GENCOUP: CROSS-CHANNEL COUPLING FOR PROBABILISTIC FORECAST TRAJECTORIES

Anonymous authors
Paper under double-blind review

ABSTRACT

Probabilistic time-series forecasters generate ensembles of future trajectories to represent predictive uncertainty. For these ensembles to form coherent multivariate scenarios, they require accurate marginals and realistic dependence across channels over time. Channel-independent generative forecasters already capture within-channel temporal dynamics, but leave cross-channel coupling unspecified. Existing plug-and-play methods recover this coupling post hoc, often at the expense of the temporal dynamics already encoded in the trajectories. We introduce **Ge**nerative **N**oise **C**oupling (**GeNCoup**), which injects cross-channel dependence directly into the sampling process of a frozen generative forecaster. Instead of drawing channel noise independently, **GeNCoup** jointly samples the noise across channels at each noise feature using a cross-channel covariance (R), while preserving each channel's original noise distribution. This introduces cross-channel dependence without changing the forecaster's marginal predictions or within-channel temporal dynamics. Building on this framework, we introduce **GeNCoup-Corr**, which derives the coupling from empirical cross-channel correlations, and **GeNCoup-Dyn**, which learns it by optimizing a joint forecast score. Across six forecasters from four generative families, eight long-horizon datasets, and four prediction horizons, **GeNCoup** outperforms baselines in joint forecast quality while preserving both marginal accuracy and temporal dynamics.

1 INTRODUCTION

Probabilistic forecasting represents uncertainty over future outcomes through a predictive distribution rather than a single deterministic trajectory (Salinas et al., 2020; Rasul et al., 2021; Kollovieh et al., 2023). This distribution may be represented explicitly through quantiles or implicitly through an ensemble of sampled future trajectories. Sampled trajectories are particularly important for downstream systems because they represent multiple plausible future scenarios while capturing the joint dependence across channels through time. This forms a coherent scenario for simulation, risk evaluation, and decision making in applications such as energy, weather, and finance (Ruszczyński & Shapiro, 2003; Löhndorf, 2016; Madeka et al., 2018; Price et al., 2025; He & Zhang, 2024)

An ensemble of joint predictive trajectories is determined by two components: the *marginal predictive distributions*, which describe the possible values of each variable separately, and their *coupling*, which determines how samples from these marginals occur together (Wen & Torkkola, 2019; Ashok et al., 2024). The coupling, therefore, represents an additional degree of freedom beyond the marginals: specifying the marginals does not determine the coupling, so the same marginal predictive distributions can induce a substantially different ensemble of projections (Drouin et al., 2022). Figure 1 illustrates this distinction: coupled and uncoupled samples share the same marginal predictive distributions, yet exhibit substantially different dependence between two strongly related channels.

Accurate marginals are therefore insufficient when the sampled values are coupled incorrectly. For multivariate forecasting, cross-channel dependence is particularly important: even when each channel already has a coherent temporal trajectory, the channels must also co-occur realistically. In an energy system, for example, modeling uncertainty in demand and renewable generation separately can overlook critical scenarios, such as high demand coinciding with low generation, which can

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.

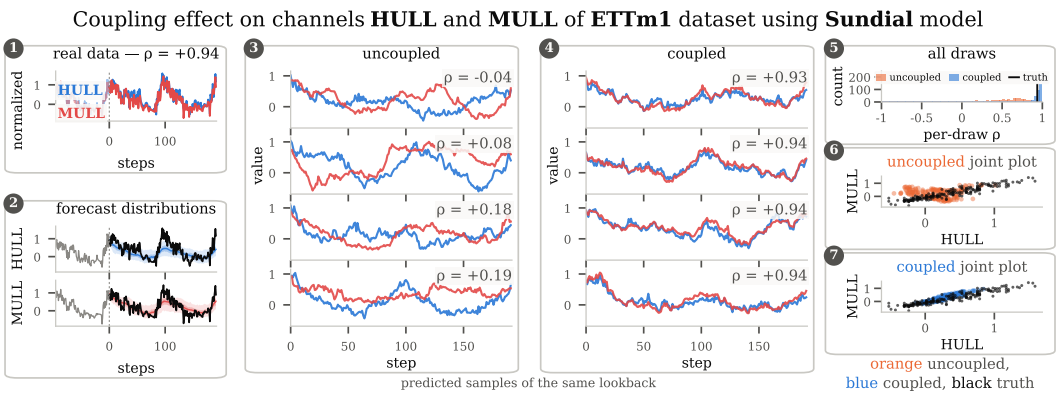


Figure 1: Sundial projections for the HULL and MULL channels of ETTm1. 1 The observed future window is strongly correlated (Pearson $\rho = 0.94$). 2 A probabilistic forecast represented by multiple projected trajectories. 3 and 4 Uncoupled and coupled trajectories exhibit different cross-channel dependence while being drawn from the same marginal predictive distributions. 5 The Pearson correlation of the coupled samples concentrates around the observed value. 6 and 7 The channel relationship at individual forecast steps, with the coupled trajectories following the observed dependence. Additional examples are provided in Appendix B.5.

lead to shortages and operational risk. (Stover et al., 2024). We therefore target cross-channel dependence while preserving the forecaster's existing within-channel temporal structure.

Unlike quantile models, which specify the marginal distributions without determining how values at different forecast times co-occur, sample-based generative forecasters directly produce complete future trajectories. These trajectories can already exhibit strong temporal structure, as supported by our analysis in Appendix B.1. In contrast, cross-channel dependence is often less explicitly structured. This is especially true for Channel-Independent (CI) forecasters, which model temporal dynamics within each channel while generating channels independently and omitting cross-channel dependence by design. This applies to both in-domain models (Kollovieh et al., 2025; Li et al., 2026) and pretrained foundation models (Liu et al., 2025; Wu et al., 2026). They can therefore provide strong marginals and coherent temporal structure within each channel while leaving cross-channel coupling unspecified. This motivates a research direction that leverages the forecaster's existing marginal and temporal structure while improving cross-channel dependence.

Existing plug-and-play approaches operate post hoc on an already generated forecast ensemble. Sample-reordering methods rearrange values across ensemble members independently at each forecast time to recover cross-channel dependence (Iman & Conover, 1982; Clark et al., 2004; Schefzik, 2016; Flos et al., 2026). This reordering often modifies the forecaster's original within-channel temporal dynamics. Copula-based methods instead reduce the generated forecasts to marginal distributions, for example, through quantiles, and resample them under a cross-channel copula (Möller et al., 2013). In doing so, they may discard the temporal dynamics encoded in the original trajectories. Thus, both approaches aim to recover cross-channel dependence at the expense of the temporal dynamics already generated by the forecaster.

We therefore introduce **Ge**nerative **N**oise **Coup**ling (GENCOUP), which integrates dependence modeling into the forecaster's sampling process rather than applying it post hoc to generated trajectories. GENCOUP applies to generative forecasters with independently drawn source noise across channels and a deterministic sampling procedure. It couples the source noise across channels while preserving each channel's marginal noise distribution. The frozen forecaster therefore receives the same marginal source distribution it was trained on, and its sampling procedure remains unchanged. This preserves the original marginal predictive distributions and the sampler's temporal dynamics, while allowing cross-channel dependence to be incorporated into the joint distribution.

Within this framework, we propose two variants that differ in how the source-noise dependence is determined: GENCOUP-CORR sets it directly from empirical cross-channel correlations, while GENCOUP-DYN starts from independent noise with $A = I$ and optimizes the dependence structure

Table 1: Comparison of dependence-reconstruction methods. ✓ yes, × no, ◐ partial: GCA preserves the marginals in distribution, while Schaake partially preserves temporal dependence through phase-aligned contiguous rank templates.

Method	Recovers joint information?	Preserves marginals?	Preserves temporal dependence?	Integrated into the forecaster?	Minimizable End-To-End?
Gaussian Copula (GCA) (Möller et al., 2013)	✓	◐	×	×	×
Iman–Conover (Iman & Conover, 1982)	✓	✓	×	×	×
SimSchaake (Schefzik, 2016)	✓	✓	×	×	×
COBASE (Flos et al., 2026)	✓	✓	×	×	×
Schaake Shuffle (Clark et al., 2004)	✓	✓	◐	×	×
GENCOUP-CORR (Ours)	✓	✓	✓	✓	×
GENCOUP-DYN (Ours)	✓	✓	✓	✓	✓

by minimizing a joint score over A, while keeping the forecaster frozen. Table 1 compares these variants with post-hoc dependence-reconstruction baselines.

Our contributions are summarized as follows:

1. We introduce GENCOUP, a general framework for injecting cross-channel dependence into originally independent source noise while preserving each channel's marginal predictive and within-channel temporal distributions.
2. We propose two variants designed to recover cross-channel joint dependence: GENCOUP-CORR, a simple empirical construction based on observed cross-channel correlations, and GENCOUP-DYN, which learns the source-noise dependence by optimizing a joint score while keeping the forecaster frozen.
3. We empirically show that GENCOUP improves joint forecast quality without sacrificing marginal quality or within-channel temporal structure. We evaluate it on six frozen forecasters from four generative families, including two pretrained foundation models, across eight long-horizon forecasting datasets and four prediction horizons. GENCOUP achieves the best variogram score on seven of eight datasets and the best cross-channel energy score on six of eight datasets.

2 RELATED WORK

Coupling marginal distributions. Quantile models predict a distribution for each time step (Wen & Torkkola, 2019; Park et al., 2022), which summarizes the marginals but not how values co-occur across time and channels, so sampling trajectories from such a forecast requires a separate additional coupling step (Wen & Torkkola, 2019; Baron et al., 2025). Other models integrate the marginal model with the copula and train them together. GP-Copula (Salinas et al., 2019) parametrizes a low-rank Gaussian copula from the recurrent state it was trained with, and TACTiS (Drouin et al., 2022) and TACTiS-2 (Ashok et al., 2024) learn marginals and an attentional copula jointly, with TACTiS-2 doing so in two stages, training the marginals first and then the copula with the marginals frozen. In all of them, the aim is to produce realistic samples. We instead aim to improve the cross-channel dependence of forecasters that already sample complete trajectories (Salinas et al., 2020; Rasul et al., 2021; Kollovieh et al., 2025).

Cross-channel dependence coupling. Several methods have been proposed to recover the missing dependence between channel-wise forecasts, most of them from ensemble weather and hydrological postprocessing. One family uses sample reordering, preserving the marginal distributions while changing which sampled values are paired across channels. The Iman–Conover correlation-induction method (Iman & Conover, 1982) rearranges the marginal samples to match a correlation computed over the data, while rank-based methods such as the Schaake shuffle and its variants (Clark et al., 2004; Schefzik, 2016) instead use the rank structure of historical multivariate trajectories. Another approach uses a parametric Gaussian copula (Möller et al., 2013), which explicitly estimates a joint dependence model separately from the marginals and samples from the resulting multivariate distribution, at the cost of marginal calibration. COBASE (Flos et al., 2026) combines the two, reordering to the ranks of a draw from a fitted copula. Most post-hoc couplers recover cross-channel dependence at the expense of the forecaster's original temporal structure: reordering can mix time

steps from different sampled trajectories, while copula-based methods resample from marginals and discard the temporal dynamics encoded in the original trajectories. The Schaake shuffle (Clark et al., 2004) partially preserves temporal ranks through phase-aligned contiguous templates, but still replaces the forecaster's own temporal structure. In contrast, GENCOUP adds cross-channel dependence while consistently preserving temporal structure. Additional implementation details for these post-hoc baselines can be found in Appendix A.8.

Source noise modification in probabilistic forecasting. Several probabilistic forecasting methods modify the generative source, mainly to improve marginal predictive performance. TMDM learns a lookback-conditioned mean (Li et al., 2024), NsDiff additionally learns the variance (Ye et al., 2025), TSFlow introduces temporal covariance (Kollovieh et al., 2025), and CW-Gen learns a joint mean and covariance across channels (Yang et al., 2026). GeNCoup has a different objective: improving cross-channel dependence while preserving the forecaster's marginal distributions. It is therefore complementary to TMDM, NsDiff, and TSFlow, as it requires only a Gaussian source with independently sampled channels and makes no assumptions about the channel-wise mean or covariance structure.

3 BACKGROUND

3.1 PROBLEM FORMULATION

Let $w \in \mathbb{R}^{L\times C}$ denote a multivariate lookback window and $Y \in \mathbb{R}^{H\times C}$ its future horizon, where L is the lookback length, H the forecast horizon, and C the number of channels. A sample-based probabilistic forecaster represents $p(Y \mid w)$ by drawing complete future trajectories $\hat{Y}^{(1)}, \dots, \hat{Y}^{(S)} \in \mathbb{R}^{H\times C}$, where each $\hat{Y}^{(s)}$ contains all channels and forecast steps of one joint forecast scenario. Given the observed future $y \in \mathbb{R}^{H\times C}$, our goal is to improve the cross-channel dependence of these trajectories while preserving the forecaster's pointwise predictive marginals and within-channel temporal dependence.

3.2 EVALUATION METRICS

Marginal metrics. Probabilistic forecasts are commonly evaluated using CRPS (Matheson & Winkler, 1976) and NMAE. Both are computed independently at each forecast coordinate (h, n), where $h \in \{1, \dots, H\}$ indexes the forecast step and $n \in \{1, \dots, C\}$ the channel, and are then averaged over forecast steps, channels, and windows. They therefore use $\{\hat{Y}^{(s)}_{h,n}\}_{s=1}^{S}$ while ignoring which values belong to the same sampled trajectory, making them insensitive to how values are coupled across channels or time.

For independent trajectories $\hat{Y}, \hat{Y}' \sim p(Y \mid w)$, $\mathrm{CRPS}_{h,n} = \mathbb{E}|\hat{Y}_{h,n} - y_{h,n}| - \frac{1}{2}\mathbb{E}|\hat{Y}_{h,n} - \hat{Y}'_{h,n}|$, and hence depends only on the marginal distribution of $\hat{Y}_{h,n}$. Normalized MAE (NMAE) similarly uses the ensemble median as the point forecast, $\mathrm{NMAE} = \sum_{h,n} \left|\mathrm{median}_s\, \hat{Y}^{(s)}_{h,n} - y_{h,n}\right| / \sum_{h,n} |y_{h,n}|$. Thus, both metrics depend only on the coordinate-wise predictive distributions and are insensitive to how values are coupled within the sampled trajectories. Preserving these marginals therefore preserves both metrics, up to finite-sample variation.

Joint scores. To evaluate dependence within the sampled trajectories, we use the variogram score (VS) (Scheuerer & Hamill, 2015) and two applications of the energy score (ES) (Gneiting & Raftery, 2007). Unlike CRPS and NMAE, these scores are computed jointly over multiple values from the same sampled trajectory, evaluating how their co-occurrence aligns with the observed future. VS and ES-Chan evaluate cross-channel dependence, whereas ES-Time evaluates the within-channel temporal structure that our method aims to preserve.

The variogram score evaluates a D-dimensional multivariate forecast by taking pairwise differences across its dimensions (Scheuerer & Hamill, 2015). In our setting, we take these dimensions to be the C channels at a fixed forecast step, i.e., $D = C$, compute VS across all channel pairs, and sum

over the horizon:

$$\mathrm{VS}_p = \sum_{h=1}^{H} \sum_{n=1}^{C} \sum_{n'=1}^{C} \left(|y_{h,n} - y_{h,n'}|^p - \mathbb{E}|\hat{Y}_{h,n} - \hat{Y}_{h,n'}|^p \right)^2, \qquad \hat{Y} \sim p(Y \mid w), \tag{1}$$

with $p = \frac{1}{2}$. This evaluates cross-channel dependence at each forecast step without mixing it with temporal dependence. The energy score is the multivariate extension of CRPS for a D-dimensional forecast vector $U \in \mathbb{R}^D$ and observation $u \in \mathbb{R}^D$:

$$\mathrm{ES}(F, u) = \mathbb{E}\|U - u\|_2 - \frac{1}{2}\mathbb{E}\|U - U'\|_2, \qquad U, U' \sim F \text{ independent.} \tag{2}$$

For $D = 1$, ES reduces to CRPS; for $D > 1$, its coordinates are evaluated jointly. **ES-Chan** applies equation 2 to $U = \hat{Y}_{h,\cdot} \in \mathbb{R}^C$ and $u = y_{h,\cdot}$ at each h and averages over the horizon, making it sensitive to cross-channel dependence. **ES-Time** applies it to $U = \hat{Y}_{\cdot,n} \in \mathbb{R}^H$ and $u = y_{\cdot,n}$ for each channel n and averages over channels, making it sensitive to within-channel temporal dependence. We estimate ES from $\{\hat{Y}^{(s)}\}_{s=1}^{S}$ using the finite-ensemble correction of Ferro (2014). ES can have limited ability to distinguish forecasts that differ in their dependence structure, with this limitation becoming harder to detect under sampling noise and higher dimensions (Pinson & Tastu, 2013). Scheuerer & Hamill (2015) further show that ES can be relatively insensitive to incorrectly specified correlations, whereas VS is more discriminative with respect to the correlation structure. We therefore use VS as our main cross-channel metric and ES-Chan as a complementary joint score.

4 METHOD

We present **GeNCoup**, a framework for injecting cross-channel dependence into probabilistic forecast trajectories by coupling the source noise during sampling. At each latent noise index, instead of drawing the channel-wise noise values independently, **GeNCoup** draws them jointly using a cross-channel covariance matrix R, while different noise indices remain sampled independently. This preserves each channel's original noise distribution, so the frozen forecaster retains its complete channel-wise predictive distribution, including marginal uncertainty and within-channel temporal dependence, while introducing dependence only across channels.

Building on this framework, we propose two approaches for choosing R. **GeNCoup-Corr** constructs R from empirical cross-channel correlations (§4.2), whereas **GeNCoup-Dyn** learns the coupling by directly optimizing a joint probabilistic score while constraining R to remain a valid correlation matrix (§4.3).

4.1 NOISE COUPLING FRAMEWORK

We consider a generative forecaster that independently draws a Gaussian source for each channel,

$$X_n \sim \mathcal{N}(\mu_n, \Sigma_n) \quad \text{independently across } n, \qquad \hat{Y}_n = \Phi_w(X_n), \qquad n = 1, \ldots, C. \tag{3}$$

Here, Φ_w is deterministic given the source draw and the lookback. More generally, this formulation applies to any generative forecaster whose randomness is fully specified by X_n and whose subsequent decoding or sampling procedure is deterministic.

Each channel may be sampled from a different Gaussian source distribution:

$$X_n = \mu_n + z_n B_n^\top, \qquad z_n \sim \mathcal{N}(0, I_d), \qquad \Sigma_n = B_n B_n^\top, \tag{4}$$

where the mean μ_n and covariance Σ_n, and hence the factor B_n, may vary across channels.

Coupling independently drawn channel noise. Stack the C independently drawn noise vectors $z_n \sim \mathcal{N}(0, I_d)$ as the rows of $Z \in \mathbb{R}^{C \times d}$. Under the original sampler, each noise coordinate is drawn independently across channels as $Z_{:,k} \sim \mathcal{N}(0, I_C)$.

GeNCoup replaces this independent cross-channel draw with

$$Z'_{:,k} \sim \mathcal{N}(0, R), \qquad k = 1, \ldots, d, \tag{5}$$

independently across k, where $R \in \mathbb{R}^{C\times C}$ is symmetric and positive semi-definite. We now determine the condition on R required to preserve the original distribution $z_i \sim \mathcal{N}(0, I_d)$ of every channel. The mean is unchanged, since each $Z'_{:,k}$ is zero-mean, so $\mathbb{E}[Z'_{i,:}] = 0$. For the covariance, by definition $\mathrm{Cov}(Z'_{i,:}) = \mathbb{E}[(Z'_{i,:} - \mathbb{E}Z'_{i,:})^\top (Z'_{i,:} - \mathbb{E}Z'_{i,:})] = \mathbb{E}[Z'^\top_{i,:} Z'_{i,:}]$. Since $Z'_{:,k} \sim \mathcal{N}(0, R)$, we have $\mathbb{E}[(Z'_{i,k})^2] = \mathrm{Var}(Z'_{i,k}) = R_{ii}$ for every k. Independence across k gives $\mathbb{E}[Z'_{i,k} Z'_{i,\ell}] = \mathbb{E}[Z'_{i,k}]\,\mathbb{E}[Z'_{i,\ell}] = 0$ for $k \neq \ell$. Hence,

$$\mathbb{E}\Big[Z'^\top_{i,:} Z'_{i,:}\Big] = \mathrm{diag}(R_{ii}, \ldots, R_{ii}) = R_{ii} I_d.$$

Thus, preserving the per-channel covariance requires imposing $R_{ii} = 1$ for every channel i (equivalently, R must have unit diagonal). Under this condition, $\mathrm{Cov}(Z'_{i,:}) = I_d$, and hence

$$Z'_{i,:} \sim \mathcal{N}(0, I_d) \overset{d}{=} z_i. \tag{6}$$

The off-diagonal entries of R can therefore introduce cross-channel dependence while leaving each channel's marginal noise distribution unchanged, provided the unit-diagonal constraint is maintained.

Channel-wise preservation. Applying equation 4 to the coupled seed gives $X'_n = \mu_n + Z'_{n,:} B_n^\top$. Since $Z'_{n,:} \overset{d}{=} z_n$, we have $X'_n \sim \mathcal{N}(\mu_n, \Sigma_n)$ and therefore $X'_n \overset{d}{=} X_n$. Because the frozen forecaster Φ_w is deterministic given its source,

$$Z'_{n,:} \overset{d}{=} z_n \implies X'_n \overset{d}{=} X_n \implies \hat{Y}'_n = \Phi_w(X'_n) \overset{d}{=} \hat{Y}_n. \tag{7}$$

Thus, GeNCoup preserves each channel's full predictive trajectory distribution, including its within-channel temporal dependence, while modifying only the dependence across channels.

4.2 Correlation coupling

GeNCoup-Corr sets R from empirical cross-channel correlations. Since time series are often non-stationary, their statistical properties, including cross-channel correlations, may change over time. To account for potential shifts between training and inference, we fuse a global correlation estimated from the training data with a local correlation estimated from the observed lookback:

$$R(\alpha) = (1-\alpha) R_{\mathrm{global}} + \alpha R_{\mathrm{local}}, \qquad \alpha \in [0, 1], \tag{8}$$

where α controls the contribution of the local estimate and is selected on the validation split. Since both terms are correlation matrices, their convex combination remains a valid correlation matrix and satisfies the preservation conditions of Section 4.1.

To estimate either term, we stack W lookback windows of shape $L\times C$ along the temporal dimension into $\mathcal{W} \in \mathbb{R}^{WL\times C}$ and compute the standard Pearson correlation matrix across channels. R_{global} is estimated once from $W = 2048$ training lookbacks, while R_{local} is estimated at inference from the observed lookbacks in the current minibatch. Since R_{local} uses only observed lookbacks and no future values, it introduces no data leakage.

4.3 Learned coupling

GeNCoup-Dyn learns the coupling directly while keeping the probabilistic forecaster frozen. Since the coupled noise is injected inside the sampling process, its parameters can be optimized through the forecaster while keeping the forecaster's weights frozen.

Rather than sampling $Z'_{:,k} \sim \mathcal{N}(0, R)$ directly, the same coupled distribution can be obtained by transforming the original independent noise. Let

$$Z' = AZ, \qquad R = AA^\top.$$

Since $Z_{:,k} \sim \mathcal{N}(0, I_C)$, we have $Z'_{:,k} = AZ_{:,k} \sim \mathcal{N}(0, AA^\top) = \mathcal{N}(0, R)$.

The preservation condition from Section 4.1 requires $R_{nn} = 1$ for every channel n. Since $R_{nn} = (AA^\top)_{nn} = \|A_{n,:}\|_2^2$, this is equivalent to requiring each row of A to have unit norm. We enforce this constraint by parameterizing A through an unconstrained matrix $V \in \mathbb{R}^{C\times C}$:

$$A_{n,:} = \frac{V_{n,:}}{\|V_{n,:}\|_2}, \qquad Z' = AZ, \qquad R = AA^\top. \tag{9}$$

Table 2: **Coupling gain per dataset, averaged over forecasters and horizons.** Percent change vs. the uncoupled arm of the same settings; negative is better. ± is the standard deviation over 3 seeds (2026–2028). All settings of a cell are first averaged per seed, and the mean and standard deviation are then taken over the seeds. Red best, blue second. *Avg. marginal & temporal* averages all cells.

Dataset	Metric (%)	GeNCoup-Dyn (ours)	GeNCoup-Corr (ours)	COBASE	Schaake	SimSch.	GCA	IC
ETTh1	VS	−12.607±0.495	−11.994±0.036	−11.057±0.020	−11.368±0.022	−10.035±0.094	−11.490±0.017	−11.130±0.030
	ES-Chan	− 1.011±0.025	− 1.404±0.007	− 1.337±0.041	− 0.747±0.012	− 1.049±0.005	− 0.805±0.005	− 1.234±0.004
ETTh2	VS	− 7.924±0.230	− 4.395±0.080	− 3.837±0.052	− 3.048±0.016	− 2.294±0.047	− 2.507±0.069	− 2.215±0.065
	ES-Chan	− 0.730±0.230	− 0.937±0.088	− 0.848±0.079	− 0.775±0.021	− 0.408±0.017	− 0.329±0.026	− 0.729±0.028
ETTm1	VS	−13.953±0.425	−11.750±0.037	−10.844±0.035	−11.281±0.024	−10.584±0.026	−11.430±0.025	−11.288±0.031
	ES-Chan	− 1.187±0.038	− 1.436±0.044	− 1.378±0.021	− 0.802±0.002	− 1.009±0.016	− 0.876±0.001	− 1.330±0.002
ETTm2	VS	− 5.690±0.206	− 3.722±0.046	− 4.069±0.029	− 3.125±0.034	− 2.517±0.142	− 2.729±0.060	− 2.592±0.071
	ES-Chan	− 0.713±0.037	− 0.841±0.002	− 0.970±0.011	− 0.743±0.015	− 0.449±0.049	− 0.332±0.017	− 0.765±0.012
Weather	VS	−12.814±0.602	− 7.723±0.037	− 6.033±0.045	− 6.448±0.005	− 6.782±0.033	− 7.763±0.018	− 7.295±0.025
	ES-Chan	− 2.413±0.137	− 1.561±0.012	− 1.160±0.030	− 1.234±0.005	− 1.698±0.022	− 1.310±0.006	− 1.603±0.023
ECL	VS	− 0.112±0.031	− 0.878±0.026	− 1.459±0.035	− 0.860±0.001	− 0.969±0.031	+ 0.053±0.005	− 0.560±0.017
	ES-Chan	+ 0.002±0.011	− 0.509±0.028	− 0.446±0.011	− 0.168±0.002	− 0.224±0.038	+ 0.188±0.003	− 0.399±0.016
Traffic	VS	− 0.365±0.015	− 1.597±0.018	− 1.468±0.003	− 1.175±0.001	− 0.681±0.031	− 1.100±0.001	–
	ES-Chan	+ 0.070±0.050	+ 0.141±0.015	− 0.848±0.004	− 0.145±0.002	+ 0.469±0.011	+ 0.700±0.010	–
Solar	VS	−39.181±0.433	−43.582±0.012	−31.215±0.219	−23.268±0.002	−35.624±0.029	−43.011±0.031	−43.082±0.030
	ES-Chan	− 6.058±0.019	− 5.970±0.029	− 5.356±0.030	− 3.176±0.001	− 5.330±0.031	− 5.054±0.005	− 5.844±0.012
1st count	VS	5	2	1	0	0	0	0
	ES-Chan	2	4	2	0	0	0	0
Avg. marginal & temporal	CRPS	+ 0.012±0.024	+ 0.017±0.015	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.671±0.001	+ 0.000±0.000
	NMAE	+ 0.012±0.043	+ 0.012±0.011	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.962±0.002	+ 0.000±0.000
	ES-Time	+ 0.013±0.016	+ 0.007±0.020	+ 2.765±0.013	+ 0.076±0.004	+ 0.412±0.021	+ 2.421±0.012	+ 3.047±0.022

Thus, $\mathrm{diag}(R) = \mathbf{1}$ by construction, while $R = AA^\top$ guarantees symmetry and positive semi-definiteness. Hence, R remains a valid correlation matrix throughout optimization and preserves the channel-wise noise distributions established in Section 4.1. We initialize $V = I_C$, yielding $A = I_C$ and $R = I_C$, which recovers the original uncoupled sampler. With the forecasting backbone frozen, we optimize only V using a smoothed variogram-score training objective. This objective is the VS of Section 3.2, where we replace the $|a|^p$ term with a smoothed approximation for numerical stability during training. See Appendix A.7 for details: $\min_V \mathcal{L}^{\delta}_{\mathrm{VS}}\left(y,\ \mathrm{forecast}\left(w; AZ\right)\right)$.

5 Experiments

We evaluate GeNCoup-Corr and GeNCoup-Dyn on eight datasets from the standard LTSF benchmark suite: ETTh1, ETTh2, ETTm1, ETTm2 (Zhou et al., 2021), Weather, ECL, Traffic (Wu et al., 2021), and Solar (Lai et al., 2018). All models use a lookback of $L = 96$ and horizons $H \in \{96, 192, 336, 720\}$. Both variants are applied plug-and-play to six frozen forecasters spanning four generative families: k2VAE (Wu et al., 2025), PPM (Li et al., 2026), FMPatch and DiffPatch, additional PatchTST-based baselines (Nie et al., 2023) described in Appendix A.4.1, and the foundation models Sundial (Liu et al., 2025) and Aurora (Wu et al., 2026). We compare against five post-hoc couplers: Schaake (Clark et al., 2004), SimSchaake (Schefzik, 2016), COBASE (Flos et al., 2026), Iman–Conover (IC) (Iman & Conover, 1982), and GCA (Möller et al., 2013). For each baseline and both GeNCoup variants, we report the relative change (in percent) of each score relative to the uncoupled run for the same cell, where a cell is a (forecaster, dataset, horizon, seed) combination. Table 2 reports the results per dataset, averaged across forecasters and horizons. Appendix D provides the per-cell results, and Appendix A gives additional training and evaluation details.

Results. Table 2 shows that the two GeNCoup variants deliver the strongest overall cross-channel performance, obtaining the best VS on most datasets and the best ES-Chan on six of eight datasets. Both variants keep the marginal CRPS/NMAE and temporal dynamics nearly unchanged, shifting ES-Time by only $+0.013\%$ and $+0.007\%$ on average, whereas post-hoc baselines change ES-Time by 0.076–3.047%. The per-dataset breakdown in Table 13 of Appendix D.1 shows that GeNCoup preserves temporal dynamics consistently across datasets, keeping ES-Time changes within about 0.08% of the uncoupled baseline; in comparison, post-hoc ES-Time degradation reaches 6.8% overall and 0.716% for Schaake. Thus, while most post-hoc coupling methods trade temporal structure for cross-channel dependence, GeNCoup improves cross-channel dependence while preserving both marginal distributions and temporal trajectories.

6 ANALYSIS

Ablation of the GeNCoup-Corr Fusion. GENCOUP-CORR fuses global and local correlation matrices to accommodate potentially non-stationary cross-channel dependence. We ablate this fusion mechanism by comparing the validation-selected blend $\alpha^\star$ against two no-fusion baselines: $\alpha = 0$, which uses only the training-set estimate R_{global}, and $\alpha = 1$, which uses only the window-specific estimate R_{local}. We evaluate these setups across all forecasters, datasets, and horizons using seeds 2026–2028. As shown in Table 3, the validation-selected fusion performs best overall, outperforms both no-fusion alternatives for every forecaster, and is best on seven of eight datasets.

Table 3: **Ablation of GeNCoup fusion.** Percent change in VS relative to the uncoupled control; lower is better and the best α is bold. Second lines show the standard deviation over 3 seeds (2026–2028); all settings of a cell are first averaged per seed. Results are reported per dataset and per forecaster. $\alpha^\star$ is selected on validation.

	per dataset								per forecaster						
	ETTh1	ETTh2	ETTm1	ETTm2	Weather	ECL	Traffic	Solar	k2VAE	PPM	FMPatch	DiffPatch	Sundial	Aurora	*avg*
$\alpha = 0$	−11.95	− 2.53	−11.63	− 2.65	− 7.68	− 0.79	− 1.45	−43.56	−13.35	− 8.83	− 6.77	−25.83	− 4.07	− 1.69	−10.10
	±0.05	±0.09	±0.07	±0.12	±0.04	±0.04	±0.03	±0.01	±0.04	±0.06	±0.04	±0.12	±0.01	±0.01	±0.02
$\alpha = 1$	−10.82	**− 4.40**	−11.02	− 3.65	**− 7.91**	− 0.46	− 1.47	−43.00	−13.53	− 9.19	− 6.69	−25.60	− 4.16	− 1.74	−10.17
	±0.04	±0.09	±0.01	±0.03	±0.03	±0.05	±0.04	±0.01	±0.05	±0.04	±0.03	±0.08	±0.02	±0.00	±0.01
$\alpha^\star$	**−11.99**	**− 4.40**	**−11.75**	**− 3.72**	− 7.72	**− 0.88**	**− 1.60**	**−43.58**	**−14.51**	**− 9.31**	**− 7.12**	**−26.10**	**− 4.24**	**− 1.82**	**−10.53**
	±0.04	±0.08	±0.04	±0.05	±0.04	±0.03	±0.02	±0.01	±0.05	±0.04	±0.02	±0.09	±0.02	±0.01	±0.01

Effect of Injected Cross-Channel Dependence on Target Trajectories. Since the variogram score in equation 1 is the sum of all cross-channel pair losses, decomposing it into pairwise contributions allows us to directly assess how the injected dependence $R_{i,j}, i \neq j$, affects each target pair through its loss. We compared each pair's loss before and after coupling, where a decrease indicates that the injected dependence moved the pair closer to the target dependence. Figure 2 illustrates this effect for GENCOUP-CORR and GENCOUP-DYN on the Weather dataset.

