KOPA: A Kuramoto Oscillator Phase Adapter for Long-Term Forecasting of Periodic Multivariate Time Series

Nhan Thai[0009−0003−3249−7496] and Tran Thanh Phuong(✉)[0009−0004−6392−6895]

Faculty of Information Technology, Lac Hong University, Bien Hoa, Dong Nai, Vietnam
926000386@lachong.edu.vn, thanhphuong@lhu.edu.vn

Abstract. Many multivariate series in traffic, energy and weather monitoring are dominated by a known cycle whose phase differs across channels and drifts within a window. We propose KOPA, a Kuramoto Oscillator Phase Adapter that can be attached to any forecasting backbone. From the look-back window KOPA estimates, in closed form, the harmonic profile and the phase drift of every channel; it then propagates the phase deviations over the horizon with Kuramoto dynamics on a learned low-rank coupling graph, so that physically coupled channels pull each other back into alignment. The coupled harmonic forecast is added to the backbone, which only models the residual. A per-channel gate initialised at zero makes the combined model start exactly at the backbone, and the adapter adds only $N(r+1)+2$ parameters for N channels and rank r. We state six elementary properties of the adapter (exact reduction, exact harmonic recovery, approximate drift recovery, gradient-flow structure, bounded phase dispersion under coupling, and a Lipschitz forecast-error bound) and verify them numerically. On six benchmarks with a pronounced dominant cycle, three backbones (DLinear, PatchTST, iTransformer) and four horizons, KOPA reduces the test MSE in 67 of 72 settings, by 2.1% on average and up to 10%, and the resulting models are competitive with the periodic baseline CycleNet. An ablation shows that harmonic removal explains most of the gain, while the Kuramoto coupling adds a further 0.6–0.9% on the two road-traffic sensor networks whose channels are tightly phase-locked but drift at different rates.

Keywords: Time series forecasting · Kuramoto model · Phase synchronisation · Seasonality · Plug-in module.

1 Introduction

Long-term multivariate forecasting has recently been dominated by two design choices: light channel-independent models such as DLinear [29] and PatchTST [19], which share one mapping across channels, and inverted Transformers such as iTransformer [13], which attend across channels but treat the

whole look-back window as a token. Both families are trained end-to-end on raw windows, so they must re-learn, from data, something that is often known in advance: the series are built around a dominant cycle. Hourly electricity consumption and road traffic follow the day, five-minute traffic-flow sensors follow the 288-step day, and chaotic oscillator networks rotate around their attractor with a characteristic period, a regularity that specialised chaotic forecasters also exploit [18]. A simple, explicit treatment of this cycle can be surprisingly competitive; CycleNet [11], which subtracts a learned periodic pattern before a linear map, is a recent example.

Explicit periodic models, however, make an assumption that is rarely true in multivariate data: that every channel sits at the same phase of the cycle and stays there. Sensors on a road network peak at different times depending on their distance from the city centre, electricity customers have different daily profiles, and the "phase" of a channel measured on one window differs from the phase measured on the next. Within a look-back window the phase of each channel *drifts*, and if the drift is extrapolated linearly over a long horizon the harmonic forecast of neighbouring channels fans out, even though physically coupled channels (adjacent road sensors, oscillators sharing a mean field) should stay aligned.

This is exactly the situation described by the Kuramoto model of coupled oscillators [9,24,1,21]: oscillators with slightly different natural frequencies, coupled through the sine of their phase differences, synchronise when the coupling is strong enough relative to the frequency spread. We use this model as a forecasting device. The Kuramoto Oscillator Phase Adapter (KOPA) estimates, from the look-back window and in closed form, the harmonic coefficients and the phase drift of every channel. It then integrates the Kuramoto equation $\dot{\phi}_i = d_i + \sum_j K_{ij} \sin(\phi_j - \phi_i)$ over the horizon, with a learned low-rank coupling matrix K, and reconstructs a harmonic forecast with the resulting phases. The harmonic part is removed from the backbone's input and the coupled harmonic forecast is added to its output, with a per-channel gate that starts at zero. KOPA therefore (i) starts exactly at the backbone, (ii) adds very few parameters, and (iii) works with any backbone, channel-independent or not.

Our contributions are the following.

1. We introduce KOPA, a plug-in adapter that combines closed-form harmonic and phase-drift estimation with Kuramoto phase propagation on a learned low-rank coupling graph (Sect. 3).
2. We state six elementary properties of the adapter: exact reduction to the backbone at initialisation, exact recovery of harmonic coefficients on whole cycles, approximate recovery of the detuning, the gradient-flow structure of the phase dynamics, bounded phase dispersion under sufficient coupling from a phase-cohesive start, and a Lipschitz bound of the harmonic forecast error in terms of the phase error (Sect. 3.6). All six are verified numerically.
3. On six benchmarks with a pronounced dominant cycle, three backbones and four horizons, KOPA improves the test MSE in 67 of 72 settings and the resulting models are competitive with CycleNet (Sect. 4). An ablation sep-

arates the contributions of harmonic removal, phase drift and Kuramoto coupling, and two window-level diagnostics (phase-locking value and drift dispersion) indicate on which datasets the coupling pays off.

2 Related Work

Long-term forecasting backbones. Transformer forecasters [31,28,33,15,30] were challenged by the observation that a one-layer linear map on a decomposed window (DLinear) is a strong baseline [29]. Since then the field has converged on patching and channel independence (PatchTST [19]), inverted attention across channels (iTransformer [13]), and multi-scale MLP mixers [2,4,26]. KOPA restores one specific kind of cross-channel information, phase coupling, without changing the backbone; we use DLinear, PatchTST and iTransformer because they span the three families.

