E}_{\mathbf{x} \sim \mathcal{D}^{\text{train}}} [|\hat{X}_k|^2], \quad \frac{\sum_{k \in \mathcal{K}} I(k)}{\sum_{k=0}^{M-1} I(k)} \geq r.$$

A fixed ratio r adapts K to each dataset's signal-to-noise ratio.

A simple backbone. An MLP predicts only the selected columns:

$$[\hat{\mathbf{A}}_{\mathbf{y}}^r, \hat{\mathbf{A}}_{\mathbf{y}}^i] = \mathbf{\Theta}_2 \text{ReLU}(\mathbf{\Theta}_1 \mathbf{x} + \mathbf{b}_1) + \mathbf{b}_2 \\ \in \mathbb{R}^{N \times 2K}$$

The output has $2NK$ entries instead of $2NM$; frequencies outside $\mathcal{K}$ carry zero amplitude.

This simple backbone suffices for state-of-the-art accuracy: the gains come from the spectral model.

Probabilistic forecasting: best CRPS on 8 of 9 datasets

17.2%

lower CRPS than NsDiff, on average

18.7%

lower MSE than NsDiff, on average

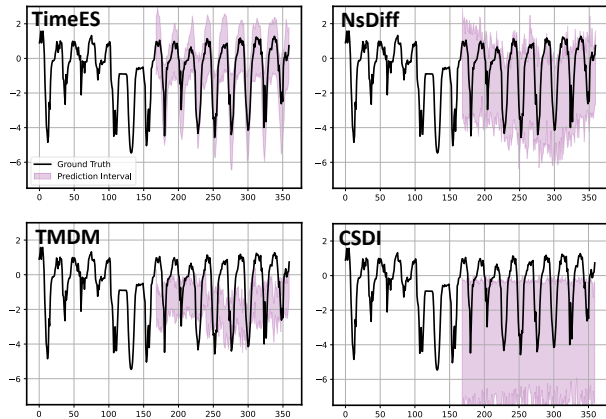
35%

lower CRPS than NsDiff on Traffic

CRPS ↓	ETTh1	ETTh2	ETTm1	ETTm2	ECL	EXG	ILI	Solar	Traffic
TimeGrad	0.606	1.212	0.647	0.775	0.397	0.826	1.140	<u>0.293</u>	0.407
CSDI	0.492	0.647	0.524	0.817	0.577	0.855	1.244	0.432	1.418
TimeDiff	0.465	0.471	0.464	0.316	0.750	0.433	1.153	0.700	0.771
DiffusionTS	0.603	1.168	0.574	1.035	0.633	1.251	1.612	0.470	0.668
TMDM	0.452	0.383	0.375	0.289	0.461	0.336	0.967	0.350	0.557
NsDiff	<u>0.392</u>	<u>0.358</u>	<u>0.346</u>	<u>0.256</u>	<u>0.290</u>	<u>0.324</u>	0.806	0.300	<u>0.378</u>
TimeES	0.349	0.320	0.310	0.230	0.205	0.270	<u>0.834</u>	0.193	0.245

Bold: best; underlined: second best. MSE, QICE and MAE are reported in the paper.

A closer look: uncertainty that follows the period



An ETTh1 test sample; shaded areas are 96% prediction intervals.

- **TimeES** recovers both the periodic structure and its evolving uncertainty.
- **NsDiff** follows the changing uncertainty but misses the spectral components.
- **TMDM** and **CSDI** miss the non-stationarity between input and output, which gives poorly calibrated variances.

Deterministic forecasting: accurate on its own and as a plug-in

Dataset	Univariate		Multivariate	
	Best baseline	TimeES	Best baseline	TimeES
ETTh1	0.755 FreTS	0.740	0.386 PatchTST	0.390
ETTh2	0.452 FreTS	0.434	0.281 PatchTST	0.278
ETTh1	0.882 FEDformer	0.878	0.440 FEDformer	0.440
ETTh2	0.596 FreTS	0.583	0.383 iTransformer	0.379
ECL	0.530 FreTS	0.516	0.178 iTransformer	0.197
EXG	0.517 FreTS	0.356	0.360 iTransformer	0.357
Traffic	0.500 iTransformer	0.431	0.428 iTransformer	0.478
Weather	0.523 FreTS	0.513	0.243 FreTS	0.239
Solar	0.191 PatchTST	0.205	0.233 iTransformer	0.243

Average MSE over horizons 96, 192, 336 and 720 (input 96). Bold: best.