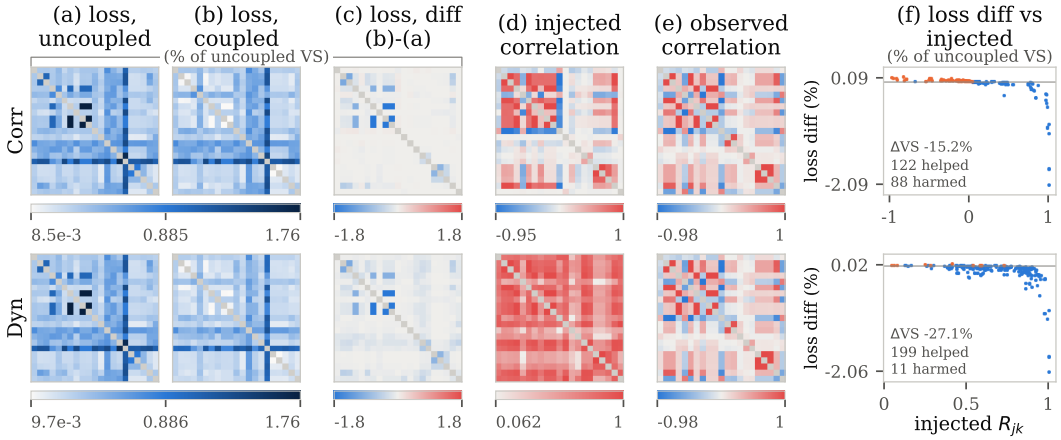


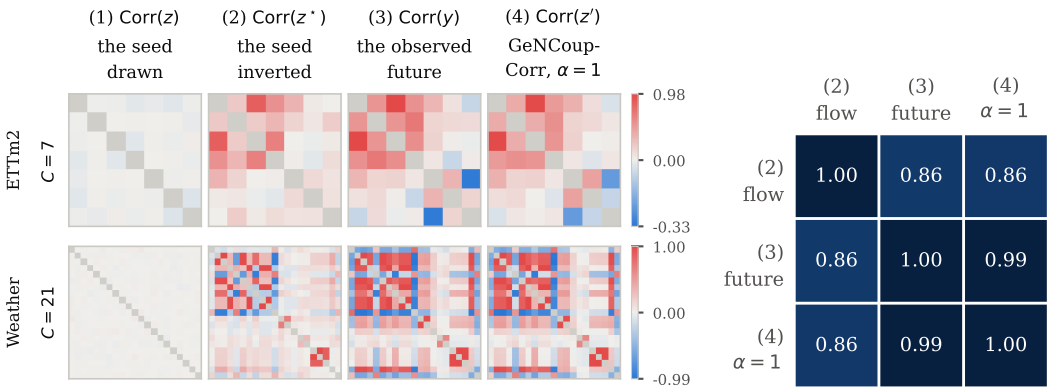
Figure 2: the VS decomposition pair by pair, Weather on k2VAE ($C = 21, 210$ pairs). GENCOUP-CORR at $\alpha = 0$ in the top row, GENCOUP-DYN in the second. Panels **(a)**–**(c)** report each pair's contribution relative to the uncoupled VS: **(a)** uncoupled losses, **(b)** coupled losses, and **(c)** their signed difference. **(d)** the correlation the coupling injected and **(e)** the one the observed futures carry. **(f)** each pair's change against the injected correlation.

Panel **(f)** shows that pairs receiving stronger injected dependence $R_{i,j}$ exhibit larger reductions in pairwise loss. This is also visible across panels **(a)**, **(b)**, and **(d)**, where pairs with large initial loss and strong injected dependence become nearly white after coupling. These reductions show that the coupling recovers cross-channel dependence that is missing or underrepresented in the original forecasts. Between the two variants, GENCOUP-DYN yields a denser off-diagonal coupling than

GENCOUP-CORR, affecting more channel pairs. Correspondingly, a larger fraction of pairs improve, yielding a greater overall reduction in VS. Additional pairwise analyses and results on further datasets are provided in Appendix B.4.

Empirical motivation for the correlation coupling. GENCOUP-CORR sets R to the empirical cross-channel correlation matrix of the observed data. The coupling constraints of Section 4.1 preserve the marginal distributions by requiring a unit diagonal, but leave the off-diagonal dependence unspecified, allowing many valid covariance matrices. We therefore asked whether the empirical correlation captures the source-noise dependence needed to reproduce the cross-channel structure observed in the actual data. To examine this, we performed an inversion experiment that maps the true target futures back to their realized source noise using FMPatch, a flow-matching-based model built on a PatchTST (Nie et al., 2023) backbone, as described in Appendix A.4.1. We integrated the learned ODE backward from the observed trajectory to recover the source-noise seed.

We validated the inversion using round-trip reconstruction tests and then compared the cross-channel correlation of the recovered noise with that of the observed futures. Figure 3 shows examples of the compared correlations, together with their cosine similarities. A Mantel-style permutation test showed that the recovered channel-pair structure is non-random and closer to the observed-future correlation than random reassignment. The recovered off-diagonal correlations also had consistently lower MSE than random-correlation and independence baselines. These results support the empirical data correlation as a suitable choice for R and suggest that dependence introduced between sampled seeds can be reflected in the resulting forecast trajectories. A detailed analysis is provided in Appendix B.2.



(a) the correlation matrices (b) cosine similarity

Figure 3: *(a)* Cross-channel correlations of the following: the sampled seed z, the seed $z^\star$ recovered by the flow-matching inversion, the observed future y, and the coupled seed z' from GENCOUP-CORR, shown for ETTm2 and Weather. *(b)* Mean cosine similarity between the off-diagonal entries of the recovered and coupled seed correlations and the observed-future correlation across eight datasets.

7 CONCLUSION

Seeking to recover cross-channel dependence in forecast trajectories for frozen probabilistic forecasters, we introduce GENCOUP. A simple plug-and-play framework that preserves marginals and temporal dynamics by modifying only the coupling of channel-wise source noise, leaving each channel's noise distribution unchanged. We proposed two variants: GENCOUP-CORR, based on empirical correlations, and GENCOUP-DYN, trained by minimizing a joint score. Pairwise analyses show that injected noise dependence directly affects target trajectory coupling, and the inversion experiment supports the empirical correlation used by GENCOUP-CORR. Across six forecasters and eight datasets, both variants improve joint forecast quality over post-hoc baselines while preserving marginal accuracy and temporal dynamics. Taken together, these results position source-noise coupling as a practical and promising plug-and-play way to improve joint forecasts without degrading the underlying forecaster's existing performance.

REFERENCES

Arjun Ashok, Étienne Marcotte, Valentina Zantedeschi, Nicolas Chapados, and Alexandre Drouin. TACTiS-2: Better, faster, simpler attentional copulas for multivariate time series. In *ICLR*, 2024.

Ethan Baron, Boris Oreshkin, Ruijun Ma, Hanyu Zhang, Kari Torkkola, Michael W. Mahoney, Andrew Gordon Wilson, and Tatiana Konstantinova. Efficiently generating correlated sample paths from multi-step time series foundation models, 2025. arXiv:2510.02224.

Xiaohui Chen, Mengyu Xu, and Wei Biao Wu. Covariance and precision matrix estimation for high-dimensional time series. *The Annals of Statistics*, 2013.

Martyn Clark, Subhrendu Gangopadhyay, Lauren Hay, Balaji Rajagopalan, and Robert Wilby. The Schaake shuffle: A method for reconstructing space–time variability in forecasted precipitation and temperature fields. *Journal of Hydrometeorology*, 2004.

Alexandre Drouin, Étienne Marcotte, and Nicolas Chapados. TACTiS: Transformer-attentional copulas for time series. In *ICML*, 2022.

Robert Engle. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 2002.

Christopher A T Ferro. Fair scores for ensemble forecasts. *Quarterly Journal of the Royal Meteorological Society*, 2014.

Maurits Flos, Bastien François, Irene Schicker, Kirien Whan, and Elisa Perrone. COBASE: A new copula-based shuffling method for ensemble weather forecast postprocessing. *Quarterly Journal of the Royal Meteorological Society*, 2026.

Tilmann Gneiting and Adrian E Raftery. Strictly proper scoring rules, prediction, and estimation. *JASA*, 2007.

Xiaolei He and Weiguo Zhang. Vine copula-based scenario tree generation approaches for portfolio optimization. *Journal of Forecasting*, 2024.

Ronald L. Iman and W. J. Conover. A distribution-free approach to inducing rank correlation among input variables. *Communications in Statistics - Simulation and Computation*, 1982.

Marcel Kollovieh, Abdul Fatir Ansari, Michael Bohlke-Schneider, Jasper Zschiegner, Hao Wang, and Yuyang Wang. Predict, refine, synthesize: Self-guiding diffusion models for probabilistic time series forecasting. In *NeurIPS*, 2023.

Marcel Kollovieh, Marten Lienen, David Lüdke, Leo Schwinn, and Stephan Günnemann. Flow matching with Gaussian process priors for probabilistic time series forecasting. In *ICLR*, 2025.

Guokun Lai, Wei-Cheng Chang, Yiming Yang, and Hanxiao Liu. Modeling long- and short-term temporal patterns with deep neural networks. In *SIGIR*, 2018.

Jinglin Li, Jun Tan, Qi Fang, and Ning Gui. Parametric prior mapping framework for non-stationary probabilistic time series forecasting. In *ICML*, 2026. arXiv:2605.23402.

Yuxin Li, Wenchao Chen, Xinyue Hu, Bo Chen, Baolin Sun, and Mingyuan Zhou. Transformer-modulated diffusion models for probabilistic multivariate time series forecasting. In *ICLR*, 2024.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. In *ICLR*, 2023.

Yong Liu, Guo Qin, Zhiyuan Shi, Zhi Chen, Caiyin Yang, Xiangdong Huang, Jianmin Wang, and Mingsheng Long. Sundial: A family of highly capable time series foundation models. In *ICML*, 2025. arXiv:2502.00816.

Nils Löhndorf. An empirical analysis of scenario generation methods for stochastic optimization. *European Journal of Operational Research*, 2016.

Dhruv Madeka, Lucas Swiniarski, Dean Foster, Leo Razoumov, Kari Torkkola, and Ruofeng Wen. Sample path generation for probabilistic demand forecasting. In *4th Workshop on Mining and Learning from Time Series*, 2018.

Nathan Mantel. The detection of disease clustering and a generalized regression approach. *Cancer Research*, 1967.

James E Matheson and Robert L Winkler. Scoring rules for continuous probability distributions. *Management Science*, 1976.

Annette Möller, Alex Lenkoski, and Thordis L Thorarinsdottir. Multivariate probabilistic forecasting using ensemble Bayesian model averaging and copulas. *Quarterly Journal of the Royal Meteorological Society*, 2013.

Yuqi Nie, Nam H Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. In *ICLR*, 2023.

Youngsuk Park, Danielle Maddix, François-Xavier Aubet, Kelvin Kan, Jan Gasthaus, and Yuyang Wang. Learning quantile functions without quantile crossing for distribution-free time series forecasting. In *AISTATS*, 2022.

Belinda Phipson and Gordon K. Smyth. Permutation *P*-values should never be zero: Calculating exact *P*-values when permutations are randomly drawn. *Statistical Applications in Genetics and Molecular Biology*, 2010.

Pierre Pinson and Julija Tastu. Discrimination ability of the energy score. *Technical University of Denmark*, 2013.

Ilan Price, Alvaro Sanchez-Gonzalez, Ferran Alet, et al. Probabilistic weather forecasting with machine learning. *Nature*, 2025.

Kashif Rasul, Calvin Seward, Ingmar Schuster, and Roland Vollgraf. Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. In *ICML*, 2021.

Andrzej Ruszczyński and Alexander Shapiro. Stochastic programming models. *Handbooks in Operations Research and Management Science*, 2003.

David Salinas, Michael Bohlke-Schneider, Laurent Callot, Roberto Medico, and Jan Gasthaus. High-dimensional multivariate forecasting with low-rank Gaussian copula processes. In *NeurIPS*, 2019.

David Salinas, Valentin Flunkert, Jan Gasthaus, and Tim Januschowski. DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, 2020.

Roman Schefzik. A similarity-based implementation of the Schaake shuffle. *Monthly Weather Review*, 2016.

Michael Scheuerer and Thomas M Hamill. Variogram-based proper scoring rules for probabilistic forecasts of multivariate quantities. *Monthly Weather Review*, 2015.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *ICLR*, 2021.

Oliver Stover, Paromita Nath, Pranav Karve, Sankaran Mahadevan, and Hiba Baroud. Dependence structure learning and joint probabilistic forecasting of stochastic power grid variables. *Applied Energy*, 2024.

Ruofeng Wen and Kari Torkkola. Deep generative quantile-copula models for probabilistic forecasting. In *ICML Time Series Workshop*, 2019. arXiv:1907.10697.

Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In *NeurIPS*, 2021.

Xingjian Wu, Xiangfei Qiu, Hongfan Gao, Jilin Hu, Bin Yang, and Chenjuan Guo. k^2vae: A koopman-kalman enhanced variational autoencoder for probabilistic time series forecasting. In *ICML*, 2025. Spotlight; arXiv:2505.23017.

Xingjian Wu, Jianxin Jin, Wanghui Qiu, Peng Chen, Yang Shu, Bin Yang, and Chenjuan Guo. Aurora: Towards universal generative multimodal time series forecasting. In *ICLR*, 2026. arXiv:2509.22295.

Yanfeng Yang, Siwei Chen, Xiaoyu Hu, Zhaoyang Shen, Le Zhang, Yuxuan Sun, Wei Li, Bo Chen, and Kenji Fukumizu. Conditionally whitened generative models for probabilistic time series forecasting. In *ICLR*, 2026.

Weiwei Ye et al. Non-stationary diffusion for probabilistic time series forecasting. In *ICML*, 2025. arXiv:2505.04278.

Jiawen Zhang, Xumeng Wen, Zhenwei Zhang, Shun Zheng, Jia Li, and Jiang Bian. ProbTS: Benchmarking point and distributional forecasting across diverse prediction horizons. In *Advances in Neural Information Processing Systems*, 2024.

Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *AAAI*, 2021.

A Implementation Details

A.1 Framework and hardware

Our experiments are built on the ProbTS benchmarking framework (Zhang et al., 2024). We use its data loading, train/val/test splits, standardization, rolling-window evaluation protocol, forecasting interface, and scoring routines as the common experimental framework for all reported methods. GENCOUP-CORR, GENCOUP-DYN, and every post-hoc coupling baseline are implemented inside this framework, so methodological differences are isolated to the coupling rule: within a cell, every arm receives the same lookback windows, forecaster weights, and sampling seed.

All runs use a single GPU. Sample generation for the uncoupled control, post-hoc baselines, and GENCOUP-CORR runs on an NVIDIA RTX 4090 (24 GB). GENCOUP-DYN requires more memory because it trains A through the frozen sampler. When 24 GB was not enough, mainly for Sundial, Aurora, ECL, and Traffic runs, we used a 48 GB NVIDIA RTX 6000. Also Aurora on Traffic at $H = 720$ was ran on 48 GB NVIDIA RTX 6000.

A.2 Benchmarks and evaluation protocol

Table 4 lists the eight long-term datasets. Their channel counts range from $C = 7$ to $C = 862$, testing the coupling from small to very wide cross-channel settings.

Dataset	Domain	C	Length	Resolution
ETTh1, ETTh2	Electricity transformer	7	17,420	1 hour
ETTm1, ETTm2	Electricity transformer	7	69,680	15 min
Weather	Meteorology	21	52,696	10 min
Solar	Photovoltaic production	137	52,560	10 min
ECL	Electricity consumption	321	26,304	1 hour
Traffic	Road-sensor occupancy	862	17,544	1 hour

Table 4: Benchmark statistics. C is the number of channels, Length the number of timesteps.

All train/validation/test splits and preprocessing follow the default ProbTS framework (Zhang et al., 2024). All runs use lookback $L = 96$, horizons $H \in \{96, 192, 336, 720\}$, and 100 predictive samples per test window. Test windows are taken from the test partition with the ProbTS default rolling stride of 96 steps.

For each setting, defined by a forecaster, dataset, horizon, and sampling seed, we run both proposed GENCOUP variants and all compared baselines. All arms in a setting use the same frozen forecaster checkpoint. To avoid repeated sampling, post-hoc baselines are applied to the locally cached forecast samples from the corresponding evaluation run, while GENCOUP arms share the same random draw as their uncoupled control before the coupling transform. Thus all reported percentages are paired comparisons relative to the uncoupled sampler.

A.3 Hyperparameter search

For GENCOUP-CORR, α is a single hyperparameter that controls the fusion between the training-set correlation R_{global} and the local correlation R_{local}. We perform a grid search over $\alpha \in \{0, 0.1, \ldots, 1.0\}$ on the validation split for each (forecaster, dataset, horizon) setting. Since the VS score is not normalized across horizons, we convert the validation scores within each horizon to ranks and, for each (forecaster, dataset) pair, choose the α with the best average validation rank. The selected α is then held fixed across horizons and sampling seeds for the test runs.

Table 5 lists the selected α values used by the reported GENCOUP-CORR runs.

A.4 Forecasting backbones

We retrain the published in-domain forecasters, K^2VAE (Wu et al., 2025) and PPM (Li et al., 2026), separately for each (dataset, horizon) cell. We keep their released hyperparameters unchanged and train each run with seed 2026 from the authors' configurations, so GENCOUP is

Table 5: **The blend parameter α of equation 8 used by every reported GeNCoup-Corr run.** One value per (forecaster, dataset), selected on the validation split and then held fixed across horizons and seeds; $\alpha = 0$ is the correlation fitted on the training split alone and $\alpha = 1$ the one estimated from the lookbacks at inference. Appendix B.3 measures what the choice is worth.

Forecaster	ETTh1	ETTh2	ETTm1	ETTm2	Weather	ECL	Traffic	Solar
k2VAE	0	1	0	1	0.5	0.1	1	0.1
PPM	1	1	0.1	1	0.5	0.1	0.2	0.1
FMPatch	1	0.8	1	0.8	0	0	0	0.1
DiffPatch	0	1	0	1	0	1	0.9	1
Sundial	1	0	1	0	0.9	0	0	0.6
Aurora	1	0	1	0	0	0	0	0.7

applied to the backbone specified by the original method. The foundation forecasters, Sundial (Liu et al., 2025) and Aurora (Wu et al., 2026), are loaded frozen from Hugging Face checkpoints `thuml/sundial-base-128m` and `DecisionIntelligence/Aurora`, with no fine-tuning. We also train FMPatch and DiffPatch, two PatchTST-based generative forecasters introduced for this paper, which are included in the main experiments; FMPatch is additionally used in the inversion analysis of Appendix B.2.1. Details are given in Appendix A.4.1. All trained forecasters are trained once with seed 2026 for each (dataset, horizon) cell. The same checkpoint is used in all experiments; for evaluation seeds 2026–2028, only the evaluation seed is changed, not the trained model checkpoint.

A.4.1 FMPatch and DiffPatch

FMPatch and DiffPatch are two generative forecasters built on the patching encoder of PatchTST (Nie et al., 2023). The lookback is supplied in an *inpainting* style: the normalized lookback is appended before the noised horizon and is never noised itself, so the network reads its condition from the same tensor it denoises and needs no separate conditioning path. Since they inherit PatchTST's channel-independent design, each channel is processed separately and draws its own noise independently, as required by equation 3.

The two differ only in the training and sampling framework. FMPatch is trained by linear-path conditional flow matching (Lipman et al., 2023): with $z \sim \mathcal{N}(0, I)$ and $t \sim U(0, 1)$ the network reads $x_t = t\,y + (1 - t)\,z$ and regresses the velocity $y - z$; sampling integrates the learned field from $t = 0$ to 1 with a 20-step Euler solver, the lookback pinned at every step. DiffPatch is trained as a DDPM with $T = 1000$ steps and a linear β schedule from 10^{-4} to 0.02, the network regressing the noise ϵ from $x_t = \sqrt{\bar{\alpha}_t}\,y + \sqrt{1 - \bar{\alpha}_t}\,\epsilon$; sampling uses deterministic DDIM (Song et al., 2021) with $\eta = 0$ and 20 steps. In both, no noise is redrawn inside the sampling process, so the forecast is a deterministic map of the seed, as equation 3 requires.

A.5 Installing GeNCoup in a generative forecaster

Installing GeNCoup requires replacing the forecaster's source-noise draw. When the original sampler draws channel-wise independent noise $z \sim \mathcal{N}(0, I)$, the coupled sampler instead passes $z' = Az$, with $AA^\top = R$, so each seed coordinate has cross-channel covariance R. All downstream sampling code is unchanged.

A.6 Where the coupling is installed

Each forecaster draws its source noise in a model-specific layout, but the draw can be reshaped and permuted into a common [groups $\times$ C $\times$ d] view. Here groups collects the independent forecast instances, typically the batch size times the number of samples S; C indexes channels; and d is the model-specific noise dimension, such as a latent dimension or a horizon coordinate in flow and diffusion models. GeNCoup applies one matrix product with A along the channel axis in this view, then inverts the same permutation and reshape before passing the seed back to the original sampler.

For the diffusion and flow-matching models, the coupled draw is the initial noise before the denoising or flow path begins. For K^{2}VAE, it is the standardized draw formed after the Koopman–Kalman

mean and standard deviation are computed and before the decoder. For PPM, it is the latent reparameterization noise before $\mu + \epsilon\sigma$. For Sundial and Aurora, it is the flow-head start point, with Aurora folding its output-token axis into the model-specific noise dimension.

A.7 TRAINING GENCOUP-DYN

GENCOUP-DYN trains A against the variogram score through the frozen forecaster, so the gradient runs from the score back through every sampling step to the seed while the forecaster weights stay fixed. Directly minimizing the original variogram score creates a problem when pairwise channel differences are close to zero. The variogram score uses $|a|^p$ with $p = \frac{1}{2}$. Its derivative is $\frac{d}{da}|a|^p = p|a|^{p-1}\operatorname{sign}(a)$, which becomes very large near $a = 0$ and is undefined at $a = 0$. Therefore, nearly equal channel values can lead to unstable gradients during training. We therefore replace $|a|^p$ during training by the smoothed power

$$\psi_\delta(a) = \left(a^2 + \delta\right)^{p/2}, \qquad \psi'_\delta(a) = p\, a \left(a^2 + \delta\right)^{p/2-1}, \tag{10}$$

which recovers $|a|^p$ when $\delta = 0$. The optimized objective is the same variogram score as equation 1, with this replacement:

$$\mathcal{L}^{\delta}_{\mathrm{VS}} = \sum_{h=1}^{H}\sum_{n=1}^{C}\sum_{n'=1}^{C}\left(\psi_\delta(y_{h,n} - y_{h,n'}) - \mathbb{E}\,\psi_\delta(\hat{Y}_{h,n} - \hat{Y}_{h,n'})\right)^2, \qquad \hat{Y} = \mathrm{forecast}(w; AZ). \tag{11}$$

We use $\delta = 10^{-6}$, this formulation is used only for optimization; every table reports the original variogram score equation 1.

We initialize V at I and optimize it with Adam at learning rate 0.05, gradient clipping at 1. The checkpoint is selected by validation score.

A.8 BASELINE IMPLEMENTATIONS

All post-hoc baselines operate on the same cached ensemble from the frozen forecaster. Thus, within each cell, the forecaster, lookback window, sampling seed, and number of predictive samples are fixed, and only the coupling operation varies. This paired design isolates the effect of the post-processing step.

These methods are described as two-stage procedures: a predictive model first produces calibrated marginal forecasts, and a dependence post-processor then imposes joint structure. In our setting the frozen forecaster supplies the first-stage predictive ensemble, so we implement only the dependence stage.

The Schaake shuffle (Clark et al., 2004), SimSchaake (Schefzik, 2016), Gaussian copula (Möller et al., 2013), and COBASE (Flos et al., 2026) were introduced for ensemble weather post-processing and originally implemented in R. Their coupling rules are not weather-specific: they require only an ensemble, forecast coordinates, marginal samples, and, when applicable, a historical template pool. We therefore adapt them to the general multivariate forecasting setting, as detailed in this subsection, and translate the available R implementations to Python so all methods run in a single evaluation pipeline.

Reordering arms. Each reordering method builds a template whose ranks are used to reorder the ensemble at each time step. For each (window, time step, channel), sorting the S forecast values defines their ranks. The template supplies the target rank of each output member, and the original forecast values are permuted to match those template ranks, with ties broken at random. Thus the operation changes only which ensemble member receives which value at each coordinate. Each channel keeps exactly the forecaster's original set of S values, so its marginal samples, and therefore CRPS, are unchanged. However, the sampled trajectories themselves are modified: a reordered member can receive values that came from different original trajectories at different time steps. Since the same template member index is used across channels, the template's cross-channel rank pattern is transferred to the reordered ensemble.

Schaake shuffle (Clark et al., 2004). Clark et al. (2004) designed the Schaake shuffle to recover both cross-location and time-step dependence in daily weather ensembles. It does so indirectly: each member is assigned a historical date within seven days of the same calendar day from other years, and observations after that date provide the ranks for all coordinates and time steps. Thus the temporal structure comes from seasonal past trajectories rather than from an explicit temporal model. In our benchmarks, sampled every 10 minutes to every hour (Table 4), we adapt this by drawing training windows whose first forecast step has the same time of day, using a ± 7-day window over previous calendar months rather than years (since our datasets span at most about two years). This makes Schaake the most favorable reordering baseline for temporal preservation: its ranks across time steps come from contiguous historical reference windows, phase-aligned by time of day, which are meant to provide an empirical temporal template. Because that structure is inherited from historical windows, Schaake is also sensitive to non-stationarity in the dataset. In practice the temporal cost is still non-negligible: Schaake has the smallest ES-Time degradation among post-hoc baselines, but its degradation remains larger than GENCOUP's in the main experiment table (Table 2). It is also less consistent across datasets, as shown by the per-dataset ES-Time breakdown in Appendix Table 13.

SimSchaake (Schefzik, 2016). Schefzik (2016) choose historical cases whose archived forecasts are most similar to the current forecast, then use their observations as the rank template. The same idea is applied to the training-reference set. Each reference window x is represented by the forecaster's ensemble mean and standard deviation at every (time step, channel), and the S most similar windows under Schefzik's criterion form the template. This selection is performed over whole windows, rather than independently per time step as in the original setting, so each selected template remains an observed trajectory across the full horizon and can also carry temporal ranks.

Iman–Conover (Iman & Conover, 1982). Iman & Conover (1982) induce a prescribed rank correlation by reordering samples according to a constructed template. Here, the target R is the Pearson correlation between channels in the training-split futures, pooled over windows and time steps. An $S \times C$ score matrix matched to R supplies the reordering template. Since the construction requires inverting a $C \times C$ empirical correlation estimated from S score samples, it is used only when $S > C + 1$ (§A.9). The reported results use the `probabilit` package.

Gaussian copula (Möller et al., 2013). Möller et al. (2013) impose cross-variable dependence with a Gaussian copula. Here the copula uses one fixed correlation matrix, the training-reference correlation R. For each (window, time step), S Gaussian copula samples are drawn over the channels and mapped back through each channel's empirical inverse CDF over the forecast ensemble. Unlike the reordering arms, this preserves marginals only in distribution rather than value-for-value. The reported results use the `statsmodels` sampler.

COBASE (Flos et al., 2026). Flos et al. (2026) fit a parametric copula and use its draws only as a rank template, so the ensemble is reordered and its marginals stay exact. In the original method, the copula is refit each day using the preceding thirty days. Here, the fit is restricted to the training split; since the days preceding a test window lie outside that split, a single copula is fit on the M randomly drawn training-reference windows. This implementation uses a Gaussian copula, estimated by inverting Kendall's τ on residuals: their observed futures, standardised by the forecaster's ensemble mean and standard deviation on each window and pooled over windows and time steps. For each (window, time step), S fresh copula draws are used to construct the rank template.

A.9 EVALUATION NOTES

All post-hoc arms are evaluated using 100 forecast trajectories per window. However, Iman–Conover is defined only for $S > C + 1$, which the default ensemble does not satisfy on the wide panels. On Solar and ECL we draw it at $S = C + 2$ and then score a uniformly random 100-member subset of the coupled ensemble, so that every arm in a cell is scored at the same ensemble size. On Traffic ($C = 862$) the corresponding draw was beyond our compute budget, and IC is reported on the remaining seven datasets.

B DETAILED ANALYSIS

B.1 DO GENERATED ENSEMBLE TRAJECTORIES ALREADY CARRY STRONG TEMPORAL DYNAMICS?

Our research focuses on sample-based probabilistic forecasting, where a model draws whole trajectories rather than a set of independent per-step marginals. Many of the backbones we consider are Channel-Independent: they process each channel separately, whether trained in-domain (Kollovieh et al., 2025; Li et al., 2026) or pretrained as foundation models (Liu et al., 2025; Wu et al., 2026), and so omit cross-channel interactions by design. They do, however, model the temporal dynamics within each channel explicitly, so the trajectories they draw should already carry that structure. The two axes, therefore, start from opposite positions: Cross-channel dependence is a component of the joint predictive distribution these models do not model, leaving a correspondingly large amount to gain, while the temporal axis is one they are built to capture. This raises the question:

> *Is the temporal structure these models already produce good enough that the remaining headroom lies mainly on the cross-channel axis?*

To explore this question, we compare the temporal structure of the ensemble a forecaster already produces against trajectories reconstructed from a quantile representation of that same ensemble. The quantiles are a summary of the ensemble: they retain its marginal distribution at every time step and channel, and discard only the coupling that ties those marginals into trajectories. Reconstructing from them, therefore, leaves marginal performance essentially unchanged and varies the temporal coupling alone, which is what makes the two ensembles comparable. For the reconstruction, we use the method of Baron et al. (2025), developed for multi-step quantile forecasters, which predict a marginal per time step and no joint at all, so a trajectory has to be assembled from those marginals after the fact. The forecasters studied here are sample-based and emit trajectories directly, so the method does not recover a joint they lack but replaces one they already have. Reducing their ensembles to quantiles is what places them in the setting the method was built for, and asking whether the replacement improves on the original is the test we target. Both ensembles are scored with the ES-Time of §3.2.

Setup. We reuse the experimental setup of §5: the same six frozen forecasters and eight benchmarks, a lookback of $L = 96$, horizons $H \in \{96, 192, 336, 720\}$, predictive distributions of 100 samples, and three sampling seeds (2026–2028). We follow the setup of Baron et al. (2025): the quantile summary uses nine knots at levels $0.1, \ldots, 0.9$, the grid CRPS is scored on, and the inversion back to samples is the incremental quantile function of Park et al. (2022). We report ES-Time, which measures the temporal dynamics directly, alongside CRPS, which measures what the quantile summary costs.

Reconstruction and baseline. The method of Baron et al. (2025) turns the marginals back into trajectories with a Gaussian copula along the horizon. It is defined for a single series, so we apply it to each channel independently: every channel gets its own correlation matrix $\Sigma_{ij} = \rho^{|i-j|}$, where ρ is the lag-1 autocorrelation of that channel's lookback, so the dependence between two time steps decays with the distance between them. We also run the independence baseline of the same work, which draws every (time step, channel) from its own marginal with no coupling at all. The two arms differ only in whether the copula is applied, so the gap between them measures the copula alone. The gap to the original ensemble measures what the AR(1) method costs against the trajectories the forecaster drew.