Periodicity and frequency. Fourier terms as regressors are standard in statistical forecasting [7,25]. In deep models, periodicity enters through frequency-domain attention [33,32], through folding the series along its dominant periods (TimesNet [27]), through basis expansions [20], through sparse period-aligned linear maps [12], or through an explicit learned cycle indexed by absolute time (CycleNet [11]). All of these treat the phase of the cycle as fixed per channel; none models how phases of different channels evolve jointly over the horizon. Koopman-based predictors [14,17] learn linear latent dynamics but not an oscillator structure.

Oscillators in neural networks. Coupled oscillators have been used as recurrent units with provably bounded gradients [23] and, very recently, as Kuramoto-type neurons whose synchronisation performs binding [16]. These works use oscillators as a computational substrate; KOPA uses the Kuramoto equation literally, as a model of how the measured phases of physically coupled channels behave, and keeps it outside the backbone as an adapter.

3 Method

3.1 Setting and notation

Let $\mathbf{x} \in \mathbb{R}^{L \times N}$ be a look-back window of N channels and $\mathbf{y} \in \mathbb{R}^{H \times N}$ the horizon to predict. A backbone f_θ maps $\mathbf{x}$ to $\hat{\mathbf{y}}$. We assume that a base period $P \leq L$ is known (24 for hourly data, 288 for five-minute data) or, for synthetic systems, estimated as the dominant period of the channel-averaged power spectrum of the training set. We write $\omega_1 = 2\pi/P$, index the look-back by $t = -(L-1), \ldots, 0$ so that $t = 0$ is the last observed point, and index the horizon by $\tau = 1, \ldots, H$.

As in RevIN [8] every window is instance-normalised, $\tilde{\mathbf{x}} = (\mathbf{x} - \mu)/\sigma$ with per-channel statistics of the window, and the prediction is de-normalised at the end. All quantities below are computed on $\tilde{\mathbf{x}}$.

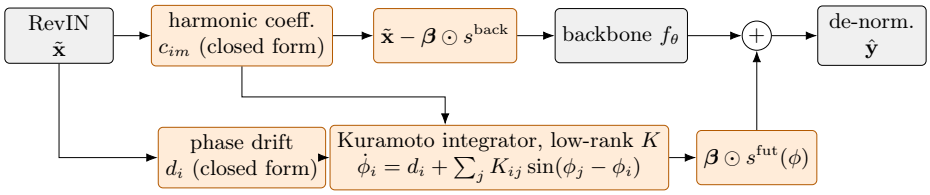


Fig. 1. KOPA wraps a backbone. Orange blocks are the adapter: closed-form harmonic and drift estimators, a Kuramoto integrator with a learned low-rank coupling K, and a per-channel gate $\boldsymbol{\beta}$ initialised at 0; grey blocks are unchanged. With $\boldsymbol{\beta} = 0$ the model is exactly the backbone.

3.2 Closed-form harmonic estimation

The last $W = \lfloor L/P \rfloor P$ points of the window cover an integer number of cycles. For every channel i and harmonic $m = 1, \ldots, M$ (we use $M = 3$) the complex coefficient

$$c_{im} = \frac{2}{W} \sum_{t=-(W-1)}^{0} \tilde{x}_{t,i}\, e^{-\mathrm{i}\, m\omega_1 t} \tag{1}$$

is a fixed linear functional of the window (two $W \times M$ matrices of cosines and sines, precomputed once). The harmonic profile of channel i over the look-back and over the horizon is

$$s^{\text{back}}_{t,i} = \sum_{m=1}^{M} \mathrm{Re}\big(c_{im} e^{\mathrm{i}m\omega_1 t}\big), \qquad s^{\text{fut}}_{\tau,i}(\phi) = \sum_{m=1}^{M} \mathrm{Re}\big(c_{im} e^{\mathrm{i}m(\omega_1 \tau + \phi_{\tau,i})}\big), \tag{2}$$

where $\phi_{\tau,i}$ is a *phase deviation* from the nominal phase $\omega_1 \tau$. With $\phi \equiv 0$, s^{fut} is the plain periodic extrapolation of the fitted harmonics; the whole point of KOPA is to predict ϕ.

3.3 Closed-form phase-drift estimation

A channel whose true frequency is $\omega_1 + \delta_i$ accumulates a phase deviation that grows linearly within the window. We estimate it with a windowed phase: for $n_o = \min(9, L - P + 1)$ offsets o_k spread uniformly over $[0, L - P]$, the complex projection $z_{ki} = \sum_{t \in \mathcal{W}_k} \tilde{x}_{t,i} e^{-\mathrm{i}\omega_1 t}$ of the P-point sub-window $\mathcal{W}_k$ gives a phase $\theta_{ki} = \arg z_{ki}$. The phases are unwrapped along k and a least-squares line is fitted against the sub-window centres; its slope d_i^{raw} (radians per step) is the drift. The raw slope is bounded and damped on channels with a weak fundamental,

$$d_i = \gamma\, d_{\max} \tanh\!\big(d_i^{\text{raw}} / d_{\max}\big)\, \frac{a_i^2}{a_i^2 + \epsilon}, \quad a_i^2 = |c_{i1}|^2,\ d_{\max} = \tfrac{\omega_1}{4},\ \epsilon = 0.05, \tag{3}$$

with a single learnable gain γ initialised at 1. The estimator is computed without gradient; only γ is trained. Since the harmonic fit (1) is centred on the fit window, the deviation at $t = 0$ is $\phi_{0,i} = d_i (W - 1)/2$.