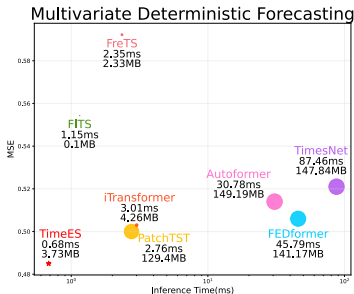
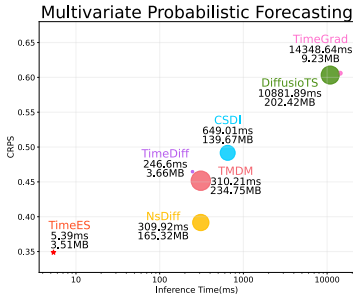
Setup. 7 baselines, from Autoformer to iTransformer; univariate and multivariate.

Results. Best on 8 of 9 datasets (univariate) and best or tied on 5 of 9 (multivariate).

As a plug-in ($w = 1$), ETTh1 MSE:

- PatchTST 0.469 $\rightarrow$ 0.438 (−6.6%)
- iTransformer 0.454 $\rightarrow$ 0.441 (−2.8%)

Efficiency: one forward pass instead of iterative denoising



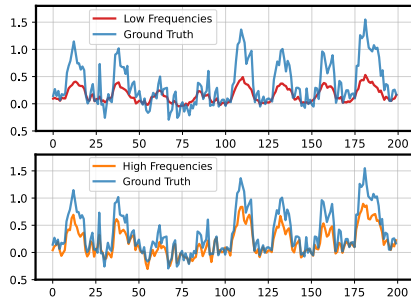
Probabilistic. 5.39 ms and 3.5 MB, against 309.9 ms and 165 MB for NsDiff and 14.3 s for TimeGrad.

Deterministic. 0.68 ms, the fastest of all compared models.

Inference time and memory on ETTh1; deterministic results use 720-step multivariate forecasting.

Interpretability: spectra you can read

Band-wise synthesis. Summing subsets of frequencies splits a forecast into components: the low band follows the trend, the high band abrupt changes.

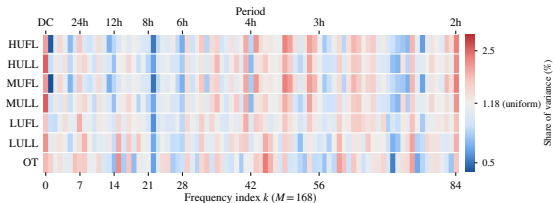


TimeES · NeurIPS 2026

Uncertainty attribution. With independent W_k , the predictive variance splits by frequency:

$$\text{Var}[X_n] = \sum_k dS(n, \omega_k),$$

$$dS(n, \omega_k) = \frac{1}{M} |A(n, \omega_k)|^2 \sigma_{W_k}^2.$$



ETTh1: load channels concentrate uncertainty at the trend; oil temperature (OT) carries more in the weekly band ($k = 1$).

Ablation and further analyses

Ablation on ETTh1

Variant	Univ. MSE	Multiv. MSE	CRPS
TimeES	0.929	0.485	0.349
w/o evolution	1.282	0.604	0.466
w/o DES	0.974	0.554	not generative

w/o evolution: the same spectrum at every forecast step. w/o DES: a linear layer outputs the forecast directly.

More in the paper

- **vs. KooNPro** (probabilistic Koopman): lower CRPS in every tested setting, e.g. 0.349 vs. 3.079 on ETTh1 at horizon 192.
- **Input length**: on ETTh1, a 24-step input stays within 2% of the default 168-step CRPS.
- **Learnable frequencies** stay on the DFT grid (at most 0.065 bins of drift) and do not improve accuracy.
- **Classification**: 62.96% average accuracy on 5 UEA tasks, the best of 8 models.

Takeaways

1. **Evolutionary spectra become learnable.** DES ties the time-varying spectrum to every observation, so a network can estimate it from data.
2. **One model, two modes.** The same amplitudes give point forecasts ($\mathbf{w} = 1$) and predictive distributions ($\mathbf{w} \sim \mathcal{CN}$).
3. **Efficient and interpretable.** $\mathcal{O}(NK)$ outputs, a single forward pass, and spectra that decompose both the forecast and its uncertainty.

Limitations. Learning evolutionary spectra needs enough data (e.g. the small ILI dataset), and energy-based selection may struggle under heavy noise.



Paper: arxiv.org/abs/2609.32384



Code: github.com/www155/TimeES