Results. Table 6 reports the ES-Time change per model and dataset with both margins, and Table 7 the same for CRPS. The AR(1) method improves ES-Time in 33% of the cells we score, and loses by 1.040 pooled against the trajectories the forecasters drew. Only PPM and Sundial improve on average, at -0.687 and -0.377. The largest share of the wins comes from Sundial, which improves in 83% of its cells but by under half a percent on average across datasets, while the competing foundation model, Aurora, is worse by more than a percent ($+1.145$) and wins no cell at all. Against the independence baseline, the AR(1) method does insert better temporal structure, improving ES-Time

Table 6: **ES-Time change per model and dataset.** Percent change vs. the forecaster's own ensemble, averaged over horizons; negative is better. ± is the standard deviation over the three seeds; all settings of a cell are first averaged per seed. *AR(1)* is the copula of Baron et al. (2025); *Indep.* is that paper's independence baseline, every (lead, channel) drawn from its own marginal with no coupling. Margins average over every setting of the row or column. *win %* is the percentage of runs in which that arm improves the score.

Model	Arm	ETTh1	ETTh2	ETTm1	ETTm2	Weather	ECL	Traffic	Solar	*mean*	*win %*
k2VAE	AR(1)	+ 2.14±0.09	+ 1.84±0.14	+ 1.33±0.02	+ 0.46±0.05	− 1.18±0.03	+ 0.62±0.00	+ 0.43±0.00	− 2.03±0.01	+ 0.45±0.00	29
	Indep.	+ 2.67±0.03	+ 2.27±0.05	+ 2.66±0.01	+ 0.76±0.00	+ 2.10±0.02	+ 1.07±0.01	+ 0.58±0.00	+ 1.22±0.01	+ 1.66±0.01	3
PPM	AR(1)	+ 0.70±0.18	+ 1.44±0.06	− 0.33±0.07	+ 0.71±0.05	− 0.12±0.10	− 1.72±0.01	− 4.39±0.01	− 1.77±0.02	− 0.69±0.03	61
	Indep.	+ 1.01±0.14	+ 2.13±0.06	+ 0.80±0.09	+ 1.66±0.02	+ 0.56±0.09	− 1.58±0.00	− 4.38±0.01	− 0.29±0.01	− 0.01±0.02	31
FMPatch	AR(1)	+ 1.10±0.04	+ 2.28±0.08	+ 1.12±0.04	+ 2.71±0.09	+ 1.72±0.08	+ 0.69±0.01	+ 1.42±0.00	+ 1.45±0.04	+ 1.56±0.04	0
	Indep.	+ 1.56±0.05	+ 2.89±0.11	+ 1.77±0.04	+ 3.65±0.05	+ 4.73±0.02	+ 0.97±0.00	+ 1.41±0.00	+ 6.20±0.03	+ 2.79±0.01	0
DiffPatch	AR(1)	+ 3.92±0.10	+11.28±0.04	+ 4.78±0.03	+13.54±0.10	+ 1.07±0.10	+ 2.18±0.01	− 4.68±0.01	+ 1.06±0.02	+ 4.14±0.02	22
	Indep.	+ 5.05±0.16	+13.03±0.06	+ 7.92±0.03	+17.39±0.12	+14.02±0.04	+ 3.73±0.00	− 3.77±0.01	+ 5.05±0.01	+ 7.80±0.03	12
Sundial	AR(1)	− 0.71±0.03	− 0.55±0.02	− 0.22±0.04	− 0.38±0.02	+ 0.21±0.02	− 0.79±0.01	− 0.53±0.01	− 0.05±0.00	− 0.38±0.01	83
	Indep.	− 0.78±0.01	− 0.58±0.04	− 0.32±0.03	− 0.51±0.01	− 0.09±0.01	− 0.85±0.00	− 0.61±0.01	+ 0.23±0.01	− 0.44±0.01	94
Aurora	AR(1)	+ 0.69±0.01	+ 0.97±0.04	+ 0.76±0.02	+ 1.74±0.03	+ 1.65±0.01	+ 0.42±0.01	+ 0.50±0.00	+ 2.43±0.01	+ 1.15±0.00	0
	Indep.	+ 0.50±0.01	+ 0.87±0.04	+ 0.19±0.00	+ 1.39±0.01	+ 0.94±0.01	+ 0.23±0.00	+ 0.34±0.00	+ 1.04±0.00	+ 0.69±0.01	9
mean	AR(1)	+ 1.31±0.06	+ 2.87±0.04	+ 1.24±0.03	+ 3.13±0.04	+ 0.56±0.04	+ 0.23±0.01	− 1.21±0.00	+ 0.13±0.01	+ 1.04±0.02	33
	Indep.	+ 1.67±0.06	+ 3.43±0.03	+ 2.17±0.03	+ 4.06±0.01	+ 3.71±0.02	+ 0.59±0.00	− 1.07±0.00	+ 2.07±0.01	+ 2.08±0.01	25

Table 7: **CRPS change per model and dataset.** Percent change vs. the forecaster's own ensemble, averaged over horizons; negative is better. ± is the standard deviation over the three seeds; all settings of a cell are first averaged per seed. *AR(1)* is the copula of Baron et al. (2025); *Indep.* is that paper's independence baseline, every (lead, channel) drawn from its own marginal with no coupling. Margins average over every setting of the row or column. *win %* is the percentage of runs in which that arm improves the score. A copula leaves marginals unchanged, so the gap between the two arms is not an effect of the temporal coupling; what both pay is the price of drawing levels afresh from the quantile summary rather than inheriting the ensemble's ranks.

Model	Arm	ETTh1	ETTh2	ETTm1	ETTm2	Weather	ECL	Traffic	Solar	*mean*	*win %*
k2VAE	AR(1)	+1.30±0.10	+0.79±0.04	+1.18±0.02	+0.54±0.05	+1.05±0.05	+0.90±0.01	+0.75±0.01	+0.93±0.02	+0.93±0.02	1
	Indep.	+1.23±0.05	+0.76±0.01	+1.15±0.01	+0.53±0.00	+1.02±0.01	+0.89±0.01	+0.75±0.00	+0.91±0.00	+0.91±0.01	0
PPM	AR(1)	+0.84±0.10	+0.79±0.01	+0.61±0.04	+0.70±0.04	+0.33±0.04	+0.54±0.00	+0.43±0.00	+0.63±0.02	+0.61±0.02	0
	Indep.	+0.79±0.05	+0.80±0.01	+0.57±0.01	+0.71±0.00	+0.36±0.00	+0.55±0.00	+0.43±0.00	+0.64±0.00	+0.61±0.01	0
FMPatch	AR(1)	+0.85±0.07	+0.60±0.04	+0.81±0.04	+0.81±0.09	+0.69±0.07	+0.63±0.01	+0.60±0.00	+0.99±0.07	+0.74±0.03	0
	Indep.	+0.82±0.06	+0.64±0.04	+0.76±0.01	+0.81±0.01	+0.76±0.00	+0.64±0.00	+0.60±0.00	+0.98±0.01	+0.74±0.00	0
DiffPatch	AR(1)	+1.63±0.08	+1.98±0.05	+1.62±0.07	+2.06±0.03	+2.19±0.06	+1.45±0.01	+1.43±0.01	+1.44±0.03	+1.73±0.01	0
	Indep.	+1.58±0.07	+2.00±0.02	+1.56±0.01	+2.02±0.02	+2.23±0.01	+1.46±0.00	+1.44±0.00	+1.44±0.01	+1.72±0.01	0
Sundial	AR(1)	+0.26±0.03	+0.14±0.03	+0.37±0.04	+0.20±0.02	+0.16±0.04	+0.27±0.01	+0.38±0.00	+0.36±0.01	+0.27±0.01	1
	Indep.	+0.23±0.02	+0.15±0.02	+0.34±0.02	+0.21±0.00	+0.17±0.00	+0.28±0.00	+0.38±0.00	+0.35±0.00	+0.26±0.00	0
Aurora	AR(1)	+0.12±0.01	+0.05±0.05	+0.14±0.02	+0.05±0.02	+0.05±0.01	+0.18±0.00	+0.23±0.00	+0.09±0.01	+0.11±0.01	3
	Indep.	+0.12±0.02	+0.07±0.01	+0.12±0.00	+0.07±0.00	+0.06±0.00	+0.18±0.00	+0.23±0.00	+0.08±0.00	+0.12±0.00	0
mean	AR(1)	+0.83±0.06	+0.73±0.02	+0.79±0.03	+0.73±0.04	+0.75±0.03	+0.66±0.00	+0.64±0.00	+0.73±0.02	+0.73±0.01	1
	Indep.	+0.79±0.04	+0.74±0.01	+0.75±0.01	+0.72±0.00	+0.77±0.00	+0.66±0.00	+0.64±0.00	+0.72±0.00	+0.72±0.00	0

by 1.061 on average, but not enough to compete with the trajectories the sample-based forecasters draw themselves.

The trajectories these forecasters already generate, therefore, appear to carry useful temporal dynamics. Summarizing the ensemble of trajectories into quantiles and restoring trajectories from that summary with the baselines tested here failed to improve those dynamics. This suggests that the larger remaining headroom lies on the cross-channel axis.

B.2 Empirical motivation for the correlation coupling

GenCoup-Corr chooses R to be the empirical cross-channel correlation matrix of the observed data. This choice satisfies the requirements of the coupling framework in Section 4.1, but it is not the only one that does: the framework admits any symmetric positive semidefinite matrix with unit diagonal, leaving many possible dependence structures that preserve the channel-wise seed marginals. The requirements alone therefore do not determine R, which raises a natural question:

Why should the empirical correlation matrix be a meaningful choice for R?

To investigate this question, we conduct an experiment that approximates the dependence that would need to be present in the source noise for the frozen forecaster to reproduce the dependence observed in the data. Specifically, we reverse the generative sampling process, mapping the true future trajectories of the test set from data space back to their corresponding source noise. We perform this experiment with FMPatch, a flow-matching-based model built on a PatchTST (Nie et al., 2023) backbone and described in Appendix A.4.1.

The reverse mapping recovers the source noise by integrating the learned ODE backward from the observed trajectory (Algorithm 1; §B.2.1). We verify this procedure with a round-trip test (§B.2.2), mapping each trajectory to noise and back to data space and measuring the reconstruction error. We then examine the cross-channel dependence of the recovered noise (§B.2.3) as a proxy for the source dependence required to reproduce the observed joint structure, and therefore for what R has to supply.

B.2.1 Reversing a real forecast into noise

FMPatch (Appendix A.4.1) draws $z_n \sim \mathcal{N}(0, I_H)$ per channel, so $\mu_n = 0$ and $B_n = I$ in equation 4 and the forecast is $Y_n = \Phi_{n,w}(z_n)$. Inverting that map on an observed test future y_n gives the true seed behind it, $z_n^\star = \Phi_{n,w}^{-1}(y_n)$: the draw an optimal coupling would have to install for the frozen sampler to return that future. The inversion itself runs with no coupling installed, and Algorithm 1 gives the procedure and its hyperparameters. The forecaster is channel-wise, so the inversion runs independently per (window, channel).

Algorithm 1 Flow matching: the sampler, and its inversion by reverse Euler

Require: lookback w, velocity field v_θ; steps $S = 20$, the count the sampler was trained and evaluated at; $\Delta t = 1/S$; double precision throughout

Sampler $\Phi_{n,w}$: seed $\to$ forecast

$x \leftarrow z$ ▷ $z \sim \mathcal{N}(0, I_H)$, drawn once

for $k = 0, 1, \ldots, S-1$ **do**

  $x \leftarrow x + \Delta t\, v_\theta(w, x, k\Delta t)$ ▷ explicit Euler, departure time

end for

return $y \leftarrow x$

Inversion $\Phi_{n,w}^{-1}$: observed future $\to$ seed

$x \leftarrow y$

for $k = S, S-1, \ldots, 1$ **do**

  $x \leftarrow x - \Delta t\, v_\theta(w, x, k\Delta t)$ ▷ same field, arrival time, negated step

end for

return $z^\star \leftarrow x$

B.2.2 Validity of the inversion

To validate the inversion we perform a round-trip test, where each true future of the test set is mapped back to a seed and pushed forward through the sampler again, and the reconstruction is compared with the future it started from. We report two per-sample error metrics and one reference scale,

$$\mathrm{MSE}_n = \frac{\|r_n\|_2^2}{H}, \qquad \text{rel-L2}_n = \frac{\|r_n\|_2}{\|y_n\|_2}, \qquad r_n = \Phi_{n,w}(\Phi_{n,w}^{-1}(y_n)) - y_n,$$

$$\%\text{ of MSE}_{\text{pred}} \;=\; 100 \cdot \frac{\overline{\text{MSE}}}{\text{MSE}_{\text{pred}}}, \qquad \text{MSE}_{\text{pred}} \;=\; \mathbb{E}\big[\|\bar{\hat{y}} - y\|_2^2/H\big], \qquad \bar{\hat{y}} \;=\; \tfrac{1}{S}\textstyle\sum_{s=1}^{S} \hat{y}^{(s)}.$$

MSE measures absolute reconstruction error, whereas rel-L2 normalizes the error by the magnitude of the true future and is therefore easier to interpret: a rel-L2 of 0.05 corresponds to a 5% relative error, while a value of 1 indicates an error comparable to the future itself. We additionally report a third quantity, MSE_{pred}, to place the round-trip error in the context of the model's forecasting error. It is the same model's error when run normally from the lookback, with its S sampled trajectories averaged into a point prediction and evaluated on the test split in the same normalized space as the main sweep. A round-trip MSE substantially below MSE_{pred} indicates that the reconstruction error is small relative to the model's intrinsic forecasting error. Table 8 reports all three quantities for each dataset.

Results. As Table 8 shows, the flow-matching inversion has small round-trip error, averaging a rel-L2 of 9.3×10^{-2}. Its reconstruction MSE remains 1.6% to 3.8% of MSE_{pred}, so the reverse map is accurate enough for the correlation analysis.

Table 8: Round-trip reconstruction error. Each true future of the test set is mapped back to a seed and pushed forward again, at a lookback and a horizon of 96; the table gives the mean squared error and the mean rel-L2 error over all samples, in normalized space. *% of* MSE_{pred} compares that error to the error the same model makes when it forecasts these windows.

Dataset	MSE	rel-L2	% of MSE_{pred}
ETTh1	6.4×10^{-3}	8.3×10^{-2}	1.57
ETTh2	1.1×10^{-2}	1.1×10^{-1}	2.33
ETTm1	7.9×10^{-3}	9.5×10^{-2}	2.53
ETTm2	4.5×10^{-3}	1.0×10^{-1}	2.57
Weather	7.1×10^{-3}	1.2×10^{-1}	3.82
ECL	6.7×10^{-3}	7.4×10^{-2}	3.79
Traffic	2.1×10^{-2}	9.1×10^{-2}	3.19
Solar	5.4×10^{-3}	7.1×10^{-2}	2.72
mean	8.7×10^{-3}	9.3×10^{-2}	2.82

B.2.3 SIMILARITY TO THE TRUE DEPENDENCE

To examine whether the dependence recovered in source-noise space matches the true data dependence, we compare the inverted-noise correlation $\text{Corr}(z^\star)$ with the observed-future correlation $\text{Corr}(y)$ over the same channel pairs. $\text{Corr}(y)$ is the cross-channel dependence the forecast should reproduce, while $\text{Corr}(z^\star)$ is obtained by inverting the observed future through the flow-matching sampler.

Figure 5 shows these matrices for each dataset, together with the correlation of the originally sampled noise, $\text{Corr}(z)$, and the correlation installed by GENCOUP-CORR at $\alpha = 1$, $\text{Corr}(z')$. The key comparison is therefore between $\text{Corr}(z^\star)$ and $\text{Corr}(y)$, since using the data correlation as R is justified only if the recovered source noise exhibits the same dependence.

We evaluate similarity in terms of structure and numerical accuracy. Structure is the primary criterion, as R should capture the same dependence pattern across channel pairs. Numerical accuracy provides a stricter check, requiring the recovered correlations to also match their actual values.

Off-diagonal representation. Each off-diagonal correlation represents one channel pair, with sign indicating direction and magnitude indicating dependence strength. We compare $\text{Corr}(y)$ and $\text{Corr}(z^\star)$ using the $C(C-1)/2$ unique upper-triangular off-diagonal entries in a fixed order. The diagonal is omitted because it is always one, and each pair is included once.

Setup. All experiments use a lookback and forecast horizon of 96. Test windows are taken with a stride of 96, so their forecast targets do not overlap, and each future observation appears exactly once. Each window is represented as a matrix in $\mathbb{R}^{H\times C}$, with $H = 96$ forecast steps and C channels.

Structure. We compare the off-diagonal entries using *cosine similarity*. Our goal is to assess whether the two matrices capture the same cross-channel dependence structure, meaning that corre-

Table 9: **Mantel permutation test.** T_{obs} is the observed cosine of $\text{Corr}(z^\star)$ to $\text{Corr}(y)$ over the upper-triangular off-diagonal entries, and *random* R is a reference null defined in the text. The permutation columns summarise the null of random channel relabelling. ✓: null hypothesis rejected, so the recovered structure is not explained by random channel relabelling.

				permutation			
Dataset	C	T_{obs}	random R	null mean	null p_{95}	p	H_0
ETTh1	7	0.816	-0.003	0.136	0.461	0.0002	✓
ETTh2	7	0.845	-0.004	0.326	0.638	0.0004	✓
ETTm1	7	0.843	-0.002	0.130	0.440	0.0003	✓
ETTm2	7	0.903	0.002	0.397	0.699	0.0003	✓
Weather	21	0.780	-0.001	0.064	0.212	0.0001	✓
ECL	321	0.777	0.000	0.519	0.534	0.0001	✓
Traffic	862	0.908	0.000	0.838	0.841	0.0001	✓
Solar	137	0.970	0.000	0.962	0.962	0.0001	✓
Mean		0.855	-0.001				

lations have similar signs and relative strengths across channel pairs. Cosine similarity is appropriate for this purpose because it compares the overall pattern of pairwise correlations while remaining invariant to a common rescaling of their magnitude. Table 9 reports the observed cosine between the recovered $\text{Corr}(z^\star)$ and the true $\text{Corr}(y)$, beside a reference null: the cosine of a random correlation matrix to that same $\text{Corr}(y)$ (off-diagonal entries drawn uniformly on $[-1, 1]$, averaged over 10^4 draws).

The correlations are estimated at the dataset level, matching the construction used by GENCOUPCORR: all test windows are pooled along the temporal dimension and a single cross-channel correlation matrix is computed from the resulting observations, so the matrices compared correspond directly to the objects the method uses. Because the test windows have no overlap, pooling them reconstructs the test sequence, so $\text{Corr}(y)$ is the correlation matrix of the test data itself and is not affected by between-window noise.

The random-R column in Table 9 shows that random correlation matrices have nearly zero cosine similarity to the true correlation matrix on average, indicating that the high similarity of the recovered noise is nontrivial. We further use a Mantel-style permutation test (Mantel, 1967) to test whether the *structure* of the recovered correlation matrix matches that of the true data correlation beyond chance.

Our target statistic T_{obs} is the cosine similarity between the off-diagonal entries of $\text{Corr}(y)$ and $\text{Corr}(z^\star)$. Under the null hypothesis, the recovered correlation pattern is not associated with the correct channel pairs. We generate $B = 10^4$ permutations by jointly permuting the rows and columns of $\text{Corr}(z^\star)$. This preserves its correlation values and matrix properties while randomly reassigning them to channel pairs. For each permutation, we recompute the cosine similarity T_b with $\text{Corr}(y)$ and report

$$p = \frac{1 + \#\{T_b \geq T_{\text{obs}}\}}{B + 1},$$

using the finite-sample correction of Phipson & Smyth (2010). We reject the null at $p < 0.05$; rejection provides evidence that the recovered correlation preserves the true cross-channel structure beyond random channel correspondence.

The classical Mantel test uses Pearson correlation between corresponding pairwise entries; here, we use the same permutation procedure with cosine similarity instead.

Result. Table 9 rejects the null on every dataset, and the observed cosine similarities consistently exceed their permutation baselines. Figure 4 shows that even when the null similarity is high for high-dimensional datasets, its distribution is narrow. Thus, the recovered dependence aligns with the correct channel correspondence beyond random permutation.

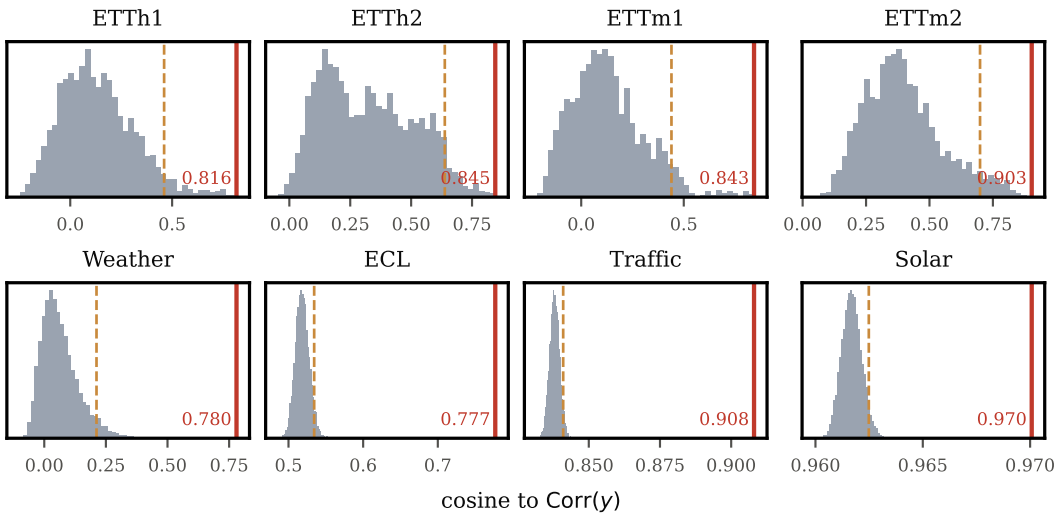


Figure 4: The permutation null of the Mantel test for the flow, one panel per dataset, over 10^4 random channel relabellings of $\mathrm{Corr}(z^\star)$. The solid red line is the observed cosine to $\mathrm{Corr}(y)$, printed beside it, and the dashed amber line the null's 95th percentile, with the null distribution itself in grey. The observed value lies beyond the null on every dataset.

Table 10: **Reconstruction error per window.** MSE between the recovered and the true correlation matrix of the same window, as mean±sd over the windows, beside two baselines scored against the same $\mathrm{Corr}(y)$: a random correlation matrix and independent channels. Every entry is computed on the $C(C-1)/2$ upper-triangular off-diagonal entries. **Bold** is the smallest and underline the second.

		baselines		$\mathrm{MSE}_{\mathrm{rec}}$
Dataset	C	random R	I	flow
ETTh1	7	0.5149 ± 0.0313	0.1806 ± 0.0309	$\mathbf{0.0615 \pm 0.0353}$
ETTh2	7	0.5146 ± 0.0698	0.1886 ± 0.0707	$\mathbf{0.0565 \pm 0.0355}$
ETTm1	7	0.5799 ± 0.0750	0.2404 ± 0.0748	$\mathbf{0.1351 \pm 0.0575}$
ETTm2	7	0.5023 ± 0.0569	0.1696 ± 0.0617	$\mathbf{0.0847 \pm 0.0433}$
Weather	21	0.6737 ± 0.0720	0.3383 ± 0.0720	$\mathbf{0.1793 \pm 0.0730}$
ECL	321	0.7430 ± 0.0249	0.4096 ± 0.0249	$\mathbf{0.3712 \pm 0.0489}$
Traffic	862	0.7709 ± 0.0493	0.4375 ± 0.0493	$\mathbf{0.2773 \pm 0.0473}$
Solar	137	1.0858 ± 0.2044	0.7533 ± 0.2045	$\mathbf{0.3578 \pm 0.1768}$

Numerical accuracy. We assess numerical accuracy using the mean squared error (MSE) between corresponding off-diagonal correlation entries. Unlike cosine similarity, which measures structural alignment, MSE directly penalizes differences in the recovered correlation values.

For each window, we compute its correlation matrix and measure the MSE between the recovered and observed matrices. Table 10 reports the mean±sd across windows, together with two baselines evaluated against the same $\mathrm{Corr}(y)$: a random correlation matrix and an independence baseline with zero off-diagonal entries. The independence baseline is the main reference, as it represents recovered noise carrying no cross-channel dependence.

Result. Table 10 shows that the recovered correlations achieve lower MSE overall than both the independence and random-correlation baselines. This indicates that the reconstructed correlation matrices are non-trivial: they recover meaningful cross-channel dependence and provide a closer numerical approximation to the observed correlations than either random dependence or assuming no dependence at all.

B.2.4 CONCLUSION

The recovered noise consistently preserves the cross-channel dependence structure, with similarities significantly above the permutation null. Its correlation matrices also achieve lower numerical error overall than both random-correlation and independence baselines, showing that the recovered dependence is non-trivial and meaningfully approximates that observed in the data.

Together, these results provide empirical support for using the data correlation as R, the cross-channel correlation of the source noise. They further indicate that cross-channel dependence is carried through the generative mapping between source-noise and data space, supporting source-noise coupling as a mechanism for controlling the dependence structure of generated forecast trajectories.

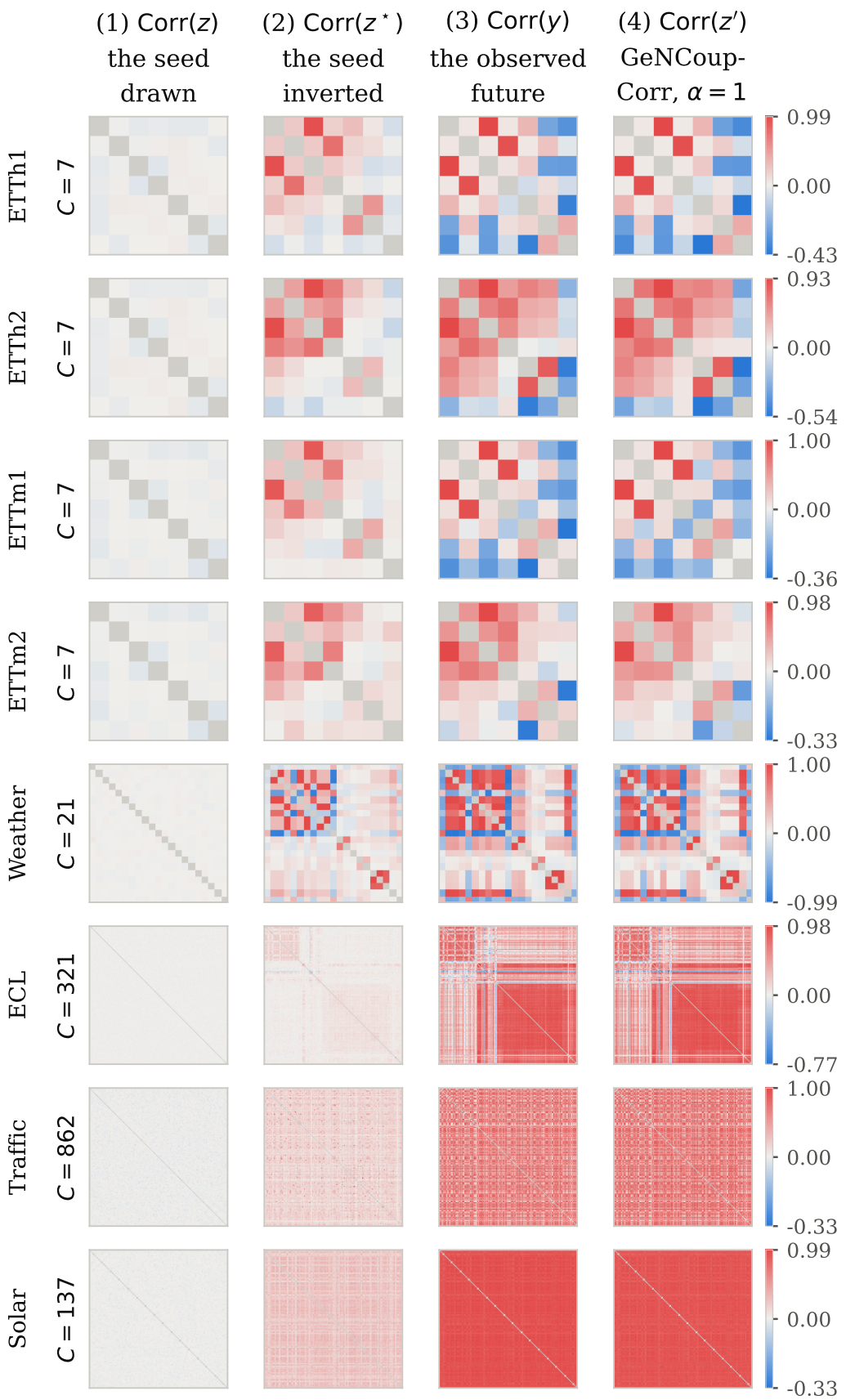


Figure 5: The seed the observed futures require, per dataset at $H = 96$. Rows are datasets; columns, left to right: the seed the sampler draws; the seed recovered by inverting FMPatch; the normalized observed future; and the drawn seed after the coupling of GeNCoup-Corr at $\alpha = 1$. One colour scale per row, native channel order; the diagonal is masked, being 1 in every matrix.

Table 11: **What each blend installs, against the target.** Cosine similarity between the off-diagonal entries of the correlation GENCOUP-CORR installs at each α and those of $\mathrm{Corr}(y)$, the correlation of the observed futures. Every draw uses the same windows and the same seed, so the blends are directly comparable with one another and with the target.

Dataset	$\alpha = 0$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.8$	$\alpha = 1$
ETTh1	0.645	0.810	0.906	0.989	0.998
ETTh2	0.497	0.742	0.873	0.977	0.995
ETTm1	0.627	0.781	0.875	0.968	0.987
ETTm2	0.439	0.645	0.792	0.951	0.980
Weather	0.872	0.939	0.969	0.993	0.998
ECL	0.976	0.985	0.989	0.990	0.988
Traffic	0.984	0.991	0.994	0.997	0.997
Solar	1.000	1.000	1.000	1.000	1.000
mean	0.755	0.862	0.925	0.983	0.993

B.3 ANALYSIS OF α

GENCOUP-CORR fuses two estimates of correlation matrix: R_{global}, fitted once on the training split, and R_{local}, estimated at inference from the current lookback window. The parameter α of equation 8 sets the fusion between them (§4.2).

The local term accounts for possible non-stationarity in cross-channel dependence: correlations and covariance structure can vary over time rather than remain fixed throughout a series (Engle, 2002; Chen et al., 2013). Consequently, the correlation in a test window may differ from that estimated over the training split. If dependence were stationary, the global estimate would be preferable, as it is fitted from substantially more data, whereas increasing α places more weight on an estimate from a single short lookback.

This motivates asking whether increasing α installs a correlation closer to that of the observed future, and whether the relative performance of the local and global estimates provides evidence of non-stationarity in the dataset's cross-channel dependence. We compare the correlation matrices produced by $\alpha \in \{0, 0.3, 0.5, 0.8, 1\}$ against the observed-future correlation. Figure 6 visualizes the matrices, and Table 11 reports their similarity to the target. The corresponding forecasting ablation of the fusion mechanism, comparing $\alpha = 0$, $\alpha = 1$, and the validation-selected $\alpha^\star$, is reported in the main analysis in Table 3.

Results. Table 11 shows that increasing α makes the produced correlation consistently closer to the observed-future correlation. Averaged over the eight datasets, the cosine similarity rises from 0.755 at $\alpha = 0$ to nearly perfect similarity, 0.993, at $\alpha = 1$. In addition, the wider datasets show relatively high similarity at $\alpha = 0$, suggesting more stationary cross-channel dependence between the train and test splits. The largest gains from increasing α appear on datasets whose correlations drift more.