3.4 Kuramoto propagation over the horizon

Without coupling, the deviation would be extrapolated linearly, $\phi_{\tau,i} = \phi_{0,i} + d_i\tau$, and channels would fan out at rates $d_i - d_j$. Instead, KOPA integrates the Kuramoto model

$$\dot{\phi}_i = d_i + \sum_{j=1}^{N} K_{ij} \sin(\phi_j - \phi_i), \qquad K = \frac{\kappa}{Nr}\, \tilde{U}\tilde{U}^\top, \quad \tilde{U} = \mathrm{softplus}(U)/\ln 2 \in \mathbb{R}_{>0}^{N\times r}, \tag{4}$$

from ϕ_0 over $\tau \in [0, H]$. The coupling matrix is symmetric, non-negative and of rank r ($r = 4$); with $U = 0$ it is the uniform all-to-all coupling κ/N, so training starts from the classical mean-field Kuramoto model and learns non-uniform coupling strengths from there. The global strength is $\kappa = \kappa_{\max}\,\mathrm{sigmoid}(\kappa_{\mathrm{raw}})$, initialised at $\min(1/P, \kappa_{\max}/2)$. Equation (4) is integrated with Heun's method on $n = \min(\max(8, H/4), 96)$ steps of size $\Delta = H/n$, with $\kappa_{\max} = \min(0.5, 1.5/\Delta)$ to keep the explicit integrator stable, and the trajectory is linearly interpolated to the integer horizon steps. The sum in (4) is never formed as an $N \times N$ matrix; with $C = \cos\phi$, $S = \sin\phi$ the coupling term equals $\kappa\,[C \odot ((S\tilde{U})\tilde{U}^\top) - S \odot ((C\tilde{U})\tilde{U}^\top)]/(Nr)$ and costs $O(Nr)$ per step.

3.5 Gated integration with the backbone

The backbone receives the window with the harmonic part removed and its output is corrected with the coupled harmonic forecast,

$$\hat{\tilde{\mathbf{y}}} = f_\theta\big(\tilde{\mathbf{x}} - \boldsymbol{\beta} \odot s^{\mathrm{back}}\big) + \boldsymbol{\beta} \odot s^{\mathrm{fut}}(\phi), \qquad \boldsymbol{\beta} \in \mathbb{R}^N,\ \boldsymbol{\beta}_{\mathrm{init}} = 0, \tag{5}$$

followed by de-normalisation (Fig. 1). The gate $\boldsymbol{\beta}$ is per channel, so a channel without a usable cycle can keep $\beta_i \approx 0$. The trainable parameters of KOPA are $\boldsymbol{\beta}$ (N), γ (1), κ_{raw} (1) and U (Nr); everything else is a fixed linear functional of the window. KOPA and the backbone are trained jointly with the MSE loss; because the adapter has so few parameters it uses a larger learning rate than the backbone.

Three variants are used in the experiments: +**Harm** ($\phi \equiv 0$: harmonic removal and periodic extrapolation only), +**Drift** ($\kappa = 0$: linear extrapolation of the estimated drift, no coupling) and +**KOPA** (drift and learned coupling).

3.6 Properties

The following statements are elementary but they say precisely what the adapter can and cannot do. Each was checked numerically (Table 1).

- **P0 Exact reduction.** With $\boldsymbol{\beta} = 0$, (5) equals $f_\theta(\tilde{\mathbf{x}})$ for any backbone, so KOPA is initialised at the backbone and cannot start worse than it.
- **P1 Exact harmonic recovery.** If $\tilde{x}_{t,i} = b + \sum_m A_m \cos(m\omega_1 t + \vartheta_m)$ on the fit window of W whole cycles, then $c_{im} = A_m e^{\mathrm{i}\vartheta_m}$ exactly; the constant b and all harmonics of other orders are annihilated by the orthogonality of the sampled exponentials.

Table 1. Numerical checks of P0–P5 (double precision; $L = 336$, $P = 24$, $M = 3$).

Property	Quantity checked	Value
P0	$\max \lvert f_\theta(\tilde{\mathbf{x}}) - \mathrm{KOPA}(\tilde{\mathbf{x}})\rvert$ at $\boldsymbol{\beta} = 0$, three backbones and CycleNet, random $\mathbf{x}$	0
P1	$\max_m \lvert c_m - A_m e^{\mathrm{i}\vartheta_m}\rvert$, random A_m, ϑ_m, offset $b = 0.7$	2.1×10^{-15}
P2	$\max \lvert d^{\mathrm{raw}} - \delta\rvert/\omega_1$ over 50 detunings $\lvert\delta\rvert \le 0.2\,\omega_1$	2.2×10^{-3}
P3	V non-increasing along 4000 Euler steps; $\max \lvert\nabla V + \dot{\phi}\rvert$ (finite diff.)	true; 4.0×10^{-10}
P4	fraction of 200 random systems ($5 \le N \le 60$, aligned start) with spread $< \pi/2$	1.00
	median uncoupled spread of the same systems at $\tau = 300$ (rad)	24.4
P5	violations of the bound in 2000 random $(c, \tau, \phi, \phi^\star)$	0

P2 Drift recovery. For a single detuned oscillation $\cos((\omega_1 + \delta)t + \vartheta)$ the unwrapped windowed phase is affine in the sub-window centre with slope δ up to the leakage of the P-point projection, so $d^{\mathrm{raw}} \approx \delta$.

P3 Gradient flow. For symmetric $K \ge 0$, the right-hand side of (4) is $-\nabla_\phi V$ with $V(\phi) = -\sum_i d_i\phi_i - \frac{1}{2}\sum_{ij} K_{ij}\cos(\phi_j - \phi_i)$, hence $\dot{V} = -\lVert\dot{\phi}\rVert^2 \le 0$: the continuous-time phase dynamics is a gradient flow and cannot blow up. (The discrete Heun scheme is kept stable separately, by the bound $\kappa \le \kappa_{\max}$ tied to the step size.)

P4 Bounded dispersion. Let $\Delta d = \max_i d_i - \min_i d_i$. With uniform coupling $K_{ij} = \kappa/N$, $\kappa \ge \Delta d$, and initial phases contained in an arc of length below $\pi/2$, the spread $\max_i \phi_i - \min_i \phi_i$ stays below $\pi/2$ for all τ: this is the phase-cohesiveness condition of the Kuramoto model [3,5], which is a statement about the synchronised region and not about arbitrary initial states. Without coupling the spread grows as $\Delta d\,\tau$.