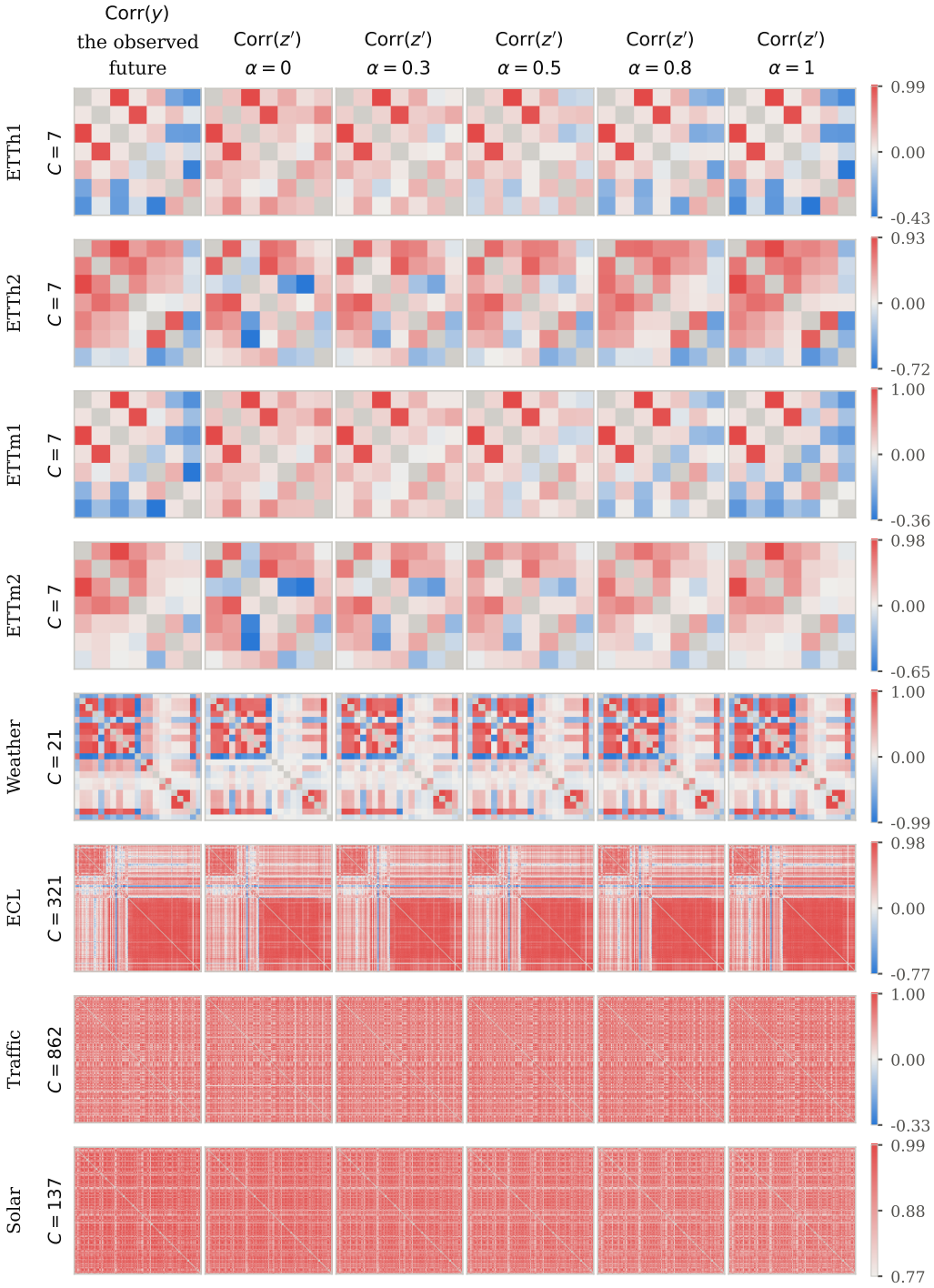


Figure 6: The correlation GENCOUP-CORR installs at each blend, beside the target it is meant to match. Rows are datasets; columns, left to right: the observed future, then the coupled seed at $\alpha = 0$, 0.3, 0.5, 0.8 and 1. One colour scale per row, native channel order; the diagonal is masked, being 1 in every matrix.

B.4 Analysis of the injected noise's effect on the target variogram score

The variogram score (equation 1) is a sum over cross-channel pairs, which lets us decompose it by pair to analyze how the coupling affects the target trajectories. We compare each pair's contribution under the uncoupled and coupled forecasts, so the difference indicates whether that pair improved or degraded and by how much. We apply this analysis to both variants. For GENCOUP-CORR, we set $\alpha = 0$, injecting the same matrix across all test windows so the observed pairwise changes can be attributed to that fixed coupling. Throughout this analysis, we use k2VAE as the forecasting backbone (Wu et al., 2025). Figure 2 shows the Weather decomposition in the main paper; Figures 7–13 show the remaining datasets, and Table 12 reports the summary per dataset.

Constructing the pair-loss plots. The per-pair loss plots are produced from the pairwise terms of the variogram score. Let ℓ^0_{jk} and ℓ^1_{jk} be the channel-pair loss for pair $j < k$ for the uncoupled and coupled forecasts. With $\mathrm{VS}_0 = \sum_{j<k} \ell^0_{jk}$, every plotted loss is normalized by the same uncoupled total:

$$a_{jk} = 100\,\frac{\ell^0_{jk}}{\mathrm{VS}_0}, \qquad b_{jk} = 100\,\frac{\ell^1_{jk}}{\mathrm{VS}_0}, \qquad c_{jk} = b_{jk} - a_{jk} = 100\,\frac{\ell^1_{jk} - \ell^0_{jk}}{\mathrm{VS}_0}. \tag{12}$$

Thus panels **(a)** and **(b)** show uncoupled and coupled pair losses on the same scale, while panel **(c)** shows their cellwise difference. Panel **(a)** sums to 1. Panel **(c)** gives the pairwise change, with total $\sum_{j<k} c_{jk}$. Panel **(b)** then sums to $1 - \sum_{j<k} c_{jk}$. In panel **(c)**, negative values mark pairs with improved loss, while positive values mark pairs with worsened loss. Panel **(f)** plots the same c_{jk} values against the injected correlation R_{jk} to show how the injected correlation contributes to the change in pair loss.

GENCOUP-CORR vs. GENCOUP-DYN comparison. Table 12 shows that GENCOUP-DYN improves both the pair-improvement percentage and the aggregate total loss. Across the eight datasets, it improves 78% of pairs on average, compared with 66% for GENCOUP-CORR, and reduces the total pairwise loss by 17.0% compared with 12.4%. It also achieves the better total on seven of eight datasets, indicating that the learned matrix improves a broader set of channel pairs.

Table 12: Where the two couplings move the variogram score. The VS of equation 1 decomposed over channel pairs, k2VAE on every dataset. Negative is better. gain and harm are the sums over the improved and the degraded pairs ($\Delta\mathrm{VS} = \mathrm{gain} + \mathrm{harm}$), each a percentage of the uncoupled total; helped % is the share of pairs improved, with the larger of the two arms in bold. The last row averages the datasets shown.

		GeNCoup-Corr ($\alpha = 0$)				GeNCoup-Dyn			
Dataset	C	ΔVS	gain	harm	helped %	ΔVS	gain	harm	helped %
ETTh1	7	−22.5	−22.7	+0.2	81	−29.5	−29.7	+0.2	**86**
ETTh2	7	+ 0.5	− 0.2	+0.6	**57**	+ 0.4	− 1.0	+1.4	48
ETTm1	7	−18.2	−18.3	+0.1	81	−27.2	−27.5	+0.2	**86**
ETTm2	7	+ 0.3	− 1.4	+1.7	48	− 6.1	− 6.8	+0.6	**76**
Weather	21	−15.2	−16.5	+1.3	58	−27.1	−27.2	+0.1	**95**
ECL	321	+ 1.9	− 1.1	+3.0	38	− 0.7	− 1.2	+0.5	**60**
Traffic	862	− 2.9	− 3.9	+1.0	68	− 2.7	− 3.1	+0.3	**77**
Solar	137	−43.2	−43.2	+0.0	100	−43.3	−43.3	+0.0	100
mean	–	−12.4	−13.4	+1.0	66	−17.0	−17.5	+0.4	**78**

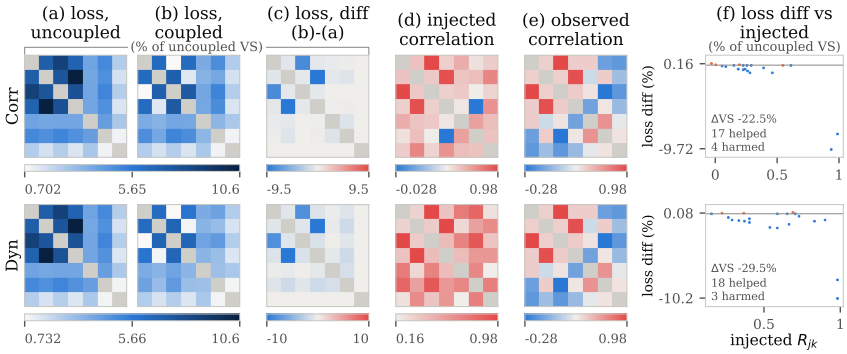


Figure 7: ETTh1, k2VAE ($C = 7$, 21 pairs). GeNCoup-Corr at $\alpha = 0$ in the top row, GeNCoup-Dyn in the second. **(a)–(c)** Each cross-channel pair's contribution, as a percentage of the uncoupled VS, in the uncoupled and coupled forecasts, and their difference. **(d)** the correlation the coupling injected and **(e)** the one the observed futures carry; **(f)** Each pair's percentage change versus the injected correlation, with improvements in blue and degradations in orange.

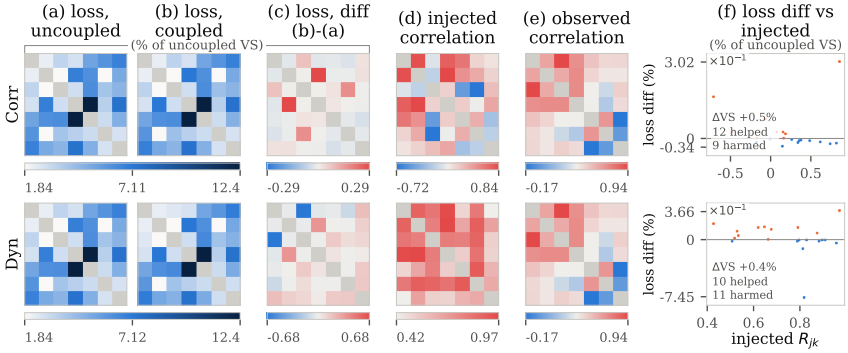


Figure 8: ETTh2, k2VAE ($C = 7$, 21 pairs). GeNCoup-Corr at $\alpha = 0$ in the top row, GeNCoup-Dyn in the second. **(a)–(c)** Each cross-channel pair's contribution, as a percentage of the uncoupled VS, in the uncoupled and coupled forecasts, and their difference. **(d)** the correlation the coupling injected and **(e)** the one the observed futures carry; **(f)** Each pair's percentage change versus the injected correlation, with improvements in blue and degradations in orange.

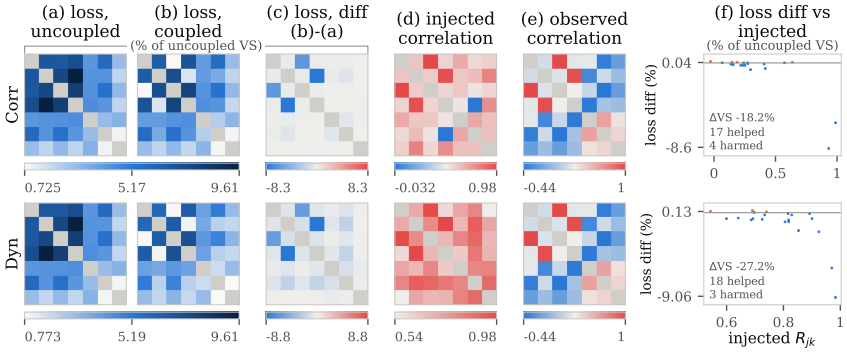


Figure 9: ETTm1, k2VAE ($C = 7$, 21 pairs). GeNCoup-Corr at $\alpha = 0$ in the top row, GeNCoup-Dyn in the second. **(a)–(c)** Each cross-channel pair's contribution, as a percentage of the uncoupled VS, in the uncoupled and coupled forecasts, and their difference. **(d)** the correlation the coupling injected and **(e)** the one the observed futures carry; **(f)** Each pair's percentage change versus the injected correlation, with improvements in blue and degradations in orange.

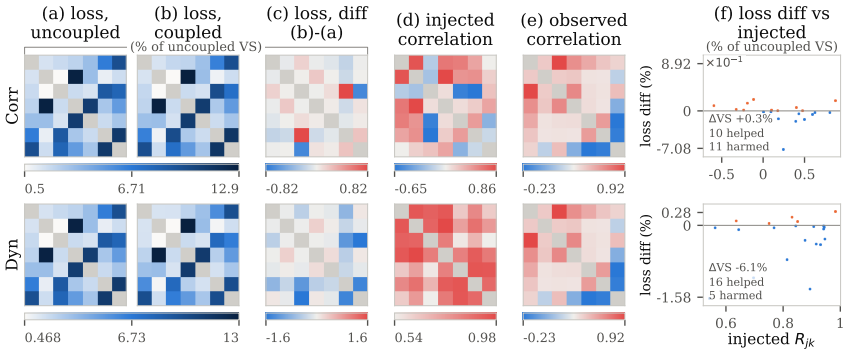


Figure 10: ETTm2, k2VAE ($C = 7$, 21 pairs). GeNCoup-Corr at $\alpha = 0$ in the top row, GeNCoup-Dyn in the second. **(a)–(c)** Each cross-channel pair's contribution, as a percentage of the uncoupled VS, in the uncoupled and coupled forecasts, and their difference. **(d)** the correlation the coupling injected and **(e)** the one the observed futures carry; **(f)** Each pair's percentage change versus the injected correlation, with improvements in blue and degradations in orange.

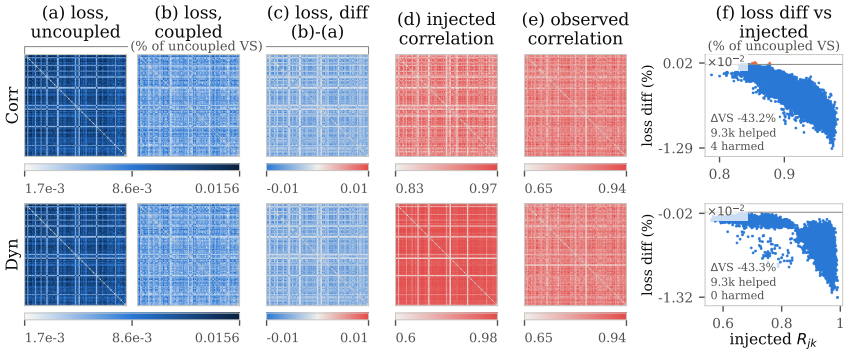


Figure 11: Solar, k2VAE ($C = 137$, 9,316 pairs). GeNCoup-Corr at $\alpha = 0$ in the top row, GeNCoup-Dyn in the second. **(a)–(c)** Each cross-channel pair's contribution, as a percentage of the uncoupled VS, in the uncoupled and coupled forecasts, and their difference. **(d)** the correlation the coupling injected and **(e)** the one the observed futures carry; **(f)** Each pair's percentage change versus the injected correlation, with improvements in blue and degradations in orange.

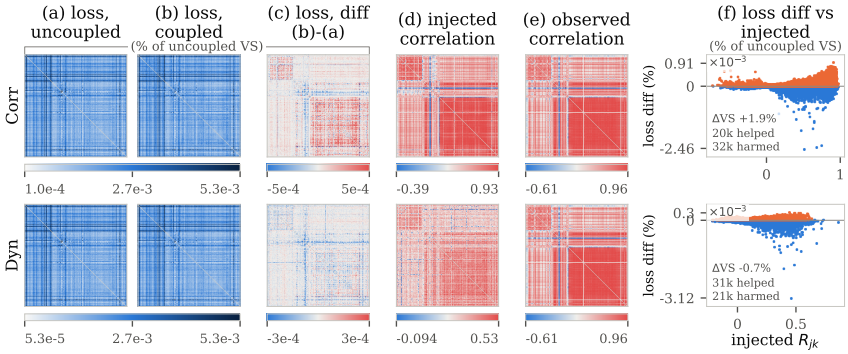


Figure 12: ECL, k2VAE ($C = 321$, 51,360 pairs). GeNCoup-Corr at $\alpha = 0$ in the top row, GeNCoup-Dyn in the second. **(a)–(c)** Each cross-channel pair's contribution, as a percentage of the uncoupled VS, in the uncoupled and coupled forecasts, and their difference. **(d)** the correlation the coupling injected and **(e)** the one the observed futures carry; **(f)** Each pair's percentage change versus the injected correlation, with improvements in blue and degradations in orange.

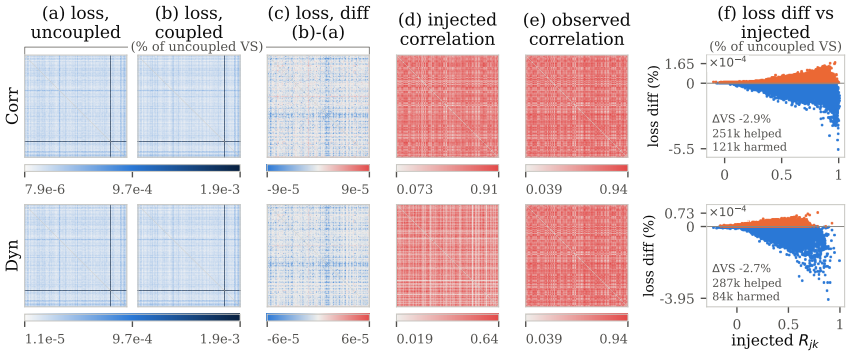


Figure 13: Traffic, k2VAE ($C = 862$, 371,091 pairs). GeNCoup-Corr at $\alpha = 0$ in the top row, GeNCoup-Dyn in the second. **(a)–(c)** Each cross-channel pair's contribution, as a percentage of the uncoupled VS, in the uncoupled and coupled forecasts, and their difference. **(d)** the correlation the coupling injected and **(e)** the one the observed futures carry; **(f)** Each pair's percentage change versus the injected correlation, with improvements in blue and degradations in orange.

B.5 What the target coupling looks like

In this section we present a comparison of coupled and uncoupled trajectories over multiple datasets and forecasters, at correlation targets stepping through $\rho \approx -1$, -0.5, 0, $+0.5$ and $+1$, similarly to Figure 1, which showed a pair whose correlation is close to $+1$.

In the presented trajectories we can see that at targets near $+1$ and -1 the coupled draws follow the observed dependence while the uncoupled ones do not, and that at targets near 0, where there is no dependence to recover, the coupling brings no gain. The coupling therefore acts on the channel pairs that are genuinely related.

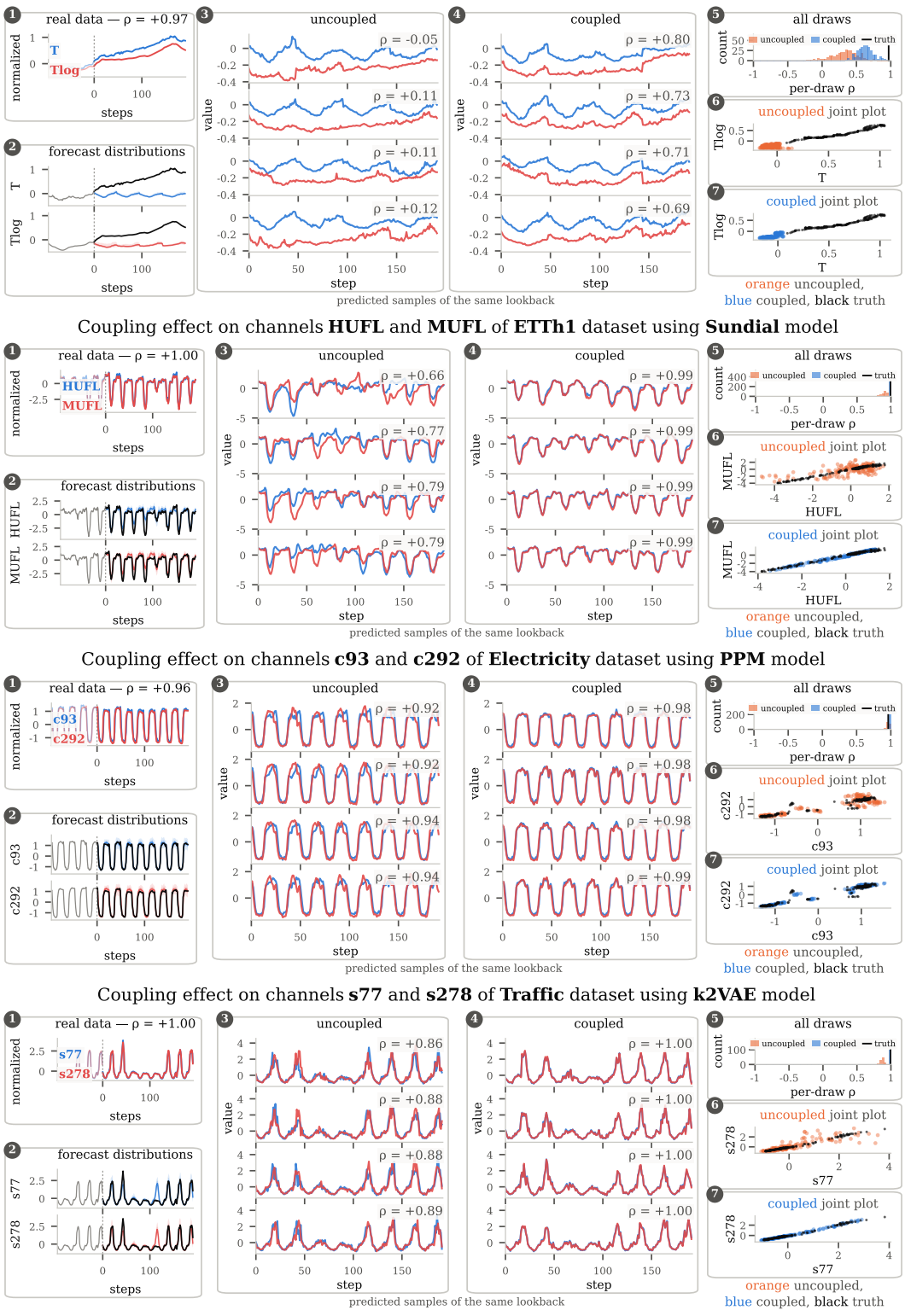


Figure 14: A target near +1. Top to bottom: Weather T/TLOG on Aurora, ETTh1 HUFL/MUFL on Sundial, two ECL clients on PPM, and two Traffic sensors on k2VAE. On the last two the uncoupled draws are already close to the observed dependence, so there is little for the coupling to add.

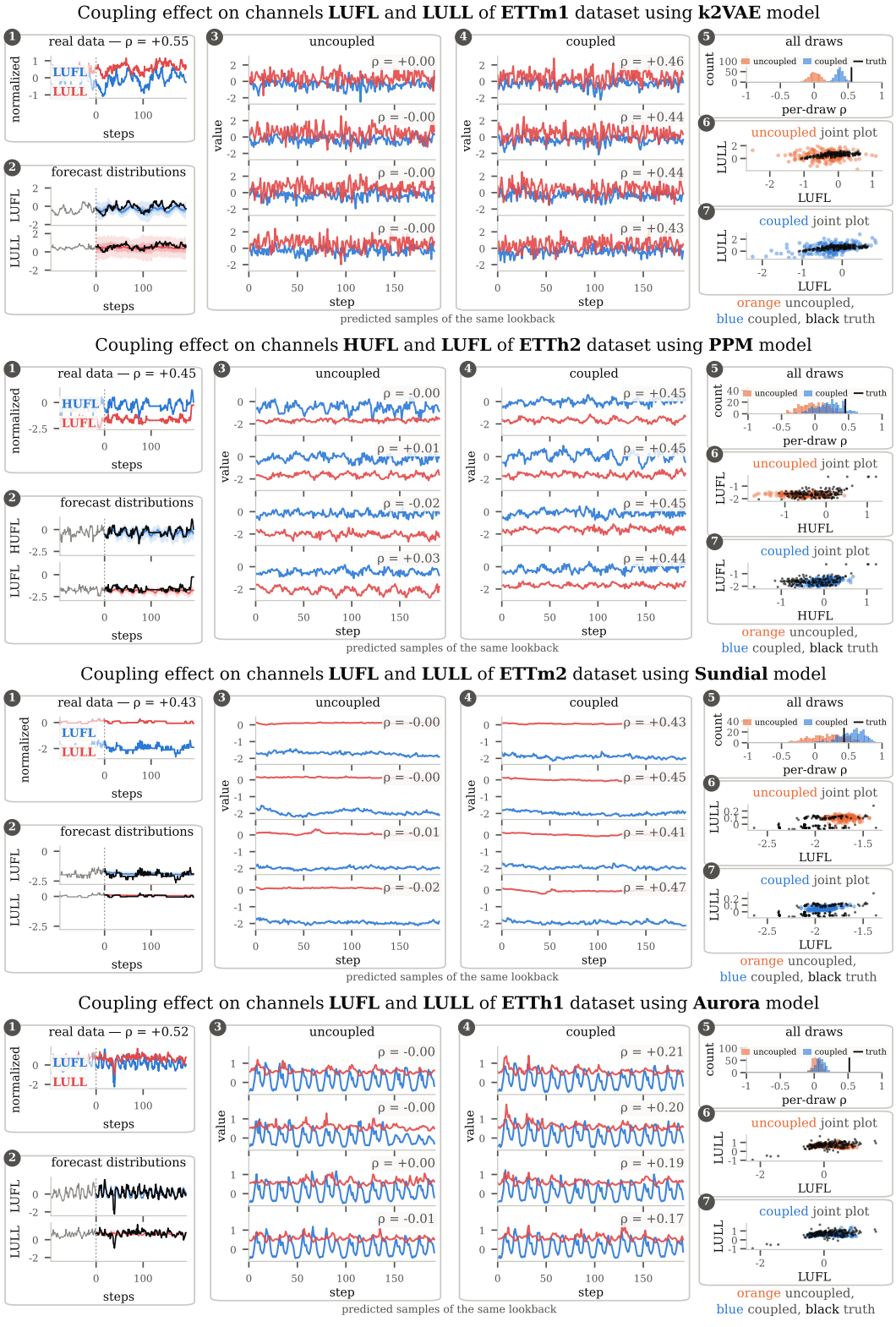


Figure 15: A target near $+0.5$**.** Top to bottom: ETTm1 LUFL/LULL on k2VAE, ETTh2 HUFL/LUFL on PPM, ETTm2 LUFL/LULL on Sundial, and ETTh1 LUFL/LULL on Aurora.

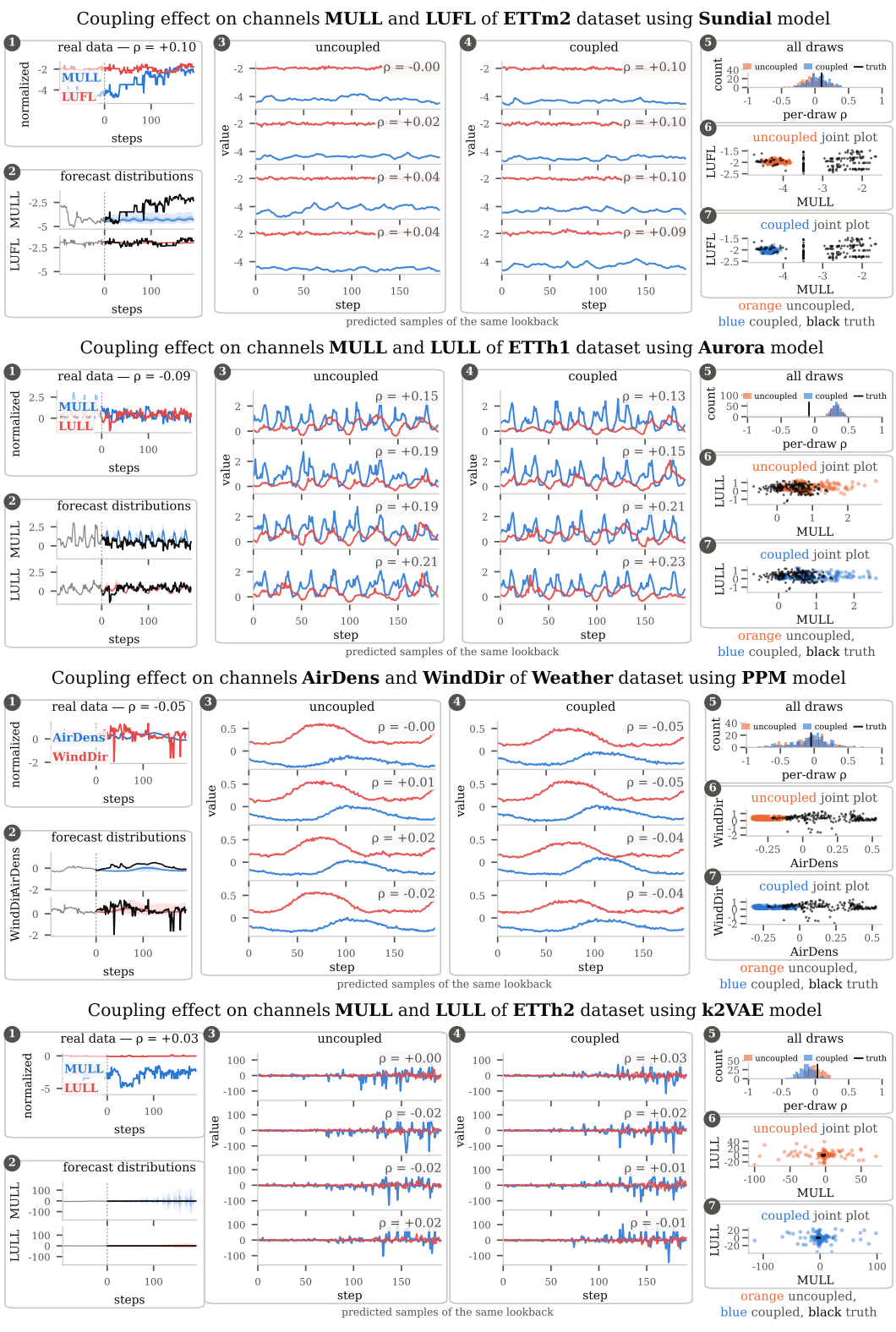


Figure 16: A target near zero, $H = 192$. Top to bottom: ETTm2 MULL/LUFL on Sundial, ETTh1 MULL/LULL on Aurora, Weather air density against wind direction on PPM, and ETTh2 MULL/LULL on k2VAE. With almost no dependence in the data there is nothing for the coupling to install, and the two arms are hard to tell apart: in (5) the per-draw correlation of the coupled ensemble sits where the uncoupled one already sat.

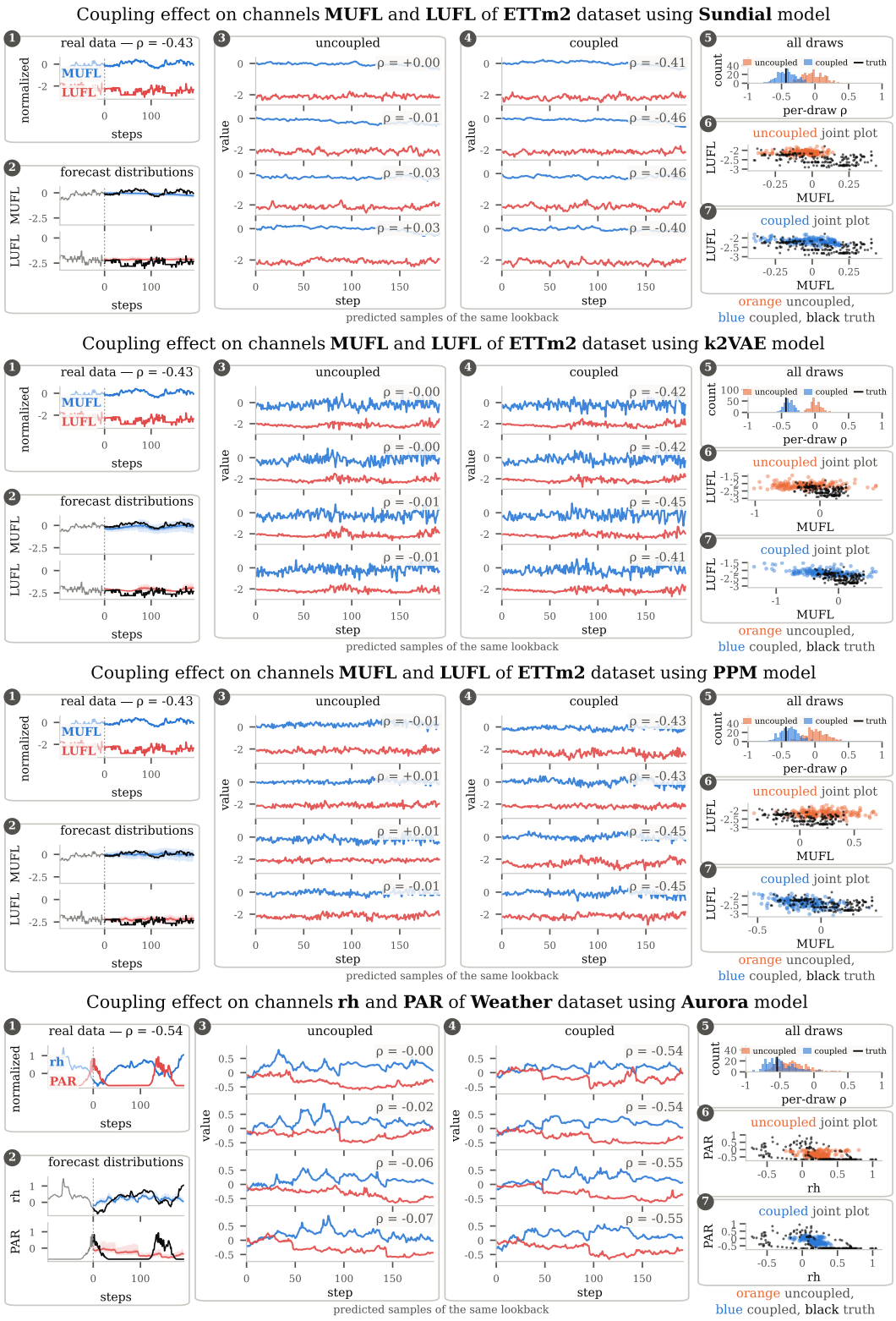


Figure 17: A target near -0.5. The first three rows are ETTm2 MUFL/LUFL in one window on Sundial, k2VAE and PPM, so they differ by the forecaster alone; the fourth is Weather relative humidity against sunlight on Aurora.

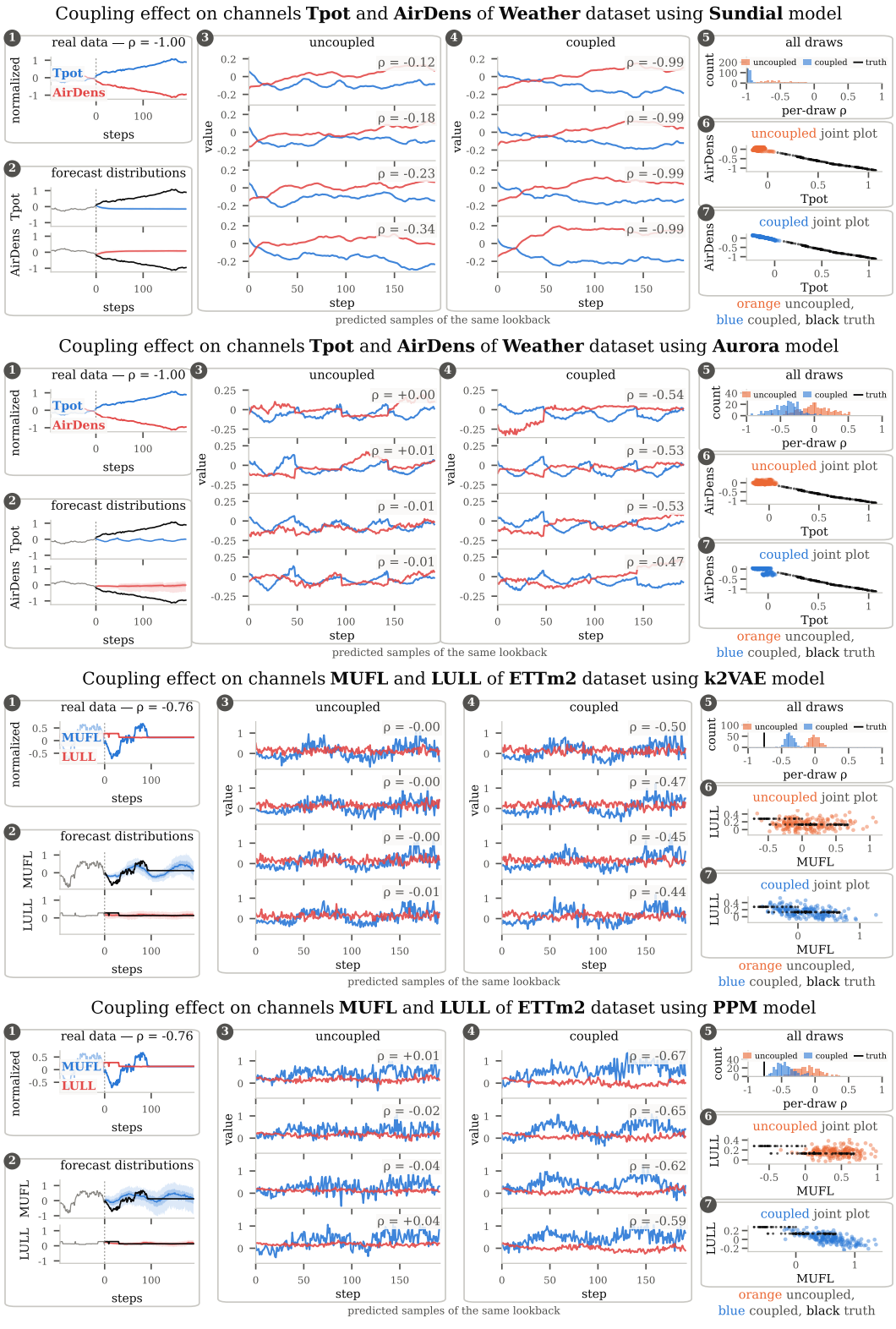


Figure 18: A target near −1. Top to bottom: Weather potential temperature against air density, which are physically anti-correlated, on Sundial and on Aurora; then ETTm2 MUFL/LULL on k2VAE and on PPM.

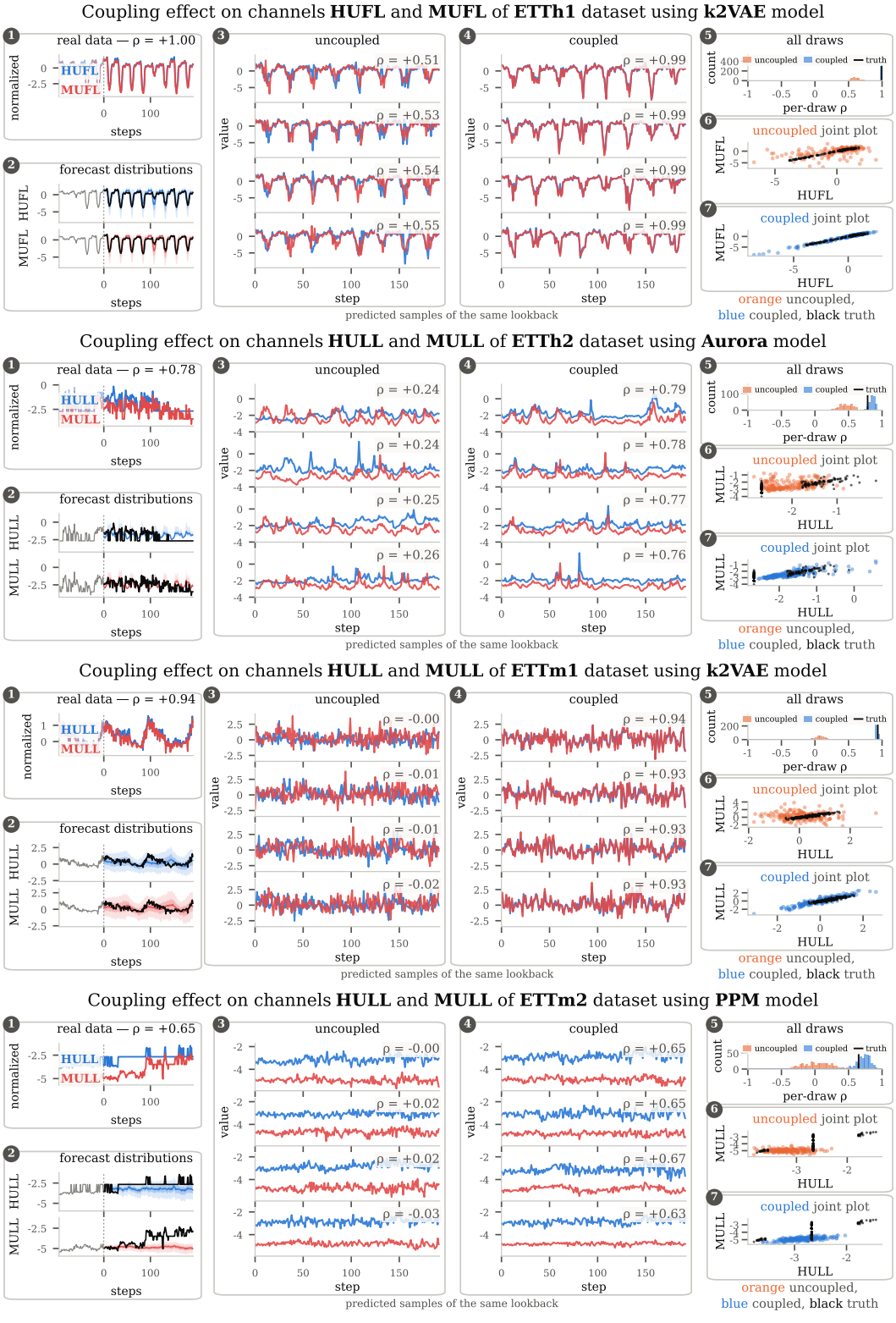


Figure 19: A target near $+1$, the ETT benchmarks. Top to bottom: ETTh1 on k2VAE, ETTh2 on Aurora, ETTm1 on k2VAE, and ETTm2 on PPM.

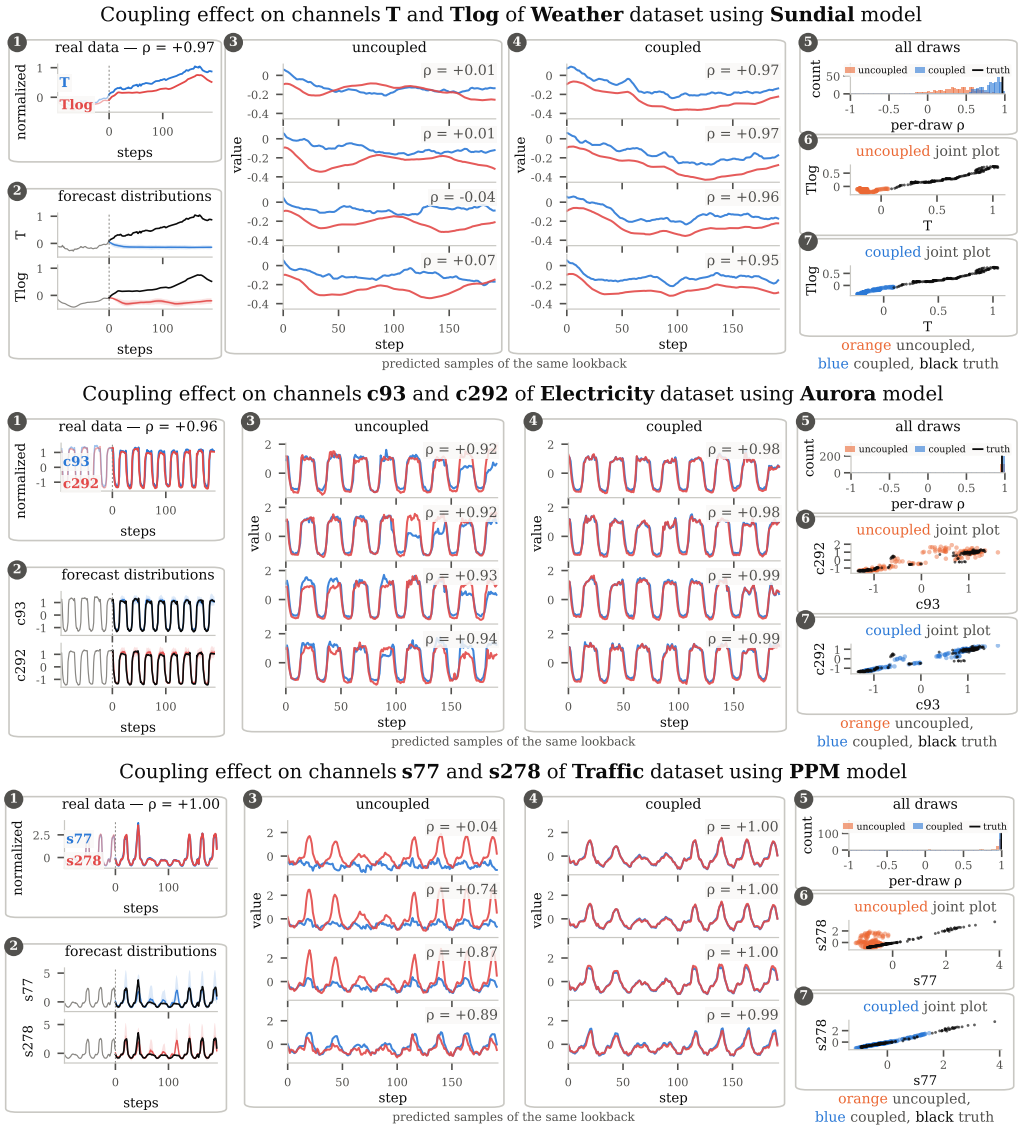


Figure 20: A target near +1, the remaining benchmarks. Top to bottom: Weather on Sundial, ECL on Aurora, and Traffic on PPM. Solar is drawn on its own in Figure 22.

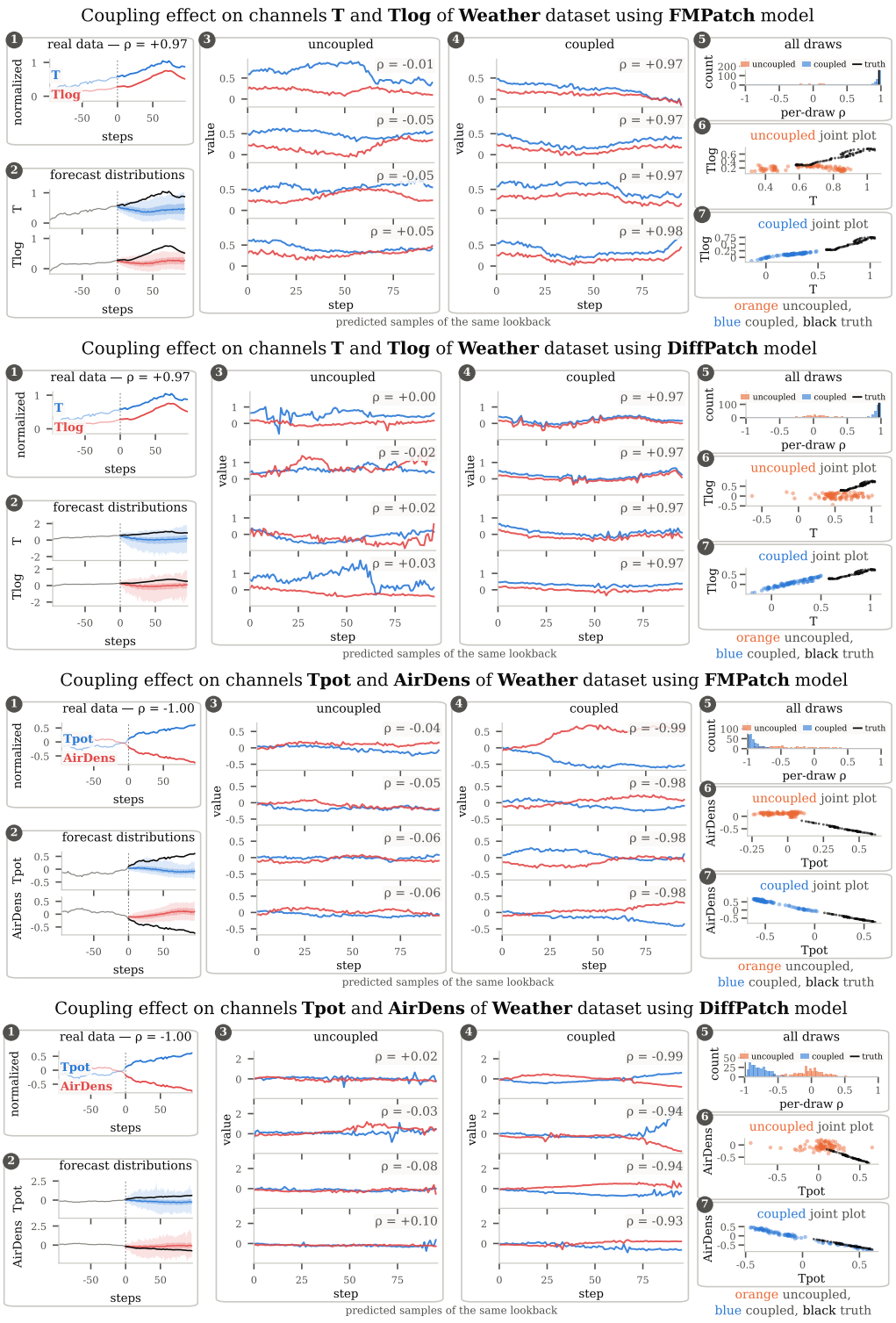


Figure 21: The two patch backbones, at $H = 96$. Weather pairs near targets $+1$ and -1, shown on FMPatch and DiffPatch.

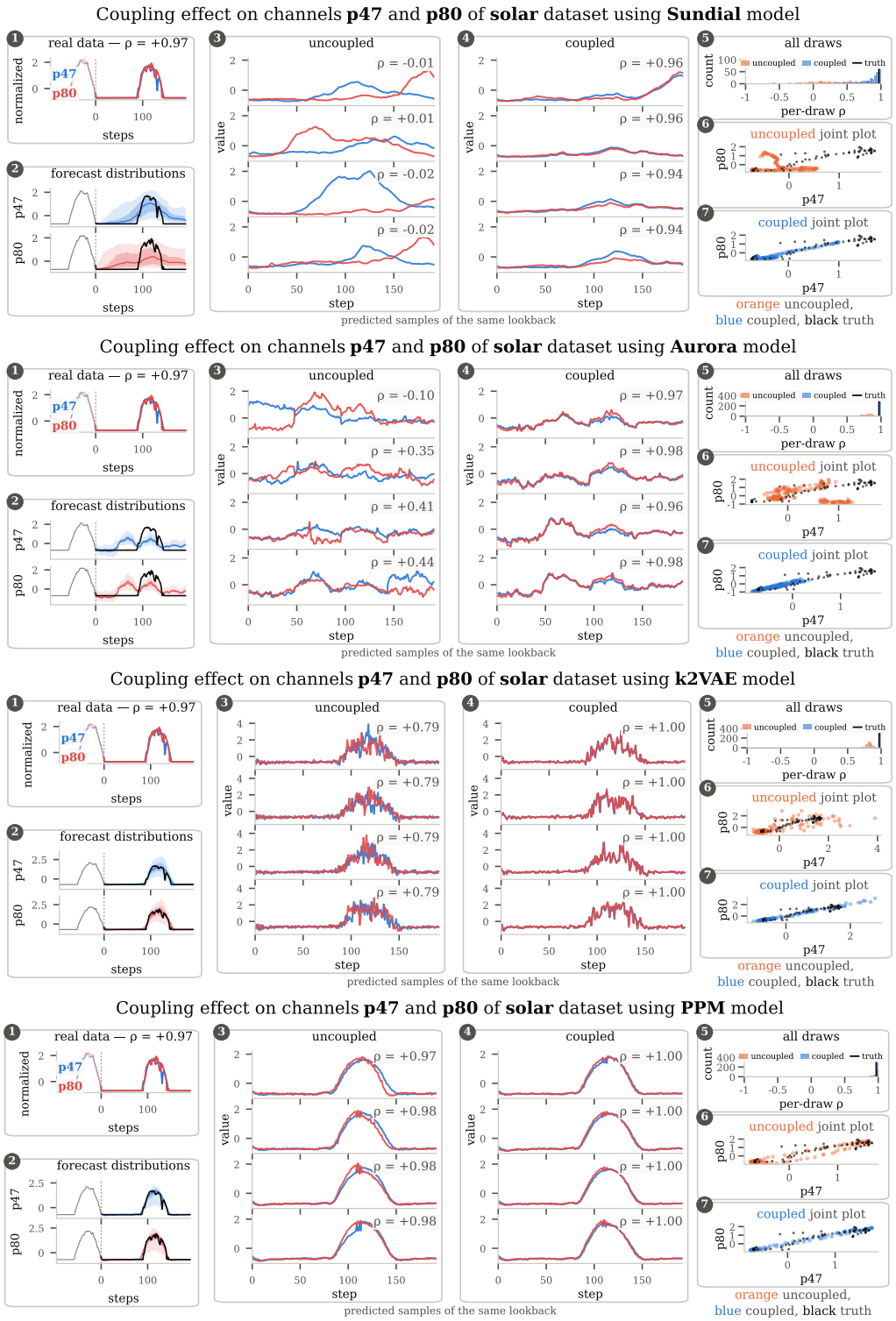


Figure 22: Solar, at a target near $+1$. One plant pair in one window, shown across Sundial, Aurora, k2VAE and PPM. Solar has only positively correlated pairs, so no near-zero or negative target is shown.

C AI USE STATEMENT

In this work, we used generative AI tools to implement some baseline methods, including translating existing implementations from R to Python. We have not used generative AI tools to generate synthetic data sets, help develop theoretical models or conceptual frameworks, formulate mathematical claims, provide critical ingredients for proving mathematical claims, assist in the writing of proofs, propose or refine hypotheses, design or provide feedback on research methodology or experiments, clean and reformat datasets, support qualitative and thematic data analysis, or interpret results, and assistance with translation is not applicable to this work.

Additionally, we used generative AI tools to aid and polish the writing of the manuscript. We have reviewed all AI-assisted work. The AI-assisted baseline implementations were manually reviewed, verified against the original methods. We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

D Detailed Experimental Results

D.1 Marginal and temporal scores per dataset

Table 13 breaks the CRPS, NMAE and ES-Time rows of Table 2 down by dataset, averaged over the six forecasters and the four horizons.

Table 13: **Marginal and temporal scores per dataset, averaged over forecasters and horizons.** Percent change vs. the uncoupled arm of the same settings; negative is better. ± is the standard deviation over 3 seeds (2026–2028); all settings of a cell are first averaged per seed. *All datasets* averages all cells.

Dataset	Metric (%)	GeNCoup-Dyn (ours)	GeNCoup-Corr (ours)	COBASE	Schaake	SimSch.	GCA	IC
ETTh1	CRPS	− 0.038±0.091	− 0.015±0.036	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.712±0.001	+ 0.000±0.000
	NMAE	− 0.035±0.102	− 0.018±0.032	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.938±0.013	+ 0.000±0.000
	ES-Time	− 0.026±0.040	+ 0.025±0.027	+ 1.320±0.020	+ 0.216±0.007	+ 0.474±0.015	+ 1.430±0.014	+ 1.305±0.014
ETTh2	CRPS	+ 0.034±0.118	− 0.004±0.095	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.655±0.003	+ 0.000±0.000
	NMAE	+ 0.044±0.154	− 0.027±0.090	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 1.372±0.014	+ 0.000±0.000
	ES-Time	+ 0.017±0.080	− 0.047±0.081	+ 4.163±0.054	+ 0.370±0.023	+ 1.230±0.074	+ 3.606±0.039	+ 4.146±0.074
ETTm1	CRPS	− 0.023±0.047	− 0.014±0.046	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.680±0.001	+ 0.000±0.000
	NMAE	− 0.004±0.027	− 0.018±0.027	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.812±0.005	+ 0.000±0.000
	ES-Time	+ 0.019±0.042	+ 0.024±0.045	+ 2.126±0.040	− 0.187±0.008	+ 0.051±0.038	+ 1.989±0.033	+ 2.083±0.036
ETTm2	CRPS	+ 0.062±0.006	+ 0.021±0.027	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.625±0.002	+ 0.000±0.000
	NMAE	+ 0.065±0.029	+ 0.036±0.027	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 1.255±0.010	+ 0.000±0.000
	ES-Time	+ 0.083±0.058	+ 0.073±0.027	+ 4.731±0.049	+ 0.716±0.003	+ 1.131±0.083	+ 4.174±0.036	+ 4.731±0.050
Weather	CRPS	− 0.012±0.031	+ 0.006±0.095	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.647±0.001	+ 0.000±0.000
	NMAE	+ 0.002±0.088	− 0.034±0.194	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 1.220±0.005	+ 0.000±0.000
	ES-Time	+ 0.011±0.075	+ 0.007±0.039	+ 6.783±0.007	+ 0.213±0.006	+ 0.332±0.032	+ 5.185±0.005	+ 6.792±0.010
ECL	CRPS	+ 0.044±0.030	+ 0.044±0.016	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.669±0.001	+ 0.000±0.000
	NMAE	+ 0.041±0.010	+ 0.045±0.014	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.694±0.003	+ 0.000±0.000
	ES-Time	− 0.002±0.013	− 0.015±0.019	+ 1.184±0.004	+ 0.282±0.003	+ 0.427±0.056	+ 1.187±0.004	+ 1.163±0.007
Traffic	CRPS	− 0.001±0.004	+ 0.030±0.044	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.649±0.000	–
	NMAE	− 0.002±0.002	+ 0.029±0.035	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.619±0.001	–
	ES-Time	− 0.003±0.011	+ 0.022±0.021	+ 0.702±0.005	− 0.013±0.003	+ 0.501±0.037	+ 0.160±0.003	–
Solar	CRPS	+ 0.032±0.048	+ 0.071±0.145	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.730±0.001	+ 0.000±0.000
	NMAE	− 0.013±0.027	+ 0.084±0.153	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.781±0.002	+ 0.000±0.000
	ES-Time	+ 0.008±0.051	− 0.039±0.052	+ 1.036±0.002	− 1.038±0.001	− 0.907±0.069	+ 1.601±0.002	+ 1.025±0.009
All datasets	CRPS	+ 0.012±0.024	+ 0.017±0.015	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.671±0.001	+ 0.000±0.000
	NMAE	+ 0.012±0.043	+ 0.012±0.011	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.962±0.002	+ 0.000±0.000
	ES-Time	+ 0.013±0.016	+ 0.007±0.020	+ 2.765±0.013	+ 0.076±0.004	+ 0.412±0.021	+ 2.421±0.012	+ 3.047±0.022

D.2 Per-model summary

Table 14 reports the joint-score performance per forecasting backbone, averaged over datasets and horizons.

Table 14: **Coupling gain per model** (repeated from the main text). Percent change vs. the uncoupled arm of the same settings; negative is better. ± is the standard deviation over 3 seeds (2026–2028). All settings of a cell are first averaged per seed, and the mean and standard deviation are then taken over the seeds. Red best, blue second. *Avg. marginal & temporal* averages all cells.

Model	Metric (%)	GeNCoup-Dyn (ours)	GeNCoup-Corr (ours)	COBASE	Schaake	SimSch.	GCA	IC
k2VAE	VS	−21.122±0.083	−16.182±0.053	−11.186±0.062	−12.096±0.004	−12.886±0.086	−14.499±0.006	−14.358±0.024
	ES-Chan	− 0.494±0.083	− 1.185±0.031	− 1.107±0.002	− 0.782±0.001	− 0.884±0.009	− 0.547±0.013	− 1.121±0.003
PPM	VS	− 7.955±0.032	−10.677±0.036	− 9.163±0.056	− 9.000±0.012	− 9.888±0.059	− 9.677±0.016	− 9.923±0.019
	ES-Chan	− 1.036±0.118	− 1.179±0.033	− 1.300±0.010	− 0.861±0.008	− 0.841±0.006	− 0.105±0.005	− 0.849±0.007
FMPatch	VS	− 8.430±0.149	− 8.608±0.018	− 7.522±0.063	− 7.084±0.007	− 7.890±0.036	− 7.001±0.010	− 7.835±0.015
	ES-Chan	− 1.243±0.062	− 1.817±0.026	− 1.756±0.012	− 0.931±0.000	− 1.204±0.025	− 0.846±0.007	− 1.601±0.009
DiffPatch	VS	−31.546±0.242	−27.567±0.095	−22.889±0.051	−20.281±0.043	−21.473±0.353	−28.934±0.083	−26.567±0.096
	ES-Chan	− 7.577±0.073	− 6.697±0.077	− 5.231±0.021	− 4.591±0.018	− 5.767±0.071	− 7.212±0.001	− 6.574±0.010
Sundial	VS	− 5.873±0.086	− 5.230±0.017	− 4.695±0.020	− 1.783±0.002	− 4.029±0.003	− 4.357±0.003	− 4.837±0.004
	ES-Chan	− 0.371±0.032	− 0.608±0.010	− 0.788±0.019	+ 0.030±0.003	− 0.390±0.004	+ 0.076±0.003	− 0.608±0.002
Aurora	VS	− 3.073±0.086	− 2.524±0.014	− 2.425±0.016	− 0.092±0.005	− 1.819±0.031	− 1.854±0.005	− 2.218±0.007
	ES-Chan	+ 0.513±0.041	+ 0.785±0.014	+ 0.459±0.112	+ 0.665±0.003	+ 0.520±0.008	+ 1.483±0.005	+ 0.702±0.005
1st count	VS	4	2	0	0	0	0	0
	ES-Chan	1	2	3	0	0	0	0
Avg. marginal & temporal	CRPS	+ 0.014±0.026	+ 0.015±0.016	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.674±0.001	+ 0.000±0.000
	NMAE	+ 0.014±0.049	+ 0.010±0.018	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 1.012±0.002	+ 0.000±0.000
	ES-Time	+ 0.016±0.017	+ 0.004±0.020	+ 3.061±0.015	+ 0.088±0.004	+ 0.399±0.020	+ 2.746±0.014	+ 3.047±0.022

Traffic is excluded so that every arm spans the same datasets: Iman–Conover needs $S > K + 1$ samples for K channels, which at $K = 862$ was beyond our compute budget.

D.3 PER-BACKBONE BREAKDOWN

Table 15: **Experiment results for k2VAE, averaged over horizons.** Percent change vs. the uncoupled arm of the same settings; negative is better. ± is the standard deviation over 3 seeds (2026–2028); all settings of a cell are first averaged per seed. Red best, blue second. *Avg. marginal & temporal* averages all cells.

Dataset	Metric (%)	GeNCoup-Dyn (ours)	GeNCoup-Corr (ours)	COBASE	Schaake	SimSch.	GCA	IC
ETTh1	VS	−29.466±0.574	−22.685±0.015	−20.650±0.081	−20.926±0.040	−18.758±0.536	−21.778±0.052	−21.624±0.082
	ES-Chan	− 1.949±0.027	− 2.258±0.031	− 2.000±0.004	− 1.108±0.004	− 1.782±0.024	− 1.667±0.015	− 2.172±0.014
ETTh2	VS	−17.155±0.036	− 6.475±0.290	− 3.365±0.057	− 3.600±0.044	− 2.991±0.093	− 2.130±0.018	− 1.593±0.031
	ES-Chan	+ 0.618±0.238	− 0.339±0.063	− 0.282±0.007	− 0.174±0.010	− 0.168±0.050	+ 0.180±0.021	− 0.446±0.019
ETTm1	VS	−32.592±0.421	−22.037±0.034	−20.849±0.101	−21.230±0.010	−20.191±0.223	−21.856±0.111	−21.127±0.117
	ES-Chan	− 1.503±0.116	− 2.121±0.017	− 1.843±0.005	− 1.321±0.002	− 1.830±0.030	− 1.559±0.002	− 1.842±0.002
ETTm2	VS	− 4.344±0.082	− 3.019±0.024	− 2.078±0.015	− 0.680±0.008	− 0.958±0.089	+ 1.325±0.027	+ 1.139±0.019
	ES-Chan	+ 1.822±0.153	+ 0.044±0.012	− 0.242±0.003	+ 0.305±0.003	+ 0.446±0.003	+ 0.658±0.020	+ 0.032±0.012
Weather	VS	−15.124±0.171	−10.159±0.022	− 8.314±0.004	− 8.384±0.006	− 8.465±0.024	− 9.441±0.045	− 9.370±0.027
	ES-Chan	+ 0.396±0.054	− 1.186±0.007	− 0.852±0.001	− 0.595±0.004	− 0.924±0.015	− 0.649±0.018	− 1.105±0.010
ECL	VS	− 0.527±0.126	+ 0.821±0.158	− 0.389±0.001	− 0.240±0.002	+ 0.446±0.011	+ 0.983±0.031	+ 0.778±0.050
	ES-Chan	− 0.083±0.132	+ 0.095±0.197	− 0.120±0.001	+ 0.158±0.001	+ 0.228±0.018	+ 0.719±0.036	+ 0.058±0.046
Traffic	VS	− 1.734±0.040	− 2.805±0.024	− 1.264±0.005	− 1.794±0.002	− 1.604±0.017	− 1.848±0.008	–
	ES-Chan	+ 0.372±0.052	+ 2.348±0.031	+ 0.075±0.001	+ 1.475±0.001	+ 3.357±0.029	+ 2.493±0.016	–
Solar	VS	−48.642±0.535	−49.720±0.015	−22.660±0.366	−29.611±0.004	−39.282±0.129	−48.593±0.076	−48.707±0.083
	ES-Chan	− 2.762±0.024	− 2.527±0.010	− 2.409±0.020	− 2.740±0.001	− 2.159±0.023	− 1.509±0.026	− 2.372±0.021
1st count	VS	6	2	0	0	0	0	0
	ES-Chan	1	3	3	0	0	0	1
Avg. marginal & temporal	CRPS	− 0.004±0.014	− 0.007±0.023	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.801±0.002	+ 0.000±0.000
	NMAE	− 0.027±0.030	− 0.024±0.006	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 1.819±0.002	+ 0.000±0.000
	ES-Time	− 0.003±0.031	− 0.019±0.053	+ 0.013±0.003	− 1.497±0.004	− 0.955±0.039	+ 0.186±0.004	+ 0.015±0.002

Table 16: **Detailed experiment results for k2VAE.** Percent change vs. the uncoupled arm of the same settings; negative is better. ± is the standard deviation over 3 seeds (2026–2028); all settings of a cell are first averaged per seed. Red best, blue second. *Avg. marginal & temporal* averages all cells.