P5 Forecast-error bound. For any two phase vectors, $\lvert s^{\mathrm{fut}}_{\tau,i}(\phi) - s^{\mathrm{fut}}_{\tau,i}(\phi^\star)\rvert \le \sum_m m\lvert c_{im}\rvert\,\lvert\phi_{\tau,i} - \phi^\star_{\tau,i}\rvert$, so the error of the harmonic forecast is controlled by the phase error, with a constant that grows with the harmonic order.

P4 and P5 together give the rationale of the adapter: for channels that start phase-cohesive, coupling keeps their phase error bounded over long horizons, and a bounded phase error gives a bounded harmonic-forecast error. They also say when KOPA should *not* help: when channels are not phase-locked, or when the fundamental is weak and (3) damps the drift.

4 Experiments

4.1 Setup

Datasets. We use five public benchmarks with a pronounced dominant cycle and one synthetic coupled-oscillator network (Table 2). *Traffic* (862 road-occupancy sensors, hourly) and *Electricity* (321 clients, hourly) follow a daily cycle, $P = 24$ [10]. *PEMS04* (307 sensors) and *PEMS08* (170 sensors) are five-minute traffic-flow series with $P = 288$ [6]. *Weather* (21 meteorological variables, 10-minute) has $P = 144$ [28]. *C-Rössler* is a network of $N = 16$ phase-coherent Rössler oscillators [22] with heterogeneous natural frequencies $\omega_i \sim \mathcal{U}(0.93, 1.07)$,

Table 2. Datasets, window-level phase diagnostics on the training set, learned adapter parameters, and average MSE reduction of +KOPA over the backbone (*Gain*), averaged over three backbones and four horizons. $\bar{a}$: mean fundamental-amplitude confidence $a^2/(a^2+\epsilon)$; PLV: inter-channel phase-locking value of the fundamental; σ_d/ω_1: cross-channel dispersion of the drift relative to the base frequency; $|\bar{\beta}|$ and κP: learned gate and coupling strength (per period) averaged over backbones and horizons.

Dataset	N	T	P	$\bar{a}$	PLV	σ_d/ω_1	$\lvert\bar{\beta}\rvert$	κP	Gain (%)
Traffic	862	17,544	24	0.95	0.98	0.005	0.52	1.59	+1.38
Electricity	321	26,304	24	0.93	0.94	0.013	0.41	1.35	+1.31
PEMS04	307	16,992	288	0.97	0.97	0.058	0.29	2.47	+1.72
PEMS08	170	17,856	288	0.97	0.97	0.058	0.25	2.93	+3.57
Weather	21	52,696	144	0.83	0.46	0.257	0.25	0.75	+1.34
C-Rössler	16	20,000	30	0.89	0.95	0.047	0.06	0.89	+3.48

parameters $(a, b, c) = (0.15, 0.2, 10)$, and mean-field diffusive coupling of strength 0.12 in the x variable; it is integrated with RK4 (step 0.05, every fourth sample kept, 20,000 points) and only x_i is observed. Its period, $P = 30$, is estimated from the training spectrum. Splits are 7:1:2 in time.

Protocol. The look-back is $L = 336$ and the horizons are $H \in \{96, 192, 336, 720\}$ for the public datasets and $\{24, 48, 96, 192\}$ for the faster-rotating Rössler network. Backbones are DLinear [29] (moving-average kernel 25), PatchTST [19] (patch 16, stride 8, 3 layers, $d = 128$) and iTransformer [13] (2 layers, $d = 128$). CycleNet/Linear [11] is run as a periodic reference with the cycle lengths recommended by its authors (168 for Traffic and Electricity, otherwise P). All models are trained with Adam on the MSE, at most 8 epochs of at most 150 batches, learning-rate decay 0.8 per epoch, early stopping on the validation MSE with patience 2, gradient clipping at 1, and the best validation checkpoint is tested. Batch sizes are set per backbone so that the number of series per batch is bounded. The backbone learning rate is 10^{-3} (DLinear, CycleNet) or 5×10^{-4} (Transformers); KOPA parameters use 5×10^{-3}. All runs use the same fixed seed and identical protocol for the backbone and its KOPA variants, so that each comparison differs only in the adapter. Metrics are MSE and MAE on the standardised test set. Code, configuration and the Rössler generator are available at https://www.thanhnhan.info/kopa.

4.2 Main results

Table 3 reports the full grid. KOPA reduces the test MSE of its backbone in 67 of the 72 (dataset, horizon, backbone) settings and the MAE in 67 of 72 as well; the average MSE reduction is 2.1% (DLinear 1.8%, PatchTST 2.9%, iTransformer 1.8%) and the largest single reductions are on PEMS08 (9% at $H = 96$ with DLinear), on the Rössler network (10% at $H = 192$ with PatchTST) and on PEMS04 (6.6% at $H = 720$ with PatchTST). A one-sided Wilcoxon signed-rank test over the 72 paired settings rejects equality of backbone and +KOPA

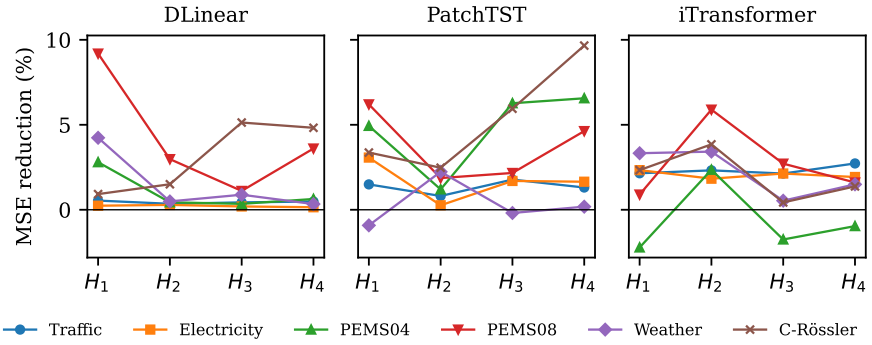


Fig. 2. MSE reduction of +KOPA over the backbone for each dataset and horizon ($H_1 < \cdots < H_4$ are 96/192/336/720, or 24/48/96/192 for C-Rössler).

at $p < 10^{-10}$. The five settings where KOPA is worse are three horizons of PEMS04 with iTransformer (at most 2.2% worse) and two horizons of Weather with PatchTST (at most 0.9%); in no case does the adapter degrade a backbone by more than that.