Dataset	h	Metric (%)	Uncoupled	GeNCoup-Dyn (ours)	GeNCoup-Corr (ours)	COBASE	Schaake	SimSch.	GCA	IC
ETTh1	96	VS	669.5	−31.754±2.004	−23.302±0.064	−21.335±0.157	−21.529±0.098	−19.990±0.557	−21.850±0.482	−21.699±0.470
		ES-Chan	1.123	− 1.866±0.201	− 2.365±0.079	− 2.107±0.021	− 1.359±0.014	− 1.849±0.053	− 1.813±0.045	− 2.226±0.017
	192	VS	1447	−28.483±1.981	−21.474±0.037	−20.380±0.200	−20.841±0.016	−19.366±1.145	−21.235±0.279	−21.030±0.302
		ES-Chan	1.177	− 1.927±0.082	− 2.201±0.058	− 1.942±0.008	− 1.275±0.004	− 1.680±0.113	− 1.647±0.021	− 2.073±0.031
	336	VS	3254	−27.937±1.619	−23.213±0.062	−21.075±0.086	−22.012±0.048	−18.673±0.278	−22.506±0.381	−22.097±0.438
		ES-Chan	1.298	− 2.224±0.060	− 2.392±0.035	− 2.055±0.017	− 1.332±0.022	− 2.023±0.044	− 1.795±0.018	− 2.178±0.016
	720	VS	7313	−29.692±1.046	−22.749±0.014	−19.809±0.158	−19.322±0.023	−17.004±0.267	−21.521±0.100	−21.670±0.092
		ES-Chan	1.329	− 1.780±0.123	− 2.074±0.042	− 1.894±0.005	− 0.464±0.007	− 1.577±0.012	− 1.415±0.020	− 2.211±0.031
	Avg	VS	3171	−29.466±0.574	−22.685±0.015	−20.650±0.081	−20.926±0.040	−18.758±0.536	−21.778±0.052	−21.624±0.082
		ES-Chan	1.232	− 1.949±0.027	− 2.258±0.031	− 2.000±0.004	− 1.108±0.004	− 1.782±0.024	− 1.667±0.015	− 2.172±0.014
ETTh2	96	VS	358.6	+ 1.658±0.804	− 0.312±0.025	− 0.090±0.011	+ 0.030±0.046	− 0.241±0.124	+ 0.754±0.008	+ 0.338±0.039
		ES-Chan	1.067	+ 1.900±0.903	+ 0.575±0.012	+ 0.212±0.006	+ 0.493±0.001	+ 0.682±0.086	+ 0.813±0.032	+ 0.135±0.022
	192	VS	1414	−16.646±0.886	− 6.307±1.111	− 3.725±0.096	− 4.284±0.188	− 2.380±0.216	− 2.092±0.108	− 2.064±0.195
		ES-Chan	1.184	+ 1.581±0.098	+ 0.003±0.241	− 0.060±0.017	+ 0.110±0.010	+ 0.438±0.024	+ 0.748±0.050	− 0.070±0.001
	336	VS	1937	−12.009±0.110	− 2.861±0.033	− 0.569±0.039	+ 1.103±0.024	+ 1.581±0.438	+ 1.801±0.125	+ 2.102±0.157
		ES-Chan	1.225	+ 3.155±0.085	− 0.030±0.034	− 0.250±0.013	+ 0.975±0.010	+ 0.526±0.069	+ 0.521±0.037	− 0.305±0.027
	720	VS	1.28×10^4	−41.623±0.143	−16.421±0.133	− 9.075±0.219	−11.249±0.015	−10.924±1.054	− 8.984±0.099	− 6.747±0.132
		ES-Chan	1.995	− 4.164±0.067	− 1.905±0.087	− 1.028±0.014	− 2.272±0.023	− 2.317±0.311	− 1.362±0.026	− 1.544±0.063
	Avg	VS	4129	−17.155±0.036	− 6.475±0.290	− 3.365±0.057	− 3.600±0.044	− 2.991±0.093	− 2.130±0.018	− 1.593±0.031
		ES-Chan	1.368	+ 0.618±0.238	− 0.339±0.063	− 0.282±0.007	− 0.174±0.010	− 0.168±0.050	+ 0.180±0.021	− 0.446±0.019
1st count		VS		7	1	0	0	0	0	0
		ES-Chan		1	3	0	0	0	0	4
Avg. marginal & temporal		CRPS		+ 0.011±0.065	− 0.034±0.041	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.824±0.004	+ 0.000±0.000
		NMAE		− 0.006±0.154	− 0.090±0.043	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 2.692±0.024	+ 0.000±0.000
		ES-Time		+ 0.013±0.105	− 0.032±0.150	+ 0.002±0.012	− 1.433±0.017	− 0.930±0.081	− 0.007±0.018	+ 0.013±0.011

Table 17: **Detailed experiment results for k2VAE.** Percent change vs. the uncoupled arm of the same settings; negative is better. ± is the standard deviation over 3 seeds (2026–2028); all settings of a cell are first averaged per seed. Red best, blue second. *Avg. marginal & temporal* averages all cells.

Dataset	h	Metric (%)	Uncoupled	GeNCoup-Dyn (ours)	GeNCoup-Corr (ours)	COBASE	Schaake	SimSch.	GCA	IC
ETTm1	96	VS	590.3	−33.120±0.529	−21.765±0.057	−20.052±0.278	−21.428±0.034	−20.312±0.513	−21.507±0.109	−20.726±0.109
		ES-Chan	0.9952	− 1.154±0.187	− 2.030±0.040	− 1.793±0.016	− 1.341±0.005	− 1.621±0.057	− 1.524±0.008	− 1.787±0.010
	192	VS	1299	−31.803±0.795	−21.562±0.067	−20.130±0.053	−20.632±0.017	−19.904±0.173	−21.468±0.202	−20.919±0.234
		ES-Chan	1.065	− 1.356±0.235	− 2.003±0.032	− 1.730±0.014	− 1.093±0.003	− 1.598±0.095	− 1.438±0.009	− 1.740±0.003
	336	VS	2644	−32.700±2.187	−21.894±0.035	−21.127±0.065	−21.339±0.006	−20.525±0.314	−21.873±0.331	−21.010±0.364
		ES-Chan	1.155	− 1.540±0.341	− 2.179±0.007	− 1.909±0.007	− 1.546±0.005	− 2.080±0.015	− 1.603±0.009	− 1.876±0.005
	720	VS	6819	−32.745±1.010	−22.926±0.035	−22.086±0.215	−21.522±0.007	−20.023±0.471	−22.574±0.311	−21.853±0.355
		ES-Chan	1.288	− 1.964±0.101	− 2.273±0.036	− 1.942±0.016	− 1.306±0.001	− 2.022±0.027	− 1.671±0.001	− 1.965±0.016
	Avg	VS	2838	−32.592±0.421	−22.037±0.034	−20.849±0.101	−21.230±0.010	−20.191±0.223	−21.856±0.111	−21.127±0.117
		ES-Chan	1.126	− 1.503±0.116	− 2.121±0.017	− 1.843±0.005	− 1.321±0.002	− 1.830±0.030	− 1.559±0.002	− 1.842±0.002
ETTm2	96	VS	233.5	− 2.816±0.282	− 2.065±0.089	− 1.312±0.041	− 0.655±0.027	− 0.251±0.139	+ 0.936±0.118	+ 0.745±0.096
		ES-Chan	0.7368	+ 1.875±0.752	+ 0.213±0.010	− 0.067±0.004	+ 0.349±0.003	+ 0.702±0.051	+ 0.755±0.046	+ 0.120±0.037
	192	VS	576.5	− 2.301±0.034	− 1.977±0.023	− 1.562±0.037	− 0.582±0.012	− 0.536±0.083	+ 0.275±0.052	+ 0.119±0.053
		ES-Chan	0.8626	+ 2.117±0.069	+ 0.102±0.009	− 0.223±0.009	+ 0.308±0.005	+ 0.463±0.027	+ 0.635±0.044	+ 0.014±0.045
	336	VS	1161	− 2.449±0.056	− 2.025±0.017	− 1.303±0.027	− 0.309±0.007	− 0.727±0.235	+ 1.211±0.039	+ 0.974±0.040
		ES-Chan	0.9509	+ 1.230±0.067	+ 0.250±0.014	− 0.053±0.009	+ 0.401±0.004	+ 0.482±0.015	+ 0.769±0.014	+ 0.131±0.019
	720	VS	3496	− 9.812±0.116	− 6.009±0.021	− 4.136±0.050	− 1.176±0.008	− 2.320±0.300	+ 2.879±0.037	+ 2.718±0.040
		ES-Chan	1.135	+ 2.065±0.032	− 0.390±0.020	− 0.625±0.004	+ 0.161±0.004	+ 0.139±0.032	+ 0.475±0.018	− 0.138±0.019
	Avg	VS	1367	− 4.344±0.082	− 3.019±0.024	− 2.078±0.015	− 0.680±0.008	− 0.958±0.089	+ 1.325±0.027	+ 1.139±0.019
		ES-Chan	0.9213	+ 1.822±0.153	+ 0.044±0.012	− 0.242±0.003	+ 0.305±0.003	+ 0.446±0.003	+ 0.658±0.020	+ 0.032±0.012
Weather	96	VS	3332	−20.884±0.067	−12.052±0.007	− 9.744±0.029	− 9.712±0.023	−10.018±0.083	−11.225±0.101	−11.227±0.069
		ES-Chan	1.286	+ 0.326±0.162	− 1.034±0.010	− 0.695±0.012	− 0.304±0.012	− 0.943±0.023	− 0.675±0.048	− 0.983±0.016
	192	VS	7659	−16.519±0.289	−10.510±0.069	− 8.607±0.015	− 8.710±0.002	− 8.726±0.061	− 9.779±0.037	− 9.723±0.034
		ES-Chan	1.425	+ 0.428±0.033	− 1.101±0.025	− 0.798±0.009	− 0.497±0.002	− 0.813±0.025	− 0.584±0.011	− 1.001±0.020
	336	VS	1.71×10^4	−13.080±0.179	− 9.015±0.029	− 7.474±0.012	− 7.466±0.004	− 7.548±0.008	− 8.575±0.023	− 8.301±0.007
		ES-Chan	1.733	+ 0.562±0.096	− 1.259±0.006	− 0.913±0.007	− 0.758±0.004	− 0.953±0.058	− 0.715±0.013	− 1.188±0.008
	720	VS	4.06×10^4	−10.014±0.305	− 9.059±0.019	− 7.430±0.014	− 7.649±0.004	− 7.567±0.046	− 8.184±0.022	− 8.230±0.014
		ES-Chan	2.001	+ 0.271±0.220	− 1.352±0.010	− 1.001±0.010	− 0.821±0.002	− 0.988±0.021	− 0.622±0.010	− 1.246±0.012
	Avg	VS	1.72×10^4	−15.124±0.171	−10.159±0.022	− 8.314±0.004	− 8.384±0.006	− 8.465±0.024	− 9.441±0.045	− 9.370±0.027
		ES-Chan	1.611	+ 0.396±0.054	− 1.186±0.007	− 0.852±0.001	− 0.595±0.004	− 0.924±0.015	− 0.649±0.018	− 1.105±0.010
1st count		VS		12	0	0	0	0	0	0
		ES-Chan		0	8	4	0	0	0	0
Avg. marginal & temporal		CRPS		− 0.002±0.008	+ 0.002±0.008	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.757±0.003	+ 0.000±0.000
		NMAE		− 0.028±0.024	− 0.014±0.023	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 1.851±0.016	+ 0.000±0.000
		ES-Time		− 0.017±0.027	− 0.021±0.021	+ 0.012±0.001	− 1.801±0.006	− 1.511±0.012	+ 0.083±0.003	+ 0.011±0.000

Table 18: **Detailed experiment results for k2VAE.** Percent change vs. the uncoupled arm of the same settings; negative is better. ± is the standard deviation over 3 seeds (2026–2028); all settings of a cell are first averaged per seed. Red best, blue second. *Avg. marginal & temporal* averages all cells.

Dataset	h	Metric (%)	Uncoupled	GeNCoup-Dyn (ours)	GeNCoup-Corr (ours)	COBASE	Schaake	SimSch.	GCA	IC
ECL	96	VS	8.11×10^{5}	+ 0.033±0.055	+ 1.520±0.091	− 0.207±0.009	+ 0.126±0.003	+ 0.970±0.009	+ 1.768±0.038	+ 1.373±0.075
		ES-Chan	5.916	+ 0.166±0.088	+ 0.384±0.063	− 0.057±0.004	+ 0.400±0.004	+ 0.553±0.039	+ 0.973±0.053	+ 0.246±0.045
	192	VS	2.10×10^{6}	− 0.664±0.312	+ 0.286±0.639	− 0.497±0.010	− 0.443±0.003	− 0.027±0.064	+ 0.208±0.054	+ 0.224±0.140
		ES-Chan	7.39	− 0.075±0.327	+ 0.130±0.683	− 0.094±0.001	+ 0.217±0.001	+ 0.178±0.080	+ 0.914±0.033	+ 0.060±0.148
	336	VS	2.69×10^{6}	− 0.536±0.102	+ 1.357±0.059	− 0.403±0.004	− 0.115±0.001	+ 0.890±0.054	+ 1.919±0.021	+ 1.320±0.010
		ES-Chan	5.591	− 0.119±0.105	+ 0.185±0.101	− 0.157±0.003	+ 0.274±0.002	+ 0.422±0.029	+ 0.669±0.016	+ 0.157±0.011
	720	VS	9.02×10^{6}	− 0.941±0.152	+ 0.119±0.114	− 0.452±0.013	− 0.529±0.000	− 0.048±0.050	+ 0.036±0.037	+ 0.196±0.061
		ES-Chan	8.265	− 0.305±0.160	− 0.320±0.139	− 0.173±0.001	− 0.257±0.002	− 0.240±0.023	+ 0.320±0.044	− 0.229±0.050
	Avg	VS	3.66×10^{6}	− 0.527±0.126	+ 0.821±0.158	− 0.389±0.001	− 0.240±0.002	+ 0.446±0.011	+ 0.983±0.031	+ 0.778±0.050
		ES-Chan	6.79	− 0.083±0.132	+ 0.095±0.197	− 0.120±0.001	+ 0.158±0.001	+ 0.228±0.018	+ 0.719±0.036	+ 0.058±0.046
Traffic	96	VS	8.29×10^{6}	− 1.115±0.069	− 3.168±0.042	− 1.392±0.004	− 1.845±0.006	− 1.683±0.046	− 1.930±0.016	–
		ES-Chan	14.62	+ 0.059±0.034	+ 2.448±0.041	+ 0.109±0.004	+ 1.506±0.002	+ 3.370±0.111	+ 2.605±0.018	–
	192	VS	1.73×10^{7}	− 1.751±0.082	− 3.630±0.053	− 1.545±0.025	− 2.372±0.003	− 2.566±0.014	− 2.636±0.017	–
		ES-Chan	15.06	+ 0.103±0.073	+ 2.315±0.064	+ 0.071±0.003	+ 1.469±0.002	+ 3.254±0.077	+ 2.412±0.022	–
	336	VS	3.07×10^{7}	− 2.159±0.086	− 2.734±0.009	− 1.143±0.030	− 1.878±0.002	− 1.627±0.035	− 1.902±0.008	–
		ES-Chan	15.28	+ 0.463±0.102	+ 2.277±0.011	+ 0.040±0.005	+ 1.466±0.001	+ 3.449±0.072	+ 2.452±0.009	–
	720	VS	6.85×10^{7}	− 1.911±0.051	− 1.688±0.031	− 0.977±0.020	− 1.081±0.000	− 0.540±0.026	− 0.925±0.009	–
		ES-Chan	15.85	+ 0.863±0.133	+ 2.352±0.031	+ 0.078±0.006	+ 1.460±0.001	+ 3.356±0.103	+ 2.503±0.028	–
	Avg	VS	3.12×10^{7}	− 1.734±0.040	− 2.805±0.024	− 1.264±0.005	− 1.794±0.002	− 1.604±0.017	− 1.848±0.008	–
		ES-Chan	15.2	+ 0.372±0.052	+ 2.348±0.031	+ 0.075±0.001	+ 1.475±0.001	+ 3.357±0.029	+ 2.493±0.016	–
Solar	96	VS	2.19×10^{5}	−48.845±1.728	−50.360±0.041	−20.517±0.485	−28.764±0.019	−36.889±0.494	−48.901±0.166	−49.014±0.147
		ES-Chan	2.37	− 2.331±0.209	− 2.010±0.036	− 2.026±0.034	− 2.377±0.005	− 1.669±0.059	− 0.914±0.060	− 1.894±0.088
	192	VS	4.77×10^{5}	−49.439±1.104	−50.606±0.036	−27.156±0.850	−28.949±0.018	−39.527±0.202	−49.364±0.153	−49.490±0.175
		ES-Chan	2.643	− 2.518±0.120	− 2.315±0.074	− 2.496±0.046	− 2.478±0.004	− 1.868±0.034	− 1.283±0.056	− 2.157±0.070
	336	VS	8.08×10^{5}	−47.190±1.354	−48.324±0.014	−19.668±0.863	−30.235±0.009	−40.037±0.204	−47.423±0.041	−47.529±0.054
		ES-Chan	2.76	− 2.871±0.232	− 2.724±0.019	− 2.327±0.077	− 3.032±0.002	− 2.361±0.105	− 1.700±0.016	− 2.478±0.030
	720	VS	1.67×10^{6}	−49.095±0.211	−49.591±0.006	−23.296±0.624	−30.497±0.005	−40.674±0.293	−48.683±0.028	−48.796±0.048
		ES-Chan	2.806	− 3.327±0.037	− 3.060±0.037	− 2.789±0.051	− 3.073±0.003	− 2.739±0.052	− 2.137±0.014	− 2.960±0.022
	Avg	VS	7.94×10^{5}	−48.642±0.535	−49.720±0.015	−22.660±0.366	−29.611±0.004	−39.282±0.129	−48.593±0.076	−48.707±0.083
		ES-Chan	2.645	− 2.762±0.024	− 2.527±0.010	− 2.409±0.020	− 2.740±0.001	− 2.159±0.023	− 1.509±0.026	− 2.372±0.021
1st count		VS		4	7	1	0	0	0	0
		ES-Chan		3	1	6	2	0	0	0
Avg. marginal & temporal		CRPS		− 0.017±0.033	+ 0.001±0.036	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.830±0.003	+ 0.000±0.000
		NMAE		− 0.040±0.063	+ 0.010±0.026	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 1.205±0.007	+ 0.000±0.000
		ES-Time		− 0.000±0.002	− 0.009±0.036	+ 0.020±0.001	− 1.235±0.001	− 0.417±0.049	+ 0.417±0.002	+ 0.024±0.005

Table 19: **Experiment results for PPM, averaged over horizons.** Percent change vs. the uncoupled arm of the same settings; negative is better. ± is the standard deviation over 3 seeds (2026–2028); all settings of a cell are first averaged per seed. Red best, blue second. *Avg. marginal & temporal* averages all cells.

Dataset	Metric (%)	GeNCoup-Dyn (ours)	GeNCoup-Corr (ours)	COBASE	Schaake	SimSch.	GCA	IC
ETTh1	VS	−13.603±0.904	−13.725±0.209	−13.655±0.057	−14.733±0.066	−12.858±0.121	−14.207±0.071	−13.349±0.085
	ES-Chan	− 1.454±0.282	− 1.734±0.082	− 1.712±0.009	− 1.190±0.002	− 1.570±0.019	− 0.998±0.015	− 1.674±0.014
ETTh2	VS	− 3.271±1.097	− 3.746±0.080	− 3.526±0.124	− 2.825±0.096	− 3.164±0.257	+ 0.231±0.156	− 0.278±0.157
	ES-Chan	− 0.729±0.718	− 1.218±0.031	− 1.217±0.034	− 0.636±0.027	− 0.591±0.071	+ 0.074±0.023	− 0.526±0.025
ETTm1	VS	−10.491±0.363	−10.224±0.119	− 9.508±0.038	−10.477±0.026	−10.518±0.164	−10.317±0.034	−10.389±0.011
	ES-Chan	− 1.328±0.067	− 1.255±0.116	− 1.284±0.007	− 0.836±0.003	− 0.952±0.010	− 0.534±0.002	− 1.301±0.006
ETTm2	VS	− 2.118±0.150	− 2.412±0.053	− 1.908±0.006	− 0.845±0.039	− 1.449±0.048	+ 0.534±0.031	− 0.166±0.031
	ES-Chan	− 0.961±0.054	− 1.294±0.037	− 1.165±0.002	− 0.510±0.002	− 0.399±0.044	+ 0.108±0.032	− 0.570±0.028
Weather	VS	− 3.078±0.128	− 3.518±0.238	− 3.274±0.007	− 3.114±0.004	− 3.164±0.018	− 2.953±0.015	− 3.250±0.006
	ES-Chan	+ 0.159±0.059	− 0.041±0.120	− 0.158±0.020	− 0.027±0.021	− 0.013±0.010	+ 0.944±0.021	− 0.003±0.020
ECL	VS	− 0.055±0.044	+ 2.350±0.013	+ 0.367±0.004	+ 0.831±0.001	+ 2.050±0.039	+ 3.948±0.009	+ 3.196±0.016
	ES-Chan	− 0.029±0.035	+ 0.221±0.043	− 0.079±0.003	+ 0.271±0.001	+ 0.465±0.009	+ 1.295±0.008	+ 0.586±0.001
Traffic	VS	− 0.082±0.004	+ 0.294±0.082	− 0.432±0.002	+ 0.589±0.002	+ 2.380±0.055	+ 1.822±0.007	–
	ES-Chan	− 0.188±0.008	+ 1.754±0.081	− 0.548±0.018	− 0.342±0.004	+ 1.341±0.079	+ 3.244±0.007	–
Solar	VS	−23.072±1.020	−43.467±0.056	−32.640±0.306	−31.835±0.015	−40.111±0.209	−44.977±0.032	−45.227±0.020
	ES-Chan	− 2.913±0.069	− 2.930±0.083	− 3.488±0.014	− 3.103±0.003	− 2.826±0.097	− 1.625±0.022	− 2.456±0.024
1st count	VS	1	3	1	1	1	0	1
	ES-Chan	1	3	4	0	0	0	0
Avg. marginal & temporal	CRPS	+ 0.035±0.095	+ 0.024±0.050	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.676±0.002	+ 0.000±0.000
	NMAE	+ 0.060±0.086	+ 0.046±0.096	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.544±0.004	+ 0.000±0.000
	ES-Time	− 0.017±0.031	+ 0.006±0.122	− 0.663±0.024	− 1.343±0.006	− 1.305±0.011	− 0.295±0.021	+ 0.093±0.024

Table 20: **Detailed experiment results for PPM.** Percent change vs. the uncoupled arm of the same settings; negative is better. ± is the standard deviation over 3 seeds (2026–2028); all settings of a cell are first averaged per seed. Red best, blue second. *Avg. marginal & temporal* averages all cells.

Dataset	h	Metric (%)	Uncoupled	GeNCoup-Dyn (ours)	GeNCoup-Corr (ours)	COBASE	Schaake	SimSch.	GCA	IC
ETTh1	96	VS	469.3	−12.115±0.481	−12.134±0.261	−12.085±0.018	−12.252±0.099	−11.753±0.197	−12.099±0.118	−11.584±0.094
		ES-Chan	1.007	− 1.372±0.204	− 1.499±0.365	− 1.581±0.014	− 1.204±0.008	− 1.426±0.064	− 0.907±0.036	− 1.416±0.012
	192	VS	1147	−14.290±1.456	−14.015±0.074	−14.540±0.145	−15.296±0.106	−13.330±0.259	−14.679±0.091	−13.490±0.141
		ES-Chan	1.107	− 1.203±0.018	− 1.595±0.016	− 1.626±0.016	− 1.329±0.004	− 1.565±0.071	− 0.946±0.026	− 1.502±0.003
	336	VS	2204	−14.880±2.364	−15.069±0.237	−15.379±0.142	−16.645±0.140	−13.421±0.070	−15.918±0.048	−14.664±0.047
		ES-Chan	1.171	− 1.347±0.198	− 1.968±0.156	− 1.813±0.038	− 1.288±0.009	− 1.637±0.010	− 1.144±0.037	− 1.735±0.030
	720	VS	5294	−13.128±1.056	−13.682±0.792	−12.617±0.274	−14.738±0.369	−12.929±0.090	−14.131±0.288	−13.657±0.245
		ES-Chan	1.278	− 1.895±0.839	− 1.874±0.580	− 1.828±0.029	− 0.937±0.027	− 1.652±0.057	− 0.994±0.028	− 2.042±0.065
	Avg	VS	2279	−13.603±0.904	−13.725±0.209	−13.655±0.057	−14.733±0.066	−12.858±0.121	−14.207±0.071	−13.349±0.085
		ES-Chan	1.141	− 1.454±0.282	− 1.734±0.082	− 1.712±0.009	− 1.190±0.002	− 1.570±0.019	− 0.998±0.015	− 1.674±0.014
ETTh2	96	VS	281.3	− 2.127±1.776	− 1.743±0.120	− 1.965±0.163	− 1.610±0.188	− 2.189±0.082	+ 1.046±0.142	+ 0.447±0.132
		ES-Chan	0.835	− 0.768±0.789	− 0.633±0.223	− 0.900±0.045	− 0.158±0.069	− 0.033±0.106	+ 0.402±0.030	− 0.227±0.022
	192	VS	690.9	− 2.954±1.032	− 4.347±0.399	− 3.924±0.087	− 3.540±0.037	− 3.624±0.220	− 0.603±0.172	− 0.884±0.114
		ES-Chan	0.9771	− 0.516±0.745	− 1.458±0.310	− 1.367±0.014	− 0.551±0.029	− 0.342±0.062	− 0.044±0.054	− 0.584±0.044
	336	VS	1378	− 3.051±1.432	− 3.622±0.029	− 3.054±0.074	− 2.299±0.092	− 2.578±0.448	+ 0.657±0.261	+ 0.167±0.321
		ES-Chan	1.085	− 0.996±1.005	− 1.363±0.159	− 1.302±0.046	− 0.772±0.023	− 0.788±0.115	− 0.010±0.068	− 0.642±0.084
	720	VS	3010	− 4.951±0.656	− 5.270±0.294	− 5.162±0.200	− 3.852±0.238	− 4.267±0.697	− 0.175±0.231	− 0.842±0.220
		ES-Chan	1.106	− 0.637±0.555	− 1.416±0.375	− 1.298±0.056	− 1.062±0.035	− 1.200±0.189	− 0.051±0.039	− 0.649±0.011
	Avg	VS	1340	− 3.271±1.097	− 3.746±0.080	− 3.526±0.124	− 2.825±0.096	− 3.164±0.257	+ 0.231±0.156	− 0.278±0.157
		ES-Chan	1.001	− 0.729±0.718	− 1.218±0.031	− 1.217±0.034	− 0.636±0.027	− 0.591±0.071	+ 0.074±0.023	− 0.526±0.025
1st count		VS		0	3	0	4	1	0	0
		ES-Chan		0	4	3	0	0	0	1
Avg. marginal & temporal		CRPS		+ 0.083±0.342	+ 0.040±0.044	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.781±0.007	+ 0.000±0.000
		NMAE		+ 0.144±0.339	+ 0.073±0.063	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.813±0.017	+ 0.000±0.000
		ES-Time		− 0.088±0.170	+ 0.039±0.347	+ 0.861±0.006	+ 0.298±0.010	+ 0.298±0.015	+ 1.282±0.001	+ 0.862±0.006

Table 21: **Detailed experiment results for PPM.** Percent change vs. the uncoupled arm of the same settings; negative is better. ± is the standard deviation over 3 seeds (2026–2028); all settings of a cell are first averaged per seed. Red best, blue second. *Avg. marginal & temporal* averages all cells.

Dataset	h	Metric (%)	Uncoupled	GeNCoup-Dyn (ours)	GeNCoup-Corr (ours)	COBASE	Schaake	SimSch.	GCA	IC
ETTm1	96	VS	369.4	− 9.002±0.647	− 8.836±0.491	− 7.580±0.148	− 8.346±0.182	− 8.939±0.044	− 8.267±0.227	− 8.670±0.211
		ES-Chan	0.8781	− 0.663±0.282	− 0.714±0.165	− 0.926±0.069	− 0.329±0.079	− 0.473±0.053	− 0.082±0.085	− 0.857±0.086
	192	VS	828	−10.384±0.336	−10.107±0.153	− 9.148±0.127	−10.230±0.046	−10.276±0.250	− 9.830±0.014	−10.061±0.056
		ES-Chan	0.9594	− 1.377±0.083	− 1.249±0.177	− 1.242±0.036	− 0.869±0.036	− 0.982±0.016	− 0.484±0.041	− 1.270±0.033
	336	VS	1635	−11.923±0.505	−11.188±0.260	−10.676±0.152	−11.621±0.164	−11.426±0.328	−11.527±0.172	−11.356±0.127
		ES-Chan	1.032	− 1.817±0.166	− 1.650±0.199	− 1.513±0.051	− 1.144±0.044	− 1.347±0.038	− 0.825±0.061	− 1.562±0.052
	720	VS	3938	−10.653±0.329	−10.763±0.175	−10.628±0.063	−11.710±0.087	−11.430±0.133	−11.645±0.123	−11.466±0.172
		ES-Chan	1.126	− 1.457±0.156	− 1.405±0.203	− 1.455±0.028	− 1.001±0.029	− 1.005±0.084	− 0.745±0.026	− 1.513±0.028
	Avg	VS	1693	−10.491±0.363	−10.224±0.119	− 9.508±0.038	−10.477±0.026	−10.518±0.164	−10.317±0.034	−10.389±0.011
		ES-Chan	0.9991	− 1.328±0.067	− 1.255±0.116	− 1.284±0.007	− 0.836±0.003	− 0.952±0.010	− 0.534±0.002	− 1.301±0.006
ETTm2	96	VS	194.2	− 1.019±0.062	− 1.253±0.065	− 0.566±0.019	+ 0.029±0.008	− 0.609±0.020	+ 1.084±0.056	+ 0.328±0.041
		ES-Chan	0.6676	− 0.722±0.079	− 1.088±0.094	− 0.994±0.013	− 0.201±0.016	− 0.021±0.041	+ 0.326±0.037	− 0.431±0.024
	192	VS	503.3	− 1.839±0.672	− 1.998±0.184	− 1.389±0.055	− 0.410±0.057	− 1.196±0.192	+ 0.796±0.069	+ 0.061±0.063
		ES-Chan	0.7885	− 1.098±0.310	− 1.339±0.034	− 1.187±0.013	− 0.490±0.022	− 0.466±0.081	+ 0.131±0.045	− 0.539±0.044
	336	VS	1022	− 2.002±0.293	− 2.413±0.033	− 2.020±0.036	− 0.956±0.033	− 1.439±0.090	+ 0.575±0.108	− 0.114±0.081
		ES-Chan	0.8726	− 0.955±0.332	− 1.275±0.046	− 1.221±0.008	− 0.574±0.006	− 0.456±0.030	+ 0.039±0.053	− 0.615±0.043
	720	VS	2777	− 3.613±0.188	− 3.985±0.068	− 3.655±0.010	− 2.042±0.069	− 2.553±0.096	− 0.320±0.090	− 0.937±0.106
		ES-Chan	1.027	− 1.068±0.087	− 1.474±0.037	− 1.257±0.004	− 0.774±0.013	− 0.655±0.071	− 0.063±0.037	− 0.697±0.034
	Avg	VS	1124	− 2.118±0.150	− 2.412±0.053	− 1.908±0.006	− 0.845±0.039	− 1.449±0.048	+ 0.534±0.031	− 0.166±0.031
		ES-Chan	0.839	− 0.961±0.054	− 1.294±0.037	− 1.165±0.002	− 0.510±0.002	− 0.399±0.044	+ 0.108±0.032	− 0.570±0.028
Weather	96	VS	2659	− 1.580±0.418	− 2.149±0.442	− 2.241±0.028	− 2.084±0.026	− 2.051±0.029	− 1.853±0.027	− 2.164±0.015
		ES-Chan	1.209	+ 0.551±0.179	+ 0.183±0.292	− 0.079±0.022	+ 0.059±0.007	+ 0.010±0.035	+ 1.007±0.015	+ 0.074±0.017
	192	VS	6589	− 3.285±0.447	− 3.557±0.273	− 3.111±0.018	− 2.948±0.005	− 2.971±0.016	− 2.621±0.043	− 3.043±0.012
		ES-Chan	1.38	+ 0.071±0.035	− 0.145±0.141	− 0.108±0.027	− 0.006±0.032	+ 0.002±0.051	+ 1.078±0.037	+ 0.077±0.018
	336	VS	1.36×10^4	− 3.256±0.408	− 3.595±0.595	− 3.484±0.032	− 3.293±0.027	− 3.398±0.023	− 3.246±0.013	− 3.471±0.025
		ES-Chan	1.575	− 0.082±0.033	− 0.002±0.260	− 0.158±0.049	− 0.031±0.046	+ 0.039±0.039	+ 0.907±0.057	− 0.017±0.052
	720	VS	3.44×10^4	− 4.190±0.075	− 4.770±0.239	− 4.259±0.036	− 4.130±0.025	− 4.238±0.055	− 4.093±0.002	− 4.322±0.025
		ES-Chan	1.801	+ 0.098±0.196	− 0.200±0.128	− 0.287±0.024	− 0.131±0.020	− 0.102±0.042	+ 0.785±0.035	− 0.148±0.016
	Avg	VS	1.43×10^4	− 3.078±0.128	− 3.518±0.238	− 3.274±0.007	− 3.114±0.004	− 3.164±0.018	− 2.953±0.015	− 3.250±0.006
		ES-Chan	1.491	+ 0.159±0.059	− 0.041±0.120	− 0.158±0.020	− 0.027±0.021	− 0.013±0.010	+ 0.944±0.021	− 0.003±0.020
1st count		VS		3	7	1	1	0	0	0
		ES-Chan		2	5	4	0	0	0	1
Avg. marginal & temporal		CRPS		− 0.003±0.028	− 0.009±0.047	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.614±0.001	+ 0.000±0.000
		NMAE		+ 0.035±0.008	− 0.035±0.075	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.400±0.004	+ 0.000±0.000
		ES-Time		+ 0.022±0.044	+ 0.026±0.007	+ 0.380±0.059	− 0.342±0.015	− 0.579±0.035	+ 0.643±0.054	+ 0.380±0.059

Table 22: **Detailed experiment results for PPM.** Percent change vs. the uncoupled arm of the same settings; negative is better. ± is the standard deviation over 3 seeds (2026–2028); all settings of a cell are first averaged per seed. Red best, blue second. *Avg. marginal & temporal* averages all cells.

Dataset	h	Metric (%)	Uncoupled	GeNCoup-Dyn (ours)	GeNCoup-Corr (ours)	COBASE	Schaake	SimSch.	GCA	IC
ECL	96	VS	6.56×10^{5}	− 0.045±0.022	+ 1.939±0.072	+ 0.444±0.031	+ 0.983±0.001	+ 2.066±0.041	+ 3.761±0.011	+ 2.984±0.007
		ES-Chan	5.069	− 0.059±0.034	+ 0.276±0.161	+ 0.016±0.005	+ 0.422±0.003	+ 0.620±0.010	+ 1.515±0.014	+ 0.733±0.015
	192	VS	1.35×10^{6}	− 0.052±0.057	+ 1.875±0.032	+ 0.098±0.012	+ 0.473±0.002	+ 1.536±0.112	+ 3.412±0.006	+ 2.640±0.022
		ES-Chan	5.271	− 0.013±0.060	+ 0.181±0.048	− 0.144±0.004	+ 0.235±0.001	+ 0.403±0.027	+ 1.214±0.028	+ 0.536±0.015
	336	VS	2.53×10^{6}	− 0.001±0.035	+ 2.855±0.042	+ 0.583±0.043	+ 1.133±0.002	+ 2.460±0.143	+ 4.439±0.009	+ 3.729±0.031
		ES-Chan	5.571	+ 0.008±0.027	+ 0.218±0.061	− 0.075±0.000	+ 0.284±0.002	+ 0.461±0.018	+ 1.270±0.006	+ 0.604±0.012
	720	VS	6.16×10^{6}	− 0.122±0.087	+ 2.730±0.030	+ 0.342±0.035	+ 0.735±0.004	+ 2.139±0.174	+ 4.179±0.020	+ 3.432±0.020
		ES-Chan	6.097	− 0.051±0.063	+ 0.209±0.012	− 0.111±0.009	+ 0.145±0.003	+ 0.377±0.035	+ 1.181±0.007	+ 0.470±0.015
	Avg	VS	2.68×10^{6}	− 0.055±0.044	+ 2.350±0.013	+ 0.367±0.004	+ 0.831±0.001	+ 2.050±0.039	+ 3.948±0.009	+ 3.196±0.016
		ES-Chan	5.502	− 0.029±0.035	+ 0.221±0.043	− 0.079±0.003	+ 0.271±0.001	+ 0.465±0.009	+ 1.295±0.008	+ 0.586±0.001
Traffic	96	VS	7.75×10^{6}	− 0.056±0.096	+ 0.338±0.161	− 0.408±0.011	+ 0.619±0.002	+ 2.453±0.094	+ 1.990±0.026	–
		ES-Chan	12.61	− 0.108±0.019	+ 1.940±0.066	− 0.503±0.029	− 0.243±0.006	+ 1.542±0.123	+ 3.403±0.049	–
	192	VS	1.55×10^{7}	− 0.092±0.068	+ 0.147±0.050	− 0.440±0.003	+ 0.551±0.008	+ 2.208±0.087	+ 1.708±0.005	–
		ES-Chan	12.63	− 0.197±0.034	+ 1.662±0.056	− 0.612±0.026	− 0.407±0.011	+ 1.183±0.044	+ 3.160±0.022	–
	336	VS	2.76×10^{7}	− 0.015±0.059	+ 0.559±0.116	− 0.313±0.006	+ 0.771±0.010	+ 2.595±0.045	+ 1.971±0.036	–
		ES-Chan	12.81	− 0.152±0.069	+ 1.939±0.176	− 0.443±0.010	− 0.297±0.009	+ 1.382±0.027	+ 3.390±0.054	–
	720	VS	6.26×10^{7}	− 0.166±0.055	+ 0.130±0.187	− 0.567±0.005	+ 0.415±0.001	+ 2.265±0.085	+ 1.618±0.015	–
		ES-Chan	13.28	− 0.295±0.088	+ 1.477±0.216	− 0.635±0.015	− 0.419±0.002	+ 1.258±0.132	+ 3.021±0.017	–
	Avg	VS	2.84×10^{7}	− 0.082±0.004	+ 0.294±0.082	− 0.432±0.002	+ 0.589±0.002	+ 2.380±0.055	+ 1.822±0.007	–
		ES-Chan	12.83	− 0.188±0.008	+ 1.754±0.081	− 0.548±0.018	− 0.342±0.004	+ 1.341±0.079	+ 3.244±0.007	–
Solar	96	VS	1.63×10^{5}	−16.452±1.826	−37.839±0.199	−30.086±0.671	−28.230±0.036	−34.238±0.208	−38.642±0.114	−38.942±0.043
		ES-Chan	2.631	− 1.616±0.165	− 1.362±0.099	− 2.303±0.023	− 1.661±0.019	− 1.232±0.201	+ 0.267±0.100	− 0.672±0.109
	192	VS	4.34×10^{5}	−21.815±1.136	−46.077±0.134	−34.575±0.372	−33.131±0.011	−42.197±0.257	−47.406±0.054	−47.523±0.026
		ES-Chan	2.996	− 3.237±0.453	− 3.848±0.414	− 4.282±0.005	− 4.006±0.003	− 3.914±0.070	− 2.748±0.020	− 3.593±0.080
	336	VS	7.39×10^{5}	−25.920±1.943	−44.836±0.292	−31.710±1.160	−32.659±0.044	−41.809±0.575	−46.660±0.068	−46.953±0.026
		ES-Chan	2.976	− 3.340±0.293	− 3.268±0.127	− 3.738±0.052	− 3.478±0.016	− 3.100±0.177	− 2.138±0.038	− 2.896±0.023
	720	VS	1.54×10^{6}	−28.101±0.650	−45.115±0.201	−34.190±0.693	−33.321±0.008	−42.201±0.326	−47.200±0.070	−47.492±0.120
		ES-Chan	3.044	− 3.458±0.107	− 3.243±0.144	− 3.628±0.018	− 3.268±0.010	− 3.058±0.079	− 1.881±0.050	− 2.665±0.071
	Avg	VS	7.19×10^{5}	−23.072±1.020	−43.467±0.056	−32.640±0.306	−31.835±0.015	−40.111±0.209	−44.977±0.032	−45.227±0.020
		ES-Chan	2.911	− 2.913±0.069	− 2.930±0.083	− 3.488±0.014	− 3.103±0.003	− 2.826±0.097	− 1.625±0.022	− 2.456±0.024
1st count		VS		4	0	4	0	0	0	4
		ES-Chan		1	0	11	0	0	0	0
Avg. marginal & temporal		CRPS		+ 0.042±0.002	+ 0.046±0.112	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.669±0.003	+ 0.000±0.000
		NMAE		+ 0.030±0.003	+ 0.110±0.164	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.507±0.007	+ 0.000±0.000
		ES-Time		− 0.009±0.075	− 0.036±0.107	− 2.723±0.006	− 3.439±0.007	− 3.101±0.034	− 2.284±0.004	− 1.107±0.008

Table 23: **Experiment results for FMPatch, averaged over horizons.** Percent change vs. the uncoupled arm of the same settings; negative is better. ± is the standard deviation over 3 seeds (2026–2028); all settings of a cell are first averaged per seed. Red best, blue second. *Avg. marginal & temporal* averages all cells.

Dataset	Metric (%)	GeNCoup-Dyn (ours)	GeNCoup-Corr (ours)	COBASE	Schaake	SimSch.	GCA	IC
ETTh1	VS	− 6.696±0.407	−10.555±0.074	− 9.411±0.045	− 8.804±0.075	− 8.950±0.207	− 8.867±0.031	− 9.132±0.016
	ES-Chan	− 0.500±0.211	− 1.779±0.087	− 1.655±0.006	− 0.640±0.010	− 1.146±0.016	− 0.703±0.006	− 1.400±0.005
ETTh2	VS	+ 0.766±0.196	+ 1.915±0.215	+ 2.924±0.064	+ 3.697±0.043	+ 3.329±0.016	+ 4.484±0.090	+ 3.559±0.143
	ES-Chan	+ 0.749±0.370	− 0.098±0.224	− 0.023±0.024	+ 0.890±0.011	+ 1.254±0.074	+ 0.996±0.047	+ 0.279±0.057
ETTm1	VS	− 7.289±0.716	− 8.803±0.059	− 8.242±0.077	− 8.066±0.042	− 8.303±0.028	− 7.952±0.066	− 8.672±0.058
	ES-Chan	− 0.768±0.174	− 1.627±0.048	− 1.515±0.005	− 0.733±0.003	− 1.024±0.023	− 0.717±0.007	− 1.432±0.009
ETTm2	VS	− 0.226±0.392	+ 0.061±0.038	+ 0.314±0.024	+ 1.112±0.032	+ 0.338±0.165	+ 1.901±0.047	+ 0.770±0.016
	ES-Chan	− 0.301±0.044	− 0.738±0.033	− 0.683±0.018	+ 0.107±0.012	+ 0.415±0.057	+ 0.366±0.001	− 0.269±0.006
Weather	VS	− 9.433±0.094	− 8.734±0.072	− 7.979±0.024	− 7.227±0.033	− 7.735±0.025	− 7.067±0.061	− 8.251±0.042
	ES-Chan	− 0.640±0.092	− 1.148±0.074	− 1.125±0.010	− 0.765±0.008	− 1.050±0.008	− 0.364±0.018	− 1.231±0.015
ECL	VS	− 0.063±0.004	+ 4.324±0.039	+ 0.462±0.012	+ 1.186±0.002	+ 2.700±0.114	+ 5.749±0.007	+ 4.623±0.034
	ES-Chan	+ 0.035±0.010	+ 0.805±0.049	− 0.151±0.003	+ 0.152±0.002	+ 0.515±0.026	+ 1.702±0.011	+ 0.843±0.025
Traffic	VS	− 0.226±0.036	+ 2.937±0.030	+ 0.500±0.007	+ 1.964±0.001	+ 5.205±0.130	+ 4.465±0.009	–
	ES-Chan	+ 0.062±0.003	+ 2.139±0.021	+ 0.357±0.006	+ 0.670±0.001	+ 2.060±0.047	+ 3.170±0.005	–
Solar	VS	−45.279±0.747	−48.412±0.089	−38.456±0.489	−39.620±0.009	−46.184±0.008	−47.342±0.084	−47.711±0.095
	ES-Chan	− 9.289±0.084	−10.241±0.074	− 8.938±0.078	− 7.063±0.003	− 9.459±0.060	− 9.322±0.030	−10.130±0.010
1st count	VS	5	3	0	0	0	0	0
	ES-Chan	1	5	1	0	0	0	1
Avg. marginal & temporal	CRPS	+ 0.021±0.097	+ 0.017±0.054	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.759±0.002	+ 0.000±0.000
	NMAE	+ 0.006±0.102	+ 0.013±0.021	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.642±0.005	+ 0.000±0.000
	ES-Time	+ 0.017±0.056	+ 0.019±0.041	+ 2.469±0.001	+ 0.748±0.003	+ 0.904±0.024	+ 2.907±0.002	+ 2.806±0.001

Table 24: **Detailed experiment results for FMPatch.** Percent change vs. the uncoupled arm of the same settings; negative is better. ± is the standard deviation over 3 seeds (2026–2028); all settings of a cell are first averaged per seed. Red best, blue second. *Avg. marginal & temporal* averages all cells.

Dataset	h	Metric (%)	Uncoupled	GeNCoup-Dyn (ours)	GeNCoup-Corr (ours)	COBASE	Schaake	SimSch.	GCA	IC
ETTh1	96	VS	471.8	− 8.734±2.165	−11.859±0.357	−11.880±0.093	−11.517±0.120	−11.500±0.181	−11.465±0.095	−11.583±0.156
		ES-Chan	1.024	− 1.331±0.415	− 1.971±0.394	− 2.062±0.055	− 1.210±0.043	− 1.761±0.035	− 1.303±0.043	− 1.839±0.053
	192	VS	984.3	− 6.940±0.174	− 9.916±0.176	− 8.729±0.071	− 8.473±0.047	− 8.800±0.184	− 8.906±0.104	− 9.265±0.090
		ES-Chan	1.069	− 0.749±0.131	− 1.636±0.223	− 1.669±0.013	− 1.085±0.018	− 1.321±0.057	− 0.943±0.018	− 1.580±0.009
	336	VS	1893	− 5.200±0.609	− 8.794±0.266	− 7.382±0.038	− 7.446±0.054	− 6.940±0.119	− 7.141±0.041	− 7.738±0.028
		ES-Chan	1.161	− 0.019±0.205	− 1.460±0.092	− 1.309±0.032	− 0.426±0.033	− 0.824±0.023	− 0.318±0.039	− 1.146±0.036
	720	VS	4855	− 5.911±0.643	−11.652±0.194	− 9.654±0.152	− 7.780±0.145	− 8.558±0.382	− 7.954±0.133	− 7.943±0.083
		ES-Chan	1.281	+ 0.101±0.597	− 2.050±0.187	− 1.580±0.064	+ 0.161±0.028	− 0.676±0.067	− 0.249±0.040	− 1.036±0.043
	Avg	VS	2051	− 6.696±0.407	−10.555±0.074	− 9.411±0.045	− 8.804±0.075	− 8.950±0.207	− 8.867±0.031	− 9.132±0.016
		ES-Chan	1.134	− 0.500±0.211	− 1.779±0.087	− 1.655±0.006	− 0.640±0.010	− 1.146±0.016	− 0.703±0.006	− 1.400±0.005
ETTh2	96	VS	346.7	+ 0.406±1.377	+ 0.674±0.687	+ 0.850±0.129	+ 1.459±0.064	+ 0.460±0.278	+ 2.077±0.111	+ 0.626±0.255
		ES-Chan	1.058	+ 1.066±1.458	+ 0.259±0.431	− 0.134±0.066	+ 0.537±0.055	+ 1.411±0.204	+ 0.716±0.060	− 0.003±0.092
	192	VS	659.8	− 0.388±1.313	+ 0.160±0.157	+ 1.375±0.076	+ 1.533±0.067	+ 1.313±0.184	+ 2.877±0.115	+ 2.149±0.070
		ES-Chan	0.98	− 0.436±0.327	− 0.727±0.171	− 0.481±0.023	+ 0.439±0.007	+ 0.694±0.109	+ 0.828±0.094	+ 0.147±0.084
	336	VS	1435	− 1.332±0.898	+ 0.543±0.289	+ 1.540±0.080	+ 1.483±0.097	+ 1.275±0.092	+ 3.086±0.068	+ 2.480±0.056
		ES-Chan	1.157	+ 1.295±0.274	− 0.004±0.233	+ 0.353±0.061	+ 1.282±0.043	+ 1.217±0.026	+ 1.419±0.032	+ 0.686±0.036
	720	VS	5076	+ 4.379±1.172	+ 6.283±0.340	+ 7.931±0.058	+10.313±0.082	+10.269±0.053	+ 9.896±0.203	+ 8.983±0.328
		ES-Chan	1.582	+ 1.071±0.375	+ 0.080±0.491	+ 0.169±0.045	+ 1.301±0.029	+ 1.696±0.016	+ 1.021±0.028	+ 0.284±0.057
	Avg	VS	1879	+ 0.766±0.196	+ 1.915±0.215	+ 2.924±0.064	+ 3.697±0.043	+ 3.329±0.016	+ 4.484±0.090	+ 3.559±0.143
		ES-Chan	1.194	+ 0.749±0.370	− 0.098±0.224	− 0.023±0.024	+ 0.890±0.011	+ 1.254±0.074	+ 0.996±0.047	+ 0.279±0.057
1st count		VS		4	3	1	0	0	0	0
		ES-Chan		0	5	3	0	0	0	0
Avg. marginal & temporal		CRPS		− 0.032±0.112	− 0.025±0.021	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.776±0.005	+ 0.000±0.000
		NMAE		+ 0.001±0.079	− 0.017±0.029	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.594±0.009	+ 0.000±0.000
		ES-Time		+ 0.007±0.072	+ 0.028±0.056	+ 2.190±0.043	+ 0.780±0.012	+ 0.955±0.021	+ 2.664±0.040	+ 2.193±0.043

Table 25: **Detailed experiment results for FMPatch.** Percent change vs. the uncoupled arm of the same settings; negative is better. ± is the standard deviation over 3 seeds (2026–2028); all settings of a cell are first averaged per seed. Red best, blue second. *Avg. marginal & temporal* averages all cells.

Dataset	h	Metric (%)	Uncoupled	GeNCoup-Dyn (ours)	GeNCoup-Corr (ours)	COBASE	Schaake	SimSch.	GCA	IC
ETTm1	96	VS	360.5	− 8.612±1.500	− 9.145±0.228	− 8.615±0.118	− 8.636±0.101	− 8.651±0.113	− 8.436±0.089	− 8.994±0.122
		ES-Chan	0.8669	− 1.501±0.389	− 1.753±0.244	− 1.518±0.020	− 0.620±0.016	− 0.955±0.034	− 0.825±0.020	− 1.498±0.021
	192	VS	799.7	− 6.216±0.820	− 7.366±0.116	− 6.701±0.045	− 6.327±0.050	− 6.792±0.203	− 6.241±0.021	− 7.081±0.036
		ES-Chan	0.9531	− 0.322±0.334	− 1.454±0.018	− 1.264±0.034	− 0.796±0.025	− 0.860±0.057	− 0.514±0.025	− 1.251±0.026
	336	VS	1507	− 7.856±1.565	− 9.725±0.255	− 9.544±0.096	− 9.448±0.076	− 9.751±0.200	− 9.262±0.109	− 9.880±0.097
		ES-Chan	1.009	− 1.037±0.186	− 1.649±0.254	− 1.739±0.012	− 0.931±0.006	− 1.289±0.021	− 0.953±0.007	− 1.641±0.005
	720	VS	3789	− 6.470±1.925	− 8.977±0.117	− 8.106±0.106	− 7.853±0.065	− 8.021±0.456	− 7.869±0.091	− 8.734±0.052
		ES-Chan	1.113	− 0.212±0.607	− 1.652±0.045	− 1.539±0.016	− 0.585±0.015	− 0.991±0.022	− 0.575±0.036	− 1.337±0.025
	Avg	VS	1614	− 7.289±0.716	− 8.803±0.059	− 8.242±0.077	− 8.066±0.042	− 8.303±0.028	− 7.952±0.066	− 8.672±0.058
		ES-Chan	0.9854	− 0.768±0.174	− 1.627±0.048	− 1.515±0.005	− 0.733±0.003	− 1.024±0.023	− 0.717±0.007	− 1.432±0.009
ETTm2	96	VS	198.4	+ 0.373±0.198	+ 0.108±0.516	+ 0.375±0.025	+ 0.954±0.055	+ 0.109±0.077	+ 1.605±0.054	+ 0.610±0.018
		ES-Chan	0.668	− 0.261±0.083	− 0.573±0.233	− 0.553±0.022	+ 0.239±0.028	+ 0.438±0.332	+ 0.720±0.046	− 0.042±0.024
	192	VS	488.6	− 0.230±0.229	+ 0.186±0.160	+ 0.781±0.025	+ 1.390±0.031	+ 0.468±0.080	+ 2.043±0.045	+ 1.069±0.027
		ES-Chan	0.7767	− 0.197±0.194	− 0.827±0.046	− 0.730±0.023	+ 0.074±0.021	+ 0.424±0.029	+ 0.517±0.042	− 0.144±0.050
	336	VS	1055	+ 0.725±1.873	+ 0.163±0.368	+ 0.305±0.026	+ 1.464±0.048	+ 0.386±0.406	+ 2.138±0.067	+ 0.959±0.077
		ES-Chan	0.9129	− 0.473±0.597	− 0.899±0.178	− 0.863±0.014	− 0.033±0.025	+ 0.691±0.137	+ 0.025±0.046	− 0.489±0.050
	720	VS	2777	− 1.773±0.674	− 0.215±0.284	− 0.206±0.155	+ 0.639±0.123	+ 0.391±0.113	+ 1.821±0.133	+ 0.442±0.100
		ES-Chan	1.051	− 0.271±0.555	− 0.654±0.197	− 0.587±0.068	+ 0.148±0.049	+ 0.107±0.147	+ 0.202±0.053	− 0.403±0.053
	Avg	VS	1130	− 0.226±0.392	+ 0.061±0.038	+ 0.314±0.024	+ 1.112±0.032	+ 0.338±0.165	+ 1.901±0.047	+ 0.770±0.016
		ES-Chan	0.8521	− 0.301±0.044	− 0.738±0.033	− 0.683±0.018	+ 0.107±0.012	+ 0.415±0.057	+ 0.366±0.001	− 0.269±0.006
Weather	96	VS	2717	− 8.402±0.107	− 8.054±0.082	− 7.134±0.046	− 6.661±0.068	− 7.033±0.086	− 6.405±0.119	− 7.458±0.070
		ES-Chan	1.217	− 0.926±0.086	− 1.336±0.154	− 1.292±0.003	− 0.937±0.022	− 1.302±0.031	− 0.667±0.029	− 1.454±0.035
	192	VS	6799	− 7.186±0.370	− 5.893±0.216	− 5.687±0.014	− 5.026±0.038	− 5.447±0.042	− 4.365±0.034	− 5.680±0.029
		ES-Chan	1.402	− 0.489±0.148	− 0.845±0.127	− 0.896±0.020	− 0.551±0.016	− 0.795±0.025	+ 0.012±0.024	− 0.935±0.012
	336	VS	1.28×10^4	− 8.835±0.129	− 8.342±0.066	− 7.702±0.045	− 6.855±0.036	− 7.413±0.052	− 6.731±0.029	− 7.854±0.033
		ES-Chan	1.499	− 0.293±0.242	− 1.094±0.043	− 1.065±0.022	− 0.702±0.023	− 0.966±0.047	− 0.296±0.024	− 1.179±0.028
	720	VS	3.27×10^4	−13.311±0.252	−12.648±0.105	−11.391±0.040	−10.368±0.047	−11.049±0.049	−10.767±0.086	−12.011±0.081
		ES-Chan	1.715	− 0.852±0.187	− 1.317±0.049	− 1.248±0.011	− 0.872±0.006	− 1.139±0.049	− 0.504±0.016	− 1.356±0.007
	Avg	VS	1.37×10^4	− 9.433±0.094	− 8.734±0.072	− 7.979±0.024	− 7.227±0.033	− 7.735±0.025	− 7.067±0.061	− 8.251±0.042
		ES-Chan	1.458	− 0.640±0.092	− 1.148±0.074	− 1.125±0.010	− 0.765±0.008	− 1.050±0.008	− 0.364±0.018	− 1.231±0.015
1st count		VS		6	5	0	0	0	0	1
		ES-Chan		0	7	1	0	0	0	4
Avg. marginal & temporal		CRPS		− 0.003±0.146	− 0.036±0.031	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.767±0.001	+ 0.000±0.000
		NMAE		− 0.040±0.185	− 0.036±0.039	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.736±0.007	+ 0.000±0.000
		ES-Time		+ 0.009±0.091	+ 0.004±0.086	+ 3.378±0.025	+ 1.034±0.009	+ 0.887±0.011	+ 3.484±0.023	+ 3.378±0.025

Table 26: **Detailed experiment results for FMPatch.** Percent change vs. the uncoupled arm of the same settings; negative is better. ± is the standard deviation over 3 seeds (2026–2028); all settings of a cell are first averaged per seed. Red best, blue second. *Avg. marginal & temporal* averages all cells.

Dataset	h	Metric (%)	Uncoupled	GeNCoup-Dyn (ours)	GeNCoup-Corr (ours)	COBASE	Schaake	SimSch.	GCA	IC
ECL	96	VS	6.36×10^{5}	− 0.160±0.078	+ 3.161±0.074	+ 0.080±0.028	+ 0.735±0.003	+ 1.934±0.177	+ 4.641±0.037	+ 3.486±0.051
		ES-Chan	5.094	− 0.053±0.061	+ 0.664±0.097	− 0.182±0.007	+ 0.123±0.005	+ 0.467±0.028	+ 1.608±0.032	+ 0.721±0.042
	192	VS	1.32×10^{6}	+ 0.027±0.041	+ 4.360±0.112	+ 0.478±0.042	+ 1.335±0.001	+ 2.817±0.218	+ 5.728±0.014	+ 4.621±0.051
		ES-Chan	5.291	+ 0.065±0.039	+ 0.849±0.049	− 0.149±0.006	+ 0.179±0.002	+ 0.522±0.047	+ 1.709±0.008	+ 0.848±0.025
	336	VS	2.49×10^{6}	− 0.044±0.104	+ 4.646±0.097	+ 0.626±0.056	+ 1.328±0.004	+ 2.808±0.117	+ 6.064±0.007	+ 4.919±0.023
		ES-Chan	5.67	+ 0.038±0.076	+ 0.840±0.128	− 0.174±0.003	+ 0.132±0.002	+ 0.458±0.027	+ 1.700±0.010	+ 0.840±0.025
	720	VS	6.03×10^{6}	− 0.073±0.047	+ 5.128±0.037	+ 0.663±0.011	+ 1.348±0.003	+ 3.241±0.272	+ 6.561±0.009	+ 5.466±0.042
		ES-Chan	6.117	+ 0.090±0.060	+ 0.867±0.023	− 0.097±0.002	+ 0.174±0.000	+ 0.613±0.088	+ 1.790±0.005	+ 0.963±0.050
	Avg	VS	2.62×10^{6}	− 0.063±0.004	+ 4.324±0.039	+ 0.462±0.012	+ 1.186±0.002	+ 2.700±0.114	+ 5.749±0.007	+ 4.623±0.034
		ES-Chan	5.543	+ 0.035±0.010	+ 0.805±0.049	− 0.151±0.003	+ 0.152±0.002	+ 0.515±0.026	+ 1.702±0.011	+ 0.843±0.025
Traffic	96	VS	7.53×10^{6}	− 0.282±0.160	+ 2.191±0.117	+ 0.037±0.023	+ 1.251±0.003	+ 4.423±0.259	+ 3.714±0.012	–
		ES-Chan	14.15	+ 0.069±0.038	+ 2.135±0.122	+ 0.305±0.022	+ 0.652±0.002	+ 2.004±0.155	+ 3.178±0.015	–
	192	VS	1.50×10^{7}	− 0.295±0.020	+ 2.566±0.067	+ 0.299±0.027	+ 1.798±0.004	+ 4.649±0.065	+ 3.978±0.044	–
		ES-Chan	13.72	+ 0.058±0.022	+ 2.181±0.045	+ 0.348±0.019	+ 0.672±0.001	+ 2.050±0.015	+ 3.159±0.025	–
	336	VS	2.66×10^{7}	− 0.102±0.017	+ 3.719±0.109	+ 0.973±0.060	+ 2.621±0.003	+ 5.943±0.116	+ 5.212±0.027	–
		ES-Chan	14	+ 0.098±0.020	+ 2.239±0.116	+ 0.378±0.032	+ 0.685±0.001	+ 2.082±0.047	+ 3.256±0.016	–
	720	VS	5.99×10^{7}	− 0.223±0.034	+ 3.273±0.026	+ 0.690±0.029	+ 2.186±0.002	+ 5.804±0.247	+ 4.955±0.011	–
		ES-Chan	14.22	+ 0.024±0.035	+ 2.000±0.018	+ 0.397±0.017	+ 0.671±0.001	+ 2.103±0.092	+ 3.088±0.005	–
	Avg	VS	2.73×10^{7}	− 0.226±0.036	+ 2.937±0.030	+ 0.500±0.007	+ 1.964±0.001	+ 5.205±0.130	+ 4.465±0.009	–
		ES-Chan	14.02	+ 0.062±0.003	+ 2.139±0.021	+ 0.357±0.006	+ 0.670±0.001	+ 2.060±0.047	+ 3.170±0.005	–
Solar	96	VS	1.55×10^{5}	−43.880±1.810	−48.493±0.083	−37.264±1.132	−39.017±0.019	−45.441±0.069	−47.135±0.020	−47.561±0.058
		ES-Chan	2.402	− 8.635±0.424	− 9.645±0.434	− 8.477±0.174	− 6.616±0.005	− 8.807±0.114	− 8.569±0.040	− 9.397±0.012
	192	VS	2.96×10^{5}	−45.473±1.764	−49.587±0.113	−38.859±0.871	−40.207±0.016	−47.131±0.191	−48.499±0.140	−48.829±0.115
		ES-Chan	2.418	− 9.704±0.088	−10.876±0.326	− 9.421±0.133	− 7.323±0.002	− 9.979±0.058	−10.082±0.048	−10.946±0.022
	336	VS	5.19×10^{5}	−46.484±1.371	−47.157±0.147	−39.244±0.632	−39.636±0.031	−45.981±0.169	−46.391±0.162	−46.744±0.165
		ES-Chan	2.487	− 9.529±0.624	−10.203±0.157	− 8.915±0.114	− 7.251±0.005	− 9.591±0.026	− 9.315±0.025	−10.047±0.033
	Avg	VS	3.23×10^{5}	−45.279±0.747	−48.412±0.089	−38.456±0.489	−39.620±0.009	−46.184±0.008	−47.342±0.084	−47.711±0.095
		ES-Chan	2.436	− 9.289±0.084	−10.241±0.074	− 8.938±0.078	− 7.063±0.003	− 9.459±0.060	− 9.322±0.030	−10.130±0.010
1st count		VS		8	3	0	0	0	0	0
		ES-Chan		4	2	4	0	0	0	1
Avg. marginal & temporal		CRPS		+ 0.085±0.158	+ 0.105±0.194	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.740±0.001	+ 0.000±0.000
		NMAE		+ 0.060±0.118	+ 0.089±0.111	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.575±0.003	+ 0.000±0.000
		ES-Time		+ 0.034±0.048	+ 0.030±0.065	+ 1.680±0.002	+ 0.412±0.000	+ 0.887±0.055	+ 2.454±0.001	+ 2.525±0.011

Table 27: **Experiment results for DiffPatch, averaged over horizons.** Percent change vs. the uncoupled arm of the same settings; negative is better. ± is the standard deviation over 3 seeds (2026–2028); all settings of a cell are first averaged per seed. Red best, blue second. *Avg. marginal & temporal* averages all cells.