Gains are largest on the two PEMS datasets and on the Rössler network (3.5–3.6% on average, Table 2), where the fundamental is strongest ($\bar{a} \geq 0.89$), and around 1.3–1.4% on Traffic, Electricity and Weather. On the Rössler network the learned gate is small ($|\bar{\beta}| = 0.06$) but the base MSE is an order of magnitude below that of the other datasets, so a small correction is a visible relative gain. Fig. 2 shows the reduction per horizon: it is present at every horizon and does not vanish at $H = 720$.

CycleNet/Linear serves as a periodic reference. The best KOPA-equipped backbone has a lower MSE than CycleNet on every horizon of five datasets and on two of the four horizons of Electricity, but part of that margin belongs to the backbones themselves, since the best plain backbone already beats CycleNet on four of the six datasets. The comparison of comparable size is DLinear+KOPA against CycleNet, two linear models of which only CycleNet knows the absolute time index: DLinear+KOPA is ahead on the Rössler network (4 of 4 horizons), Weather (3 of 4) and PEMS08 (2 of 4), and behind on Traffic and Electricity, where CycleNet's weekly cycle of 168 carries information that a daily adapter does not have.

4.3 Ablation: harmonic removal, drift, coupling

Table 4 and Fig. 3 separate the three ingredients on the DLinear backbone, averaging the MSE over the four horizons. Harmonic removal alone (+Harm) already improves every dataset and, on Weather, Electricity and Traffic, it accounts for essentially the whole gain; on Weather it is even better than the full adapter (3.0% versus 1.2%), i.e. the drift estimated on the 21 heterogeneous

Table 3. Test MSE / MAE (lower is better) on six datasets, four horizons H and three backbones, with and without KOPA. Bold marks the +KOPA entry when it is at least as good as its backbone. CycleNet/Linear is a periodic reference. All runs share one fixed seed and one protocol.

Data	H	DLinear		+KOPA		PatchTST		+KOPA		iTransformer		+KOPA		CycleNet	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Traffic	96	0.422	0.300	**0.420**	**0.296**	0.415	0.297	**0.408**	**0.290**	0.424	0.315	**0.415**	**0.305**	0.414	0.296
	192	0.436	0.307	**0.435**	**0.305**	0.427	0.302	**0.424**	**0.295**	0.453	0.332	**0.443**	**0.323**	0.428	0.302
	336	0.448	0.311	**0.446**	**0.308**	0.442	0.309	**0.434**	**0.299**	0.470	0.341	**0.460**	**0.332**	0.440	0.307
	720	0.478	0.328	**0.476**	**0.325**	0.464	0.319	**0.458**	**0.311**	0.510	0.364	**0.496**	**0.353**	0.469	0.324
Electricity	96	0.145	0.243	**0.145**	**0.242**	0.143	0.240	**0.138**	**0.233**	0.147	0.248	**0.144**	**0.244**	0.140	0.239
	192	0.159	0.256	**0.159**	**0.255**	0.158	0.253	**0.157**	**0.252**	0.167	0.267	**0.164**	**0.264**	0.155	0.253
	336	0.176	0.271	**0.175**	**0.271**	0.176	0.272	**0.173**	**0.267**	0.186	0.285	**0.182**	**0.281**	0.172	0.268
	720	0.215	0.304	**0.215**	**0.303**	0.213	0.301	**0.209**	**0.296**	0.229	0.321	**0.225**	**0.317**	0.211	0.301
PEMS04	96	0.203	0.296	**0.197**	**0.293**	0.192	0.305	**0.182**	**0.288**	0.143	0.252	0.146	0.254	0.202	0.301
	192	0.224	0.306	**0.223**	**0.306**	0.213	0.308	**0.211**	**0.304**	0.175	0.283	**0.170**	**0.277**	0.221	0.313
	336	0.245	0.326	**0.244**	**0.325**	0.230	0.324	**0.215**	**0.309**	0.197	0.303	0.200	**0.303**	0.244	0.332
	720	0.350	0.398	**0.348**	**0.397**	0.267	0.353	**0.249**	**0.342**	0.255	0.354	0.257	**0.353**	0.315	0.384
PEMS08	96	0.259	0.319	**0.235**	**0.312**	0.193	0.286	**0.181**	**0.279**	0.153	0.254	**0.152**	**0.251**	0.257	0.319
	192	0.302	0.345	**0.293**	**0.343**	0.243	0.311	**0.239**	0.312	0.198	0.278	**0.186**	**0.267**	0.310	0.352
	336	0.326	0.360	**0.322**	**0.359**	0.262	0.320	**0.256**	**0.314**	0.226	0.300	**0.220**	**0.297**	0.322	0.362
	720	0.411	0.432	**0.397**	**0.424**	0.299	0.363	**0.286**	**0.354**	0.286	0.356	**0.282**	**0.356**	0.367	0.412
Weather	96	0.176	0.226	**0.169**	**0.221**	0.148	0.197	0.149	0.199	0.163	0.213	**0.158**	**0.207**	0.177	0.227
	192	0.217	0.260	**0.216**	**0.259**	0.194	0.240	**0.190**	**0.237**	0.207	0.253	**0.200**	**0.245**	0.218	0.262
	336	0.266	0.295	**0.263**	**0.293**	0.243	0.280	0.243	0.280	0.254	0.288	**0.252**	**0.284**	0.264	0.294
	720	0.334	0.343	**0.333**	**0.343**	0.320	0.333	**0.319**	0.333	0.329	0.341	**0.324**	**0.335**	0.333	0.341
C-Rössler	24	0.012	0.068	**0.012**	**0.068**	0.007	0.055	**0.007**	**0.054**	0.007	0.055	**0.006**	**0.054**	0.014	0.077
	48	0.020	0.093	**0.020**	**0.092**	0.009	0.064	**0.009**	**0.063**	0.009	0.065	**0.009**	**0.064**	0.022	0.099
	96	0.041	0.131	**0.039**	**0.127**	0.014	0.083	**0.014**	**0.080**	0.012	0.079	**0.012**	**0.079**	0.043	0.136
	192	0.085	0.184	**0.081**	**0.178**	0.030	0.116	**0.027**	**0.110**	0.023	0.106	**0.022**	**0.105**	0.088	0.188

weather channels is not worth propagating. Propagating the drift without coupling (+Drift) adds on the Rössler network and on the two PEMS datasets but costs a little on Weather and Electricity. The Kuramoto coupling (+KOPA) adds a further 0.6–0.9% on PEMS04 and PEMS08 and is neutral elsewhere, including the Rössler network (+0.1%); it has the best average rank (1.71) ahead of +Harm (2.04) and +Drift (2.25).

Over all 72 settings of the main grid, +KOPA is better than +Drift in 57% of the cases and the paired Wilcoxon test gives $p \approx 0.01$; the improvement is concentrated on the two PEMS datasets, where the coupling gain is +0.8% and +0.3% averaged over the three backbones. The division of labour is therefore clear: harmonic removal provides the bulk of the gain on every dataset, and the Kuramoto coupling is the component that pays off on sensor networks whose channels are physically coupled.

4.4 When does the coupling help? Two diagnostics

Equation (4) can only change a forecast if (i) channels drift at different rates, otherwise there is nothing to synchronise, and (ii) channels are phase-locked in the data, otherwise synchronising them is wrong. Table 2 reports two window-level statistics of the training set that measure exactly this: the inter-channel phase-locking value (PLV) of the fundamental, $|\mathbb{E}\, e^{\mathrm{i}(\vartheta_i - \vartheta_j)}|$ over random channel pairs,

Table 4. Ablation on the DLinear backbone: test MSE averaged over the four horizons, and relative gain over the backbone. +Harm: harmonic removal with periodic extrapolation ($\phi \equiv 0$); +Drift: linear extrapolation of the estimated drift ($\kappa = 0$); +KOPA: drift and learned coupling. Average rank over the 24 (dataset, horizon) pairs.

Dataset	Average test MSE				Gain over backbone (%)		
	Backbone	+Harm	+Drift	+KOPA	+Harm	+Drift	+KOPA
Traffic	0.4459	**0.4438**	0.4442	0.4439	+0.48	+0.40	+0.45
Electricity	0.1739	**0.1734**	0.1734	0.1735	+0.27	+0.23	+0.21
PEMS04	0.2556	0.2550	0.2547	**0.2532**	+0.23	+0.34	+0.95
PEMS08	0.3247	0.3176	0.3147	**0.3118**	+2.18	+3.06	+3.95
Weather	0.2483	**0.2408**	0.2452	0.2453	+3.02	+1.22	+1.21
C-Rössler	0.0395	0.0384	0.0379	**0.0379**	+2.74	+4.11	+4.19
Average rank	4.00	2.04	2.25	1.71			

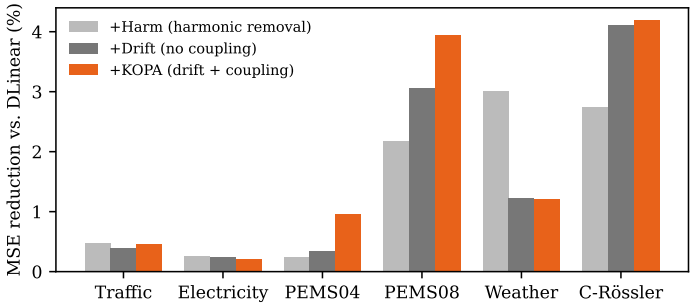


Fig. 3. Ablation on DLinear: MSE reduction of the three variants, averaged over horizons.

and the cross-channel dispersion σ_d of the estimated drift relative to ω_1. The two datasets with the largest coupling gain in Table 4, PEMS08 and PEMS04, combine a high PLV (0.97) with a sizeable drift dispersion (0.058): sensors on the same freeway are locked to the same day but their peaks move at different rates within a window. Traffic and Electricity are locked (PLV 0.98, 0.94) but barely drift (0.005, 0.013), so +Drift and +KOPA coincide with +Harm. Weather drifts strongly (0.26) but is not locked (PLV 0.46), so coupling is not justified and indeed does not help. The Rössler network is the exception that bounds the rule: it is locked (PLV 0.95) and drifts (0.047), yet the coupling adds only 0.1% beyond +Drift. Its 16 oscillators are generated with a global mean-field coupling, so their phases are already cohesive and the drift-only forecast captures most of their aligned motion, leaving little for the Kuramoto term to correct; the learned κP (0.9) is accordingly small. The learned coupling strength follows the same pattern on the other datasets: κP is 2.9 and 2.5 on PEMS08 and PEMS04, 1.3–1.6 on Traffic and Electricity, and smallest (0.75) on Weather.

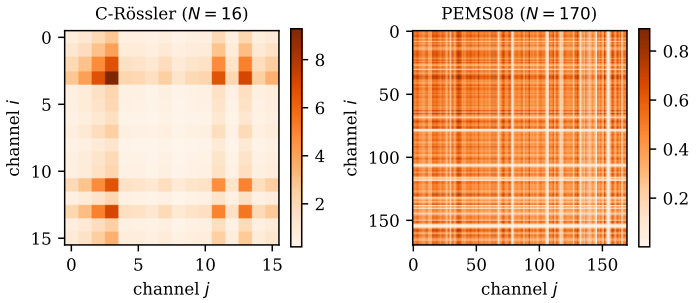


Fig. 4. Learned coupling matrices K ($\times 10^3$) with the DLinear backbone at $H = 96$. Left: the 16-oscillator Rössler network. Right: the 170 PEMS08 sensors.

The two diagnostics therefore give a necessary, not a sufficient, condition: coupling can only help when PLV is high and the drift dispersion is not negligible; when either fails, the cheaper +Harm variant is the better choice.

4.5 Learned coupling graphs

Fig. 4 shows the learned $K = \kappa \tilde{U}\tilde{U}^\top/(Nr)$ for the Rössler network and for PEMS08 (DLinear, $H = 96$). Both start from the uniform matrix and remain dense, as expected from a rank-4 non-negative factorisation, but they are no longer uniform: on the Rössler network the row sums range over a factor of seven between the least and the most coupled oscillator, and on PEMS08 over a factor of 25, i.e. the adapter learns which sensors act as hubs of the phase dynamics. We do not claim that K recovers the road graph; the rank-4 constraint is a capacity control, not a structural prior.

4.6 Cost

KOPA adds $N(r + 1) + 2$ parameters: 82 for the Rössler network, 1,607 for Electricity, 4,312 for Traffic, against 2.2×10^5 (DLinear), 3.5×10^5 (iTransformer) and 2.2×10^6 (PatchTST) for the backbones. The harmonic and drift estimators are fixed linear maps of the window. The dominant cost is the Heun integration of (4), $2n$ evaluations of an $O(Nr)$ coupling term per window with $n \leq 96$, i.e. $O(nNr)$ per window independently of the backbone size and linear in the number of channels. The +Harm variant has no integrator and costs the same as the backbone.

5 Limitations

The study uses a single training seed per setting and an identical protocol for every model; the 2% average effect is consistent across 72 settings and three

backbones, but individual entries of Table 3 carry the usual run-to-run variability. KOPA assumes one known base period and is designed for series with a pronounced dominant cycle; on series without one the learned gate is expected to stay near zero (P0 guarantees only the starting point), and we have not pursued multi-period extensions. The training budget is deliberately reduced (at most 8 epochs of 150 batches for every model), so absolute errors are above the best published numbers for these backbones and all comparisons are within-protocol. Finally, the coupling helps only when channels are phase-locked and drift; the diagnostics of Table 2 are our proposal for deciding this in advance, validated on six datasets only.

6 Conclusion

KOPA treats the dominant cycle of a multivariate series as what it physically is: a population of oscillators with slightly different frequencies that are coupled to each other. Closed-form estimation of harmonics and drift, Kuramoto propagation on a learned low-rank coupling graph, and a zero-initialised gate make the adapter exact at initialisation, nearly parameter-free, and applicable to any backbone. On six periodic benchmarks it improves DLinear, PatchTST and iTransformer in 67 of 72 settings and makes a linear backbone competitive with CycleNet. The ablation gives a clear division of labour, harmonic removal for the bulk of the gain and Kuramoto coupling for sensor networks whose channels are physically locked, and two cheap diagnostics tell the two situations apart. Multi-period adapters, coupling matrices informed by a known sensor graph, and a faster integrator are natural next steps.

References

1. Acebrón, J.A., Bonilla, L.L., Pérez Vicente, C.J., Ritort, F., Spigler, R.: The Kuramoto model: A simple paradigm for synchronization phenomena. Reviews of Modern Physics **77**(1), 137–185 (2005). https://doi.org/10.1103/RevModPhys.77.137
2. Chen, S.A., Li, C.L., Yoder, N., Arik, S.O., Pfister, T.: TSMixer: An all-mlp architecture for time series forecasting. Transactions on Machine Learning Research (2023). https://doi.org/10.48550/arXiv.2303.06053
3. Chopra, N., Spong, M.W.: On exponential synchronization of Kuramoto oscillators. IEEE Transactions on Automatic Control **54**(2), 353–357 (2009). https://doi.org/10.1109/TAC.2008.2007884
4. Das, A., Kong, W., Leach, A., Mathur, S., Sen, R., Yu, R.: Long-term forecasting with TiDE: Time-series dense encoder. Transactions on Machine Learning Research (2023). https://doi.org/10.48550/arXiv.2304.08424
5. Dörfler, F., Bullo, F.: Synchronization in complex networks of phase oscillators: A survey. Automatica **50**(6), 1539–1564 (2014). https://doi.org/10.1016/j.automatica.2014.04.012
6. Guo, S., Lin, Y., Feng, N., Song, C., Wan, H.: Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In: Proceedings of the

AAAI Conference on Artificial Intelligence. vol. 33, pp. 922–929 (2019). https://doi.org/10.1609/aaai.v33i01.3301922
7. Hyndman, R.J., Athanasopoulos, G.: Forecasting: Principles and Practice. OTexts, Melbourne, Australia, 3rd edn. (2021)
8. Kim, T., Kim, J., Tae, Y., Park, C., Choi, J.H., Choo, J.: Reversible instance normalization for accurate time-series forecasting against distribution shift. In: International Conference on Learning Representations (ICLR) (2022), https://openreview.net/forum?id=cGDAkQo1C0p
9. Kuramoto, Y.: Self-entrainment of a population of coupled non-linear oscillators. In: International Symposium on Mathematical Problems in Theoretical Physics, Lecture Notes in Physics, vol. 39, pp. 420–422. Springer (1975). https://doi.org/10.1007/BFb0013365
10. Lai, G., Chang, W.C., Yang, Y., Liu, H.: Modeling long- and short-term temporal patterns with deep neural networks. In: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. pp. 95–104 (2018). https://doi.org/10.1145/3209978.3210006
11. Lin, S., Lin, W., Hu, X., Wu, W., Mo, R., Zhong, H.: CycleNet: Enhancing time series forecasting through modeling periodic patterns. In: Advances in Neural Information Processing Systems (NeurIPS). vol. 37 (2024). https://doi.org/10.48550/arXiv.2409.18479
12. Lin, S., Lin, W., Wu, W., Chen, H., Yang, J.: SparseTSF: Modeling long-term time series forecasting with 1k parameters. In: International Conference on Machine Learning (ICML) (2024). https://doi.org/10.48550/arXiv.2405.00946
13. Liu, Y., Hu, T., Zhang, H., Wu, H., Wang, S., Ma, L., Long, M.: iTransformer: Inverted transformers are effective for time series forecasting. In: International Conference on Learning Representations (ICLR) (2024). https://doi.org/10.48550/arXiv.2310.06625
14. Liu, Y., Li, C., Wang, J., Long, M.: Koopa: Learning non-stationary time series dynamics with Koopman predictors. In: Advances in Neural Information Processing Systems (NeurIPS). vol. 36 (2023). https://doi.org/10.48550/arXiv.2305.18803
15. Liu, Y., Wu, H., Wang, J., Long, M.: Non-stationary transformers: Exploring the stationarity in time series forecasting. In: Advances in Neural Information Processing Systems (NeurIPS). vol. 35, pp. 9881–9893 (2022). https://doi.org/10.48550/arXiv.2205.14415
16. Miyato, T., Löwe, S., Geiger, A., Welling, M.: Artificial Kuramoto oscillatory neurons. In: International Conference on Learning Representations (ICLR) (2025). https://doi.org/10.48550/arXiv.2410.13821
17. Nguyen, N.P.: WaveKoop-KAN: A wavelet-decomposed Koopman-gated temporal convolution hybrid with Kolmogorov–Arnold fusion for univariate time series forecasting. IEEE Access **14**, 96322–96342 (2026). https://doi.org/10.1109/ACCESS.2026.3706582
18. Nguyen, N.P., Duong, T.A., Platos, J.: T2S-EDNN: A segmentation-driven specialized ensemble framework with cluster-adaptive routing for robust chaotic time series forecasting. IEEE Access **14**, 119950–119971 (2026). https://doi.org/10.1109/ACCESS.2026.3719509
19. Nie, Y., Nguyen, N.H., Sinthong, P., Kalagnanam, J.: A time series is worth 64 words: Long-term forecasting with transformers. In: International Conference on Learning Representations (ICLR) (2023). https://doi.org/10.48550/arXiv.2211.14730

20. Oreshkin, B.N., Carpov, D., Chapados, N., Bengio, Y.: N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. In: International Conference on Learning Representations (ICLR) (2020). https://doi.org/10.48550/arXiv.1905.10437
21. Pikovsky, A., Rosenblum, M., Kurths, J.: Synchronization: A Universal Concept in Nonlinear Sciences. Cambridge University Press (2001). https://doi.org/10.1017/CBO9780511755743
22. Rössler, O.E.: An equation for continuous chaos. Physics Letters A **57**(5), 397–398 (1976). https://doi.org/10.1016/0375-9601(76)90101-8
23. Rusch, T.K., Mishra, S.: Coupled oscillatory recurrent neural network (coRNN): An accurate and (gradient) stable architecture for learning long time dependencies. In: International Conference on Learning Representations (ICLR) (2021). https://doi.org/10.48550/arXiv.2010.00951
24. Strogatz, S.H.: From Kuramoto to Crawford: Exploring the onset of synchronization in populations of coupled oscillators. Physica D: Nonlinear Phenomena **143**(1–4), 1–20 (2000). https://doi.org/10.1016/S0167-2789(00)00094-4
25. Taylor, S.J., Letham, B.: Forecasting at scale. The American Statistician **72**(1), 37–45 (2018). https://doi.org/10.1080/00031305.2017.1380080
26. Wang, S., Wu, H., Shi, X., Hu, T., Luo, H., Ma, L., Zhang, J.Y., Zhou, J.: TimeMixer: Decomposable multiscale mixing for time series forecasting. In: International Conference on Learning Representations (ICLR) (2024). https://doi.org/10.48550/arXiv.2405.14616
27. Wu, H., Hu, T., Liu, Y., Zhou, H., Wang, J., Long, M.: TimesNet: Temporal 2d-variation modeling for general time series analysis. In: International Conference on Learning Representations (ICLR) (2023). https://doi.org/10.48550/arXiv.2210.02186
28. Wu, H., Xu, J., Wang, J., Long, M.: Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In: Advances in Neural Information Processing Systems (NeurIPS). vol. 34, pp. 22419–22430 (2021). https://doi.org/10.48550/arXiv.2106.13008
29. Zeng, A., Chen, M., Zhang, L., Xu, Q.: Are transformers effective for time series forecasting? In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 11121–11128 (2023). https://doi.org/10.1609/aaai.v37i9.26317
30. Zhang, Y., Yan, J.: Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting. In: International Conference on Learning Representations (ICLR) (2023), https://openreview.net/forum?id=vSVLM2j9eie
31. Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., Zhang, W.: Informer: Beyond efficient transformer for long sequence time-series forecasting. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 35, pp. 11106–11115 (2021). https://doi.org/10.1609/aaai.v35i12.17325
32. Zhou, T., Ma, Z., Wen, Q., Sun, L., Yao, T., Yin, W., Jin, R.: FiLM: Frequency improved Legendre memory model for long-term time series forecasting. In: Advances in Neural Information Processing Systems (NeurIPS). vol. 35, pp. 12677–12690 (2022). https://doi.org/10.48550/arXiv.2205.08897
33. Zhou, T., Ma, Z., Wen, Q., Wang, X., Sun, L., Jin, R.: FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting. In: International Conference on Machine Learning (ICML). PMLR, vol. 162, pp. 27268–27286 (2022). https://doi.org/10.48550/arXiv.2201.12740