Dataset	Metric (%)	GeNCoup-Dyn (ours)	GeNCoup-Corr (ours)	COBASE	Schaake	SimSch.	GCA	IC
ETTh1	VS	−22.071± 1.960	−23.086± 0.111	−22.013± 0.089	−23.406± 0.035	−18.404± 0.361	−23.854± 0.091	−21.327± 0.145
	ES-Chan	− 3.251± 0.250	− 3.250± 0.097	− 2.857± 0.010	− 2.639± 0.057	− 2.492± 0.028	− 3.196± 0.016	− 2.390± 0.020
ETTh2	VS	−28.044± 0.274	−18.908± 0.401	−20.187± 0.024	−16.744± 0.093	−12.078± 0.572	−19.461± 0.187	−16.119± 0.157
	ES-Chan	− 6.354± 0.147	− 4.735± 0.409	− 4.195± 0.075	− 6.255± 0.063	− 4.469± 0.112	− 5.466± 0.050	− 4.623± 0.055
ETTm1	VS	−25.367± 1.203	−23.570± 0.033	−22.499± 0.040	−24.095± 0.059	−19.867± 0.336	−24.909± 0.105	−22.719± 0.166
	ES-Chan	− 4.192± 0.175	− 3.354± 0.046	− 3.128± 0.009	− 2.011± 0.011	− 2.323± 0.102	− 3.502± 0.007	− 2.987± 0.010
ETTm2	VS	−27.500± 1.764	−17.798± 0.204	−22.149± 0.141	−19.560± 0.176	−13.980± 1.026	−21.884± 0.414	−18.254± 0.454
	ES-Chan	− 6.135± 0.174	− 3.738± 0.083	− 4.295± 0.060	− 5.711± 0.086	− 4.321± 0.214	− 5.268± 0.039	− 4.492± 0.031
Weather	VS	−47.254± 3.570	−22.815± 0.035	−15.712± 0.029	−19.105± 0.018	−20.390± 0.205	−26.988± 0.061	−21.898± 0.123
	ES-Chan	−15.016± 0.764	− 7.707± 0.053	− 5.395± 0.027	− 6.606± 0.018	− 8.956± 0.139	− 9.849± 0.074	− 7.945± 0.157
ECL	VS	− 0.188± 0.140	−20.249± 0.110	−11.096± 0.185	− 9.929± 0.003	−16.259± 0.212	−19.404± 0.014	−19.804± 0.019
	ES-Chan	− 0.160± 0.061	− 6.358± 0.053	− 2.679± 0.060	− 2.425± 0.012	− 3.901± 0.177	− 6.535± 0.022	− 6.115± 0.006
Traffic	VS	− 0.047± 0.079	−15.817± 0.044	−11.041± 0.033	−12.171± 0.011	−18.795± 0.035	−18.419± 0.027	–
	ES-Chan	− 0.001± 0.253	− 8.025± 0.166	− 5.980± 0.010	− 3.284± 0.007	− 6.477± 0.065	− 8.970± 0.043	–
Solar	VS	−70.396± 0.470	−66.544± 0.064	−46.571± 0.646	−29.125± 0.007	−49.333± 0.212	−66.034± 0.185	−65.847± 0.199
	ES-Chan	−17.932± 0.072	−17.740± 0.037	−14.068± 0.119	− 6.491± 0.003	−13.908± 0.124	−16.671± 0.018	−17.469± 0.027
1st count	VS	5	1	0	0	1	1	0
	ES-Chan	6	0	0	0	0	2	0
Avg. marginal & temporal	CRPS	+ 0.038± 0.136	+ 0.049± 0.061	+ 0.000± 0.000	+ 0.000± 0.000	+ 0.000± 0.000	+ 1.190± 0.000	+ 0.000± 0.000
	NMAE	+ 0.066± 0.205	+ 0.020± 0.039	+ 0.000± 0.000	+ 0.000± 0.000	+ 0.000± 0.000	+ 2.527± 0.014	+ 0.000± 0.000
	ES-Time	+ 0.058± 0.087	+ 0.011± 0.037	+15.388± 0.091	+ 0.857± 0.015	+ 2.966± 0.093	+10.629± 0.058	+16.101± 0.110

Table 28: **Detailed experiment results for DiffPatch.** Percent change vs. the uncoupled arm of the same settings; negative is better. ± is the standard deviation over 3 seeds (2026–2028); all settings of a cell are first averaged per seed. Red best, blue second. *Avg. marginal & temporal* averages all cells.

Dataset	h	Metric (%)	Uncoupled	GeNCoup-Dyn (ours)	GeNCoup-Corr (ours)	COBASE	Schaake	SimSch.	GCA	IC
ETTh1	96	VS	591.1	−23.734±1.641	−19.463±0.156	−18.331±0.083	−18.843±0.040	−15.622±0.274	−19.177±0.326	−17.657±0.433
		ES-Chan	1.1	− 3.355±0.128	− 2.716±0.397	− 2.765±0.020	− 1.650±0.028	− 2.150±0.081	− 2.573±0.032	− 2.438±0.008
	192	VS	1484	−29.493±1.354	−20.827±0.494	−19.959±0.038	−21.559±0.160	−18.944±0.433	−21.780±0.250	−19.361±0.275
		ES-Chan	1.253	− 4.373±0.407	− 3.297±0.243	− 2.774±0.026	− 3.453±0.077	− 2.699±0.112	− 3.210±0.032	− 2.422±0.024
	336	VS	2976	−22.947±6.734	−24.144±0.286	−22.958±0.075	−24.569±0.157	−18.977±0.679	−25.121±0.048	−22.346±0.032
		ES-Chan	1.31	− 3.217±0.213	− 3.264±0.187	− 2.848±0.023	− 2.813±0.094	− 2.504±0.077	− 3.260±0.047	− 2.301±0.038
	720	VS	1.04×10^4	−12.111±1.326	−27.908±0.220	−26.804±0.347	−28.654±0.152	−20.071±0.814	−29.339±0.260	−25.942±0.400
		ES-Chan	1.632	− 2.060±0.350	− 3.723±0.289	− 3.039±0.009	− 2.638±0.079	− 2.614±0.181	− 3.739±0.019	− 2.400±0.034
	Avg	VS	3869	−22.071±1.960	−23.086±0.111	−22.013±0.089	−23.406±0.035	−18.404±0.361	−23.854±0.091	−21.327±0.145
		ES-Chan	1.324	− 3.251±0.250	− 3.250±0.097	− 2.857±0.010	− 2.639±0.057	− 2.492±0.028	− 3.196±0.016	− 2.390±0.020
ETTh2	96	VS	542.5	−36.740±0.829	−16.079±0.512	−19.106±0.243	−17.812±0.305	−11.982±0.839	−19.768±0.348	−17.617±0.300
		ES-Chan	1.325	−10.403±0.251	− 4.603±0.040	− 4.816±0.177	− 6.386±0.076	− 4.380±0.128	− 5.605±0.131	− 4.963±0.261
	192	VS	1844	−51.398±0.558	−20.887±0.496	−21.609±0.322	−20.888±0.047	−15.939±1.178	−22.836±0.216	−18.953±0.029
		ES-Chan	1.749	−11.627±0.270	− 6.139±0.587	− 4.685±0.057	− 7.476±0.068	− 5.685±0.410	− 6.328±0.094	− 5.491±0.256
	336	VS	3318	−19.630±1.745	−19.549±0.793	−23.281±0.251	−20.003±0.239	−11.592±0.870	−21.692±0.185	−18.450±0.455
		ES-Chan	1.798	− 2.418±0.636	− 4.642±0.724	− 3.983±0.060	− 6.672±0.155	− 4.152±0.097	− 5.181±0.135	− 4.714±0.162
	720	VS	7468	− 4.410±3.232	−19.117±0.090	−16.750±0.236	− 8.273±0.310	− 8.800±0.897	−13.548±0.120	− 9.458±0.050
		ES-Chan	1.986	− 0.968±0.595	− 3.556±0.481	− 3.294±0.010	− 4.487±0.085	− 3.658±0.177	− 4.749±0.045	− 3.322±0.138
	Avg	VS	3293	−28.044±0.274	−18.908±0.401	−20.187±0.024	−16.744±0.093	−12.078±0.572	−19.461±0.187	−16.119±0.157
		ES-Chan	1.714	− 6.354±0.147	− 4.735±0.409	− 4.195±0.075	− 6.255±0.063	− 4.469±0.112	− 5.466±0.050	− 4.623±0.055
1st count		VS		4	1	1	0	0	2	0
		ES-Chan		4	1	0	1	0	2	0
Avg. marginal & temporal		CRPS		− 0.138±0.322	− 0.051±0.216	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 1.195±0.010	+ 0.000±0.000
		NMAE		− 0.144±0.366	− 0.091±0.178	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 2.686±0.019	+ 0.000±0.000
		ES-Time		+ 0.024±0.270	− 0.085±0.140	+13.907±0.281	− 0.414±0.058	+ 3.811±0.153	+10.009±0.185	+13.918±0.279

Table 29: **Detailed experiment results for DiffPatch.** Percent change vs. the uncoupled arm of the same settings; negative is better. ± is the standard deviation over 3 seeds (2026–2028); all settings of a cell are first averaged per seed. Red best, blue second. *Avg. marginal & temporal* averages all cells.

Dataset	h	Metric (%)	Uncoupled	GeNCoup-Dyn (ours)	GeNCoup-Corr (ours)	COBASE	Schaake	SimSch.	GCA	IC
ETTm1	96	VS	540.4	−31.235± 0.414	−21.880± 0.297	−19.833± 0.214	−21.820± 0.041	−19.048± 0.428	−22.552± 0.209	−20.792± 0.213
		ES-Chan	1.037	− 4.907± 0.424	− 3.321± 0.145	− 2.838± 0.020	− 2.745± 0.069	− 2.974± 0.230	− 3.415± 0.051	− 2.845± 0.057
	192	VS	1258	−30.716± 2.190	−22.606± 0.061	−20.944± 0.089	−22.293± 0.101	−19.556± 1.236	−23.318± 0.101	−21.158± 0.118
		ES-Chan	1.147	− 4.795± 0.266	− 3.437± 0.079	− 2.928± 0.022	− 1.858± 0.044	− 2.426± 0.378	− 3.382± 0.024	− 2.844± 0.037
	336	VS	2999	−35.848± 1.418	−25.112± 0.162	−24.101± 0.352	−25.337± 0.132	−18.820± 0.970	−25.850± 0.333	−23.689± 0.332
		ES-Chan	1.446	− 5.815± 0.399	− 3.638± 0.207	− 3.720± 0.052	− 0.886± 0.029	− 1.652± 0.305	− 3.622± 0.082	− 3.299± 0.082
	720	VS	7948	− 3.668± 5.459	−24.682± 0.087	−25.118± 0.359	−26.929± 0.026	−22.043± 0.282	−27.914± 0.558	−25.239± 0.729
		ES-Chan	1.472	− 1.253± 0.150	− 3.021± 0.038	− 3.025± 0.040	− 2.556± 0.024	− 2.238± 0.056	− 3.591± 0.039	− 2.960± 0.076
	Avg	VS	3186	−25.367± 1.203	−23.570± 0.033	−22.499± 0.040	−24.095± 0.059	−19.867± 0.336	−24.909± 0.105	−22.719± 0.166
		ES-Chan	1.275	− 4.192± 0.175	− 3.354± 0.046	− 3.128± 0.009	− 2.011± 0.011	− 2.323± 0.102	− 3.502± 0.007	− 2.987± 0.010
ETTm2	96	VS	458.7	−32.586± 0.111	−13.598± 0.522	−17.597± 0.196	−16.967± 0.551	−11.977± 2.054	−18.315± 0.172	−15.828± 0.100
		ES-Chan	1.122	− 8.586± 0.128	− 3.993± 0.134	− 4.341± 0.062	− 5.503± 0.084	− 4.634± 0.927	− 4.876± 0.107	− 4.344± 0.114
	192	VS	1288	−51.192± 0.980	−19.177± 0.477	−25.824± 0.300	−23.349± 0.230	−14.755± 1.432	−25.858± 1.058	−22.789± 1.147
		ES-Chan	1.411	−10.222± 0.191	− 3.989± 0.266	− 4.743± 0.143	− 5.653± 0.109	− 3.863± 0.423	− 5.460± 0.055	− 4.597± 0.080
	336	VS	3036	−15.880± 7.387	−18.884± 0.051	−24.280± 0.433	−22.385± 0.160	−14.939± 1.029	−24.603± 0.781	−20.478± 0.746
		ES-Chan	1.614	− 3.415± 0.310	− 3.717± 0.127	− 4.440± 0.086	− 6.717± 0.092	− 4.819± 0.351	− 5.608± 0.072	− 4.988± 0.115
	720	VS	8674	−10.342± 3.120	−19.535± 0.039	−20.895± 0.403	−15.540± 0.242	−14.249± 1.239	−18.761± 0.403	−13.920± 0.534
		ES-Chan	2.067	− 2.316± 0.137	− 3.255± 0.200	− 3.655± 0.094	− 4.969± 0.093	− 3.967± 0.602	− 5.126± 0.081	− 4.038± 0.064
	Avg	VS	3364	−27.500± 1.764	−17.798± 0.204	−22.149± 0.141	−19.560± 0.176	−13.980± 1.026	−21.884± 0.414	−18.254± 0.454
		ES-Chan	1.553	− 6.135± 0.174	− 3.738± 0.083	− 4.295± 0.060	− 5.711± 0.086	− 4.321± 0.214	− 5.268± 0.039	− 4.492± 0.031
Weather	96	VS	7595	−53.273± 0.666	−20.883± 0.383	−13.692± 0.210	−17.235± 0.048	−17.958± 0.405	−25.747± 0.151	−19.681± 0.302
		ES-Chan	2.083	−22.812± 0.395	− 9.818± 0.129	− 6.388± 0.160	− 8.699± 0.063	−10.680± 0.275	−12.594± 0.240	−10.041± 0.515
	192	VS	1.56×10^4	−56.846± 0.226	−23.034± 0.116	−16.429± 0.159	−19.280± 0.079	−21.122± 0.144	−27.538± 0.135	−22.308± 0.184
		ES-Chan	2.129	−19.743± 0.241	− 8.579± 0.086	− 6.260± 0.068	− 7.607± 0.016	−10.365± 0.056	−11.198± 0.074	− 9.007± 0.034
	336	VS	3.15×10^4	−54.431± 0.742	−23.739± 0.274	−17.884± 0.114	−19.922± 0.071	−20.966± 0.095	−27.080± 0.091	−22.934± 0.127
		ES-Chan	2.254	−11.986± 0.159	− 5.436± 0.026	− 4.289± 0.040	− 4.474± 0.008	− 6.293± 0.138	− 6.729± 0.006	− 5.502± 0.024
	720	VS	1.48×10^5	−24.467±14.491	−23.605± 0.438	−14.843± 0.190	−19.982± 0.024	−21.513± 0.400	−27.587± 0.136	−22.667± 0.146
		ES-Chan	3.179	− 5.521± 3.300	− 6.995± 0.241	− 4.641± 0.091	− 5.646± 0.024	− 8.483± 0.147	− 8.877± 0.049	− 7.230± 0.121
	Avg	VS	5.08×10^4	−47.254± 3.570	−22.815± 0.035	−15.712± 0.029	−19.105± 0.018	−20.390± 0.205	−26.988± 0.061	−21.898± 0.123
		ES-Chan	2.411	−15.016± 0.764	− 7.707± 0.053	− 5.395± 0.027	− 6.606± 0.018	− 8.956± 0.139	− 9.849± 0.074	− 7.945± 0.157
1st count		VS		8	0	1	0	0	3	0
		ES-Chan		8	0	0	1	0	3	0
Avg. marginal & temporal		CRPS		+ 0.144± 0.081	+ 0.086± 0.119	+ 0.000± 0.000	+ 0.000± 0.000	+ 0.000± 0.000	+ 1.244± 0.005	+ 0.000± 0.000
		NMAE		+ 0.232± 0.219	+ 0.059± 0.267	+ 0.000± 0.000	+ 0.000± 0.000	+ 0.000± 0.000	+ 3.411± 0.026	+ 0.000± 0.000
		ES-Time		+ 0.146± 0.130	+ 0.110± 0.031	+24.229± 0.102	+ 0.014± 0.012	+ 2.978± 0.133	+17.539± 0.070	+24.230± 0.106

Table 30: **Detailed experiment results for DiffPatch.** Percent change vs. the uncoupled arm of the same settings; negative is better. ± is the standard deviation over 3 seeds (2026–2028); all settings of a cell are first averaged per seed. Red best, blue second. *Avg. marginal & temporal* averages all cells.

Dataset	h	Metric (%)	Uncoupled	GeNCoup-Dyn (ours)	GeNCoup-Corr (ours)	COBASE	Schaake	SimSch.	GCA	IC
ECL	96	VS	8.82×10^{5}	− 0.498±0.365	−15.903±0.046	− 7.990±0.126	− 7.805±0.006	−12.316±0.451	−14.799±0.078	−15.287±0.062
		ES-Chan	6.464	− 0.422±0.082	− 7.200±0.075	− 2.597±0.032	− 2.614±0.022	− 3.972±0.118	− 6.793±0.055	− 6.488±0.041
	192	VS	1.86×10^{6}	− 0.206±0.037	−13.319±0.204	− 7.899±0.145	− 7.422±0.004	−10.931±0.128	−12.154±0.009	−13.085±0.029
		ES-Chan	6.862	− 0.116±0.078	− 4.481±0.179	− 1.845±0.062	− 1.702±0.013	− 2.167±0.192	− 4.265±0.030	− 4.039±0.024
	336	VS	3.77×10^{6}	− 0.103±0.141	−22.085±0.058	−13.586±0.517	−10.721±0.012	−17.777±0.297	−21.369±0.052	−21.718±0.042
		ES-Chan	7.25	− 0.140±0.031	− 7.687±0.080	− 3.804±0.162	− 3.208±0.017	− 5.317±0.319	− 8.261±0.043	− 7.639±0.035
	720	VS	9.78×10^{6}	+ 0.056±0.074	−29.690±0.160	−14.909±0.247	−13.769±0.003	−24.012±0.623	−29.297±0.023	−29.128±0.053
		ES-Chan	7.706	+ 0.037±0.184	− 6.065±0.016	− 2.469±0.027	− 2.176±0.007	− 4.150±0.283	− 6.823±0.012	− 6.293±0.087
	Avg	VS	4.07×10^{6}	− 0.188±0.140	−20.249±0.110	−11.096±0.185	− 9.929±0.003	−16.259±0.212	−19.404±0.014	−19.804±0.019
		ES-Chan	7.071	− 0.160±0.061	− 6.358±0.053	− 2.679±0.060	− 2.425±0.012	− 3.901±0.177	− 6.535±0.022	− 6.115±0.006
Traffic	96	VS	1.34×10^{7}	− 0.222±0.317	− 9.560±0.106	− 6.307±0.116	− 6.517±0.018	− 9.045±0.039	− 9.805±0.068	–
		ES-Chan	28.26	− 0.271±0.664	− 7.078±0.163	− 4.735±0.030	− 2.208±0.025	− 4.846±0.189	− 6.204±0.131	–
	192	VS	3.67×10^{7}	+ 0.088±0.108	−11.978±0.118	− 7.729±0.021	− 9.343±0.017	−13.712±0.072	−13.937±0.058	–
		ES-Chan	40.28	+ 0.169±0.371	− 7.562±0.599	− 6.365±0.045	− 3.407±0.015	− 7.713±0.189	− 8.929±0.053	–
	336	VS	6.34×10^{7}	− 0.182±0.217	−16.218±0.205	−12.128±0.080	−12.633±0.010	−18.861±0.129	−18.789±0.076	–
		ES-Chan	31.35	− 0.240±0.508	− 7.632±0.451	− 5.384±0.053	− 3.074±0.017	− 4.982±0.149	− 7.874±0.095	–
	720	VS	1.87×10^{8}	+ 0.130±0.174	−25.511±0.358	−18.000±0.128	−20.188±0.008	−33.563±0.242	−31.146±0.058	–
		ES-Chan	32.43	+ 0.337±0.386	− 9.826±0.590	− 7.433±0.031	− 4.447±0.008	− 8.369±0.068	−12.872±0.015	–
	Avg	VS	7.51×10^{7}	− 0.047±0.079	−15.817±0.044	−11.041±0.033	−12.171±0.011	−18.795±0.035	−18.419±0.027	–
		ES-Chan	33.08	− 0.001±0.253	− 8.025±0.166	− 5.980±0.010	− 3.284±0.007	− 6.477±0.065	− 8.970±0.043	–
Solar	96	VS	2.80×10^{5}	−65.257±0.364	−63.355±0.075	−44.261±1.361	−28.815±0.019	−46.701±1.099	−61.974±0.258	−61.743±0.297
		ES-Chan	3.114	−13.149±0.355	−13.490±0.199	−10.676±0.233	− 4.682±0.001	−10.633±0.238	−12.144±0.007	−13.079±0.022
	192	VS	7.50×10^{5}	−68.151±1.042	−66.254±0.066	−46.417±1.162	−30.371±0.012	−50.262±0.778	−65.455±0.458	−65.297±0.519
		ES-Chan	3.426	−19.032±0.220	−18.858±0.249	−15.042±0.244	− 7.162±0.003	−14.763±0.224	−17.891±0.034	−18.658±0.070
	336	VS	1.63×10^{6}	−73.294±0.520	−68.704±0.100	−47.687±0.384	−30.514±0.008	−52.468±0.187	−68.172±0.335	−68.007±0.369
		ES-Chan	3.905	−19.707±0.191	−19.421±0.219	−15.274±0.080	− 7.339±0.006	−15.649±0.271	−18.434±0.037	−19.164±0.031
	720	VS	3.91×10^{6}	−74.882±0.252	−67.863±0.137	−47.918±0.117	−26.798±0.007	−47.900±0.278	−68.536±0.257	−68.343±0.281
		ES-Chan	3.918	−19.840±0.089	−19.189±0.119	−15.282±0.035	− 6.781±0.007	−14.586±0.096	−18.215±0.039	−18.974±0.030
	Avg	VS	1.64×10^{6}	−70.396±0.470	−66.544±0.064	−46.571±0.646	−29.125±0.007	−49.333±0.212	−66.034±0.185	−65.847±0.199
		ES-Chan	3.591	−17.932±0.072	−17.740±0.037	−14.068±0.119	− 6.491±0.003	−13.908±0.124	−16.671±0.018	−17.469±0.027
1st count		VS		4	4	0	0	2	2	0
		ES-Chan		3	4	0	0	0	5	0
Avg. marginal & temporal		CRPS		+ 0.051±0.106	+ 0.077±0.161	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 1.134±0.001	+ 0.000±0.000
		NMAE		+ 0.040±0.135	+ 0.055±0.134	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 1.537±0.004	+ 0.000±0.000
		ES-Time		− 0.008±0.019	− 0.025±0.039	+ 7.533±0.017	+ 2.548±0.011	+ 2.391±0.094	+ 4.132±0.009	+ 6.091±0.002

Table 31: **Experiment results for Sundial, averaged over horizons.** Percent change vs. the uncoupled arm of the same settings; negative is better. ± is the standard deviation over 3 seeds (2026–2028); all settings of a cell are first averaged per seed. Red best, blue second. *Avg. marginal & temporal* averages all cells.

Dataset	Metric (%)	GeNCoup-Dyn (ours)	GeNCoup-Corr (ours)	COBASE	Schaake	SimSch.	GCA	IC
ETTh1	VS	− 2.542±0.180	− 1.530±0.023	− 0.624±0.031	− 0.592±0.016	− 1.061±0.036	− 0.537±0.001	− 1.091±0.007
	ES-Chan	+ 0.530±0.057	+ 0.047±0.030	− 0.089±0.008	+ 0.198±0.008	+ 0.088±0.007	+ 0.616±0.010	− 0.147±0.009
ETTh2	VS	+ 0.238±0.120	+ 0.422±0.146	+ 0.771±0.040	+ 0.627±0.020	+ 0.639±0.021	+ 0.962±0.021	+ 0.640±0.015
	ES-Chan	+ 0.750±0.155	+ 0.225±0.079	+ 0.221±0.010	+ 0.599±0.005	+ 0.605±0.014	+ 0.978±0.009	+ 0.350±0.016
ETTm1	VS	− 5.808±0.054	− 4.589±0.131	− 3.261±0.057	− 3.395±0.025	− 3.721±0.078	− 3.191±0.027	− 3.876±0.039
	ES-Chan	+ 0.050±0.185	− 0.589±0.110	− 0.638±0.010	− 0.480±0.013	− 0.426±0.030	+ 0.099±0.016	− 0.650±0.011
ETTm2	VS	− 0.078±0.136	+ 0.437±0.130	+ 0.788±0.005	+ 0.651±0.017	+ 0.496±0.030	+ 0.932±0.011	+ 0.518±0.012
	ES-Chan	+ 0.639±0.051	+ 0.160±0.051	+ 0.053±0.008	+ 0.495±0.000	+ 0.441±0.012	+ 0.866±0.021	+ 0.195±0.019
Weather	VS	− 1.282±0.048	− 0.791±0.053	− 0.633±0.241	− 0.628±0.007	− 0.692±0.002	− 0.272±0.009	− 0.724±0.012
	ES-Chan	+ 0.294±0.052	+ 0.197±0.038	+ 0.116±0.149	+ 0.217±0.003	+ 0.289±0.003	+ 0.802±0.007	+ 0.204±0.006
ECL	VS	+ 0.154±0.012	+ 3.906±0.006	+ 0.908±0.017	+ 1.639±0.000	+ 2.880±0.018	+ 4.743±0.010	+ 4.164±0.020
	ES-Chan	+ 0.133±0.023	+ 0.825±0.014	+ 0.024±0.003	+ 0.350±0.000	+ 0.586±0.008	+ 1.640±0.004	+ 0.894±0.018
Traffic	VS	− 0.037±0.029	+ 2.703±0.010	+ 0.934±0.011	+ 2.123±0.001	+ 4.367±0.076	+ 3.577±0.005	–
	ES-Chan	+ 0.039±0.008	+ 0.895±0.006	− 0.051±0.004	+ 0.265±0.001	+ 1.179±0.034	+ 1.742±0.005	–
Solar	VS	−31.792±0.724	−34.462±0.161	−30.816±0.155	−10.783±0.009	−26.743±0.058	−33.138±0.030	−33.487±0.048
	ES-Chan	− 4.992±0.072	− 5.120±0.091	− 5.207±0.004	− 1.172±0.001	− 4.310±0.024	− 4.469±0.004	− 5.100±0.008
1st count	VS	7	1	0	0	0	0	0
	ES-Chan	0	0	6	0	0	0	2
Avg. marginal & temporal	CRPS	− 0.026±0.017	− 0.000±0.049	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.371±0.001	+ 0.000±0.000
	NMAE	− 0.032±0.049	+ 0.001±0.036	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.165±0.003	+ 0.000±0.000
	ES-Time	+ 0.017±0.045	+ 0.018±0.040	− 0.052±0.013	+ 0.729±0.003	+ 0.397±0.011	+ 0.600±0.003	− 0.033±0.005

Table 32: **Detailed experiment results for Sundial.** Percent change vs. the uncoupled arm of the same settings; negative is better. ± is the standard deviation over 3 seeds (2026–2028); all settings of a cell are first averaged per seed. Red best, blue second. *Avg. marginal & temporal* averages all cells.

Dataset	h	Metric (%)	Uncoupled	GeNCoup-Dyn (ours)	GeNCoup-Corr (ours)	COBASE	Schaake	SimSch.	GCA	IC
ETTh1	96	VS	427.6	−2.761±0.367	−1.512±0.164	−0.687±0.094	−0.635±0.029	−1.070±0.149	−0.586±0.060	−1.086±0.074
		ES-Chan	1.093	+0.404±0.226	−0.008±0.124	−0.181±0.021	+0.102±0.018	+0.017±0.009	+0.512±0.012	−0.208±0.014
	192	VS	966.6	−2.695±0.286	−1.511±0.075	−0.478±0.002	−0.413±0.011	−0.799±0.050	−0.438±0.058	−0.971±0.046
		ES-Chan	1.196	+0.467±0.030	+0.003±0.094	−0.090±0.011	+0.178±0.007	+0.093±0.012	+0.611±0.015	−0.148±0.013
	336	VS	1769	−2.325±0.474	−1.418±0.091	−0.545±0.027	−0.530±0.023	−1.001±0.041	−0.415±0.012	−0.972±0.016
		ES-Chan	1.242	+0.594±0.230	+0.100±0.078	−0.048±0.005	+0.228±0.008	+0.100±0.021	+0.651±0.007	−0.115±0.007
	720	VS	3893	−2.389±0.239	−1.680±0.042	−0.785±0.044	−0.792±0.019	−1.376±0.066	−0.708±0.006	−1.334±0.014
		ES-Chan	1.247	+0.655±0.140	+0.093±0.064	−0.036±0.011	+0.283±0.005	+0.140±0.012	+0.689±0.008	−0.117±0.008
	Avg	VS	1764	−2.542±0.180	−1.530±0.023	−0.624±0.031	−0.592±0.016	−1.061±0.036	−0.537±0.001	−1.091±0.007
		ES-Chan	1.194	+0.530±0.057	+0.047±0.030	−0.089±0.008	+0.198±0.008	+0.088±0.007	+0.616±0.010	−0.147±0.009
ETTh2	96	VS	329.7	+0.315±0.638	+0.623±0.363	+0.885±0.057	+0.754±0.017	+0.731±0.158	+1.230±0.035	+0.887±0.054
		ES-Chan	0.9952	+0.855±0.482	+0.218±0.227	+0.123±0.035	+0.606±0.012	+0.635±0.026	+0.990±0.002	+0.323±0.029
	192	VS	775	−0.011±0.529	+0.323±0.155	+0.814±0.061	+0.727±0.050	+0.880±0.103	+1.054±0.042	+0.732±0.037
		ES-Chan	1.138	+0.618±0.166	+0.157±0.072	+0.205±0.006	+0.599±0.006	+0.648±0.046	+1.013±0.028	+0.373±0.036
	336	VS	1470	+0.296±0.211	+0.379±0.137	+0.671±0.057	+0.513±0.044	+0.515±0.008	+0.788±0.029	+0.476±0.009
		ES-Chan	1.244	+0.809±0.186	+0.274±0.081	+0.239±0.006	+0.567±0.004	+0.532±0.018	+0.950±0.008	+0.330±0.017
	720	VS	3117	+0.353±0.262	+0.364±0.259	+0.712±0.031	+0.515±0.010	+0.431±0.041	+0.777±0.012	+0.465±0.011
		ES-Chan	1.262	+0.719±0.175	+0.250±0.153	+0.319±0.006	+0.625±0.004	+0.607±0.004	+0.958±0.016	+0.375±0.013
	Avg	VS	1423	+0.238±0.120	+0.422±0.146	+0.771±0.040	+0.627±0.020	+0.639±0.021	+0.962±0.021	+0.640±0.015
		ES-Chan	1.16	+0.750±0.155	+0.225±0.079	+0.221±0.010	+0.599±0.005	+0.605±0.014	+0.978±0.009	+0.350±0.016
1st count		VS		8	0	0	0	0	0	0
		ES-Chan		0	2	2	0	0	0	4
Avg. marginal & temporal		CRPS		+0.047±0.019	+0.005±0.037	+0.000±0.000	+0.000±0.000	+0.000±0.000	+0.322±0.003	+0.000±0.000
		NMAE		+0.028±0.074	−0.010±0.005	+0.000±0.000	+0.000±0.000	+0.000±0.000	+0.104±0.005	+0.000±0.000
		ES-Time		−0.003±0.053	−0.015±0.048	−0.251±0.015	+0.882±0.008	+0.347±0.029	+0.454±0.009	−0.250±0.015

Table 33: **Detailed experiment results for Sundial.** Percent change vs. the uncoupled arm of the same settings; negative is better. ± is the standard deviation over 3 seeds (2026–2028); all settings of a cell are first averaged per seed. Red best, blue second. *Avg. marginal & temporal* averages all cells.

Dataset	h	Metric (%)	Uncoupled	GeNCoup-Dyn (ours)	GeNCoup-Corr (ours)	COBASE	Schaake	SimSch.	GCA	IC
ETTm1	96	VS	436.8	−5.652±0.318	−4.174±0.360	−3.214±0.038	−3.236±0.033	−3.834±0.092	−3.073±0.089	−3.774±0.099
		ES-Chan	1.051	−0.011±0.179	−0.690±0.180	−0.783±0.027	−0.593±0.032	−0.552±0.015	−0.035±0.031	−0.778±0.027
	192	VS	965.9	−5.980±0.257	−4.812±0.143	−3.329±0.042	−3.423±0.029	−3.691±0.060	−3.197±0.094	−3.920±0.081
		ES-Chan	1.147	+0.017±0.410	−0.686±0.118	−0.700±0.011	−0.536±0.013	−0.483±0.055	+0.055±0.024	−0.697±0.017
	336	VS	1756	−6.003±0.430	−4.725±0.040	−3.362±0.083	−3.482±0.012	−3.659±0.106	−3.295±0.031	−3.977±0.045
		ES-Chan	1.198	+0.107±0.292	−0.583±0.107	−0.605±0.009	−0.453±0.011	−0.376±0.032	+0.141±0.012	−0.618±0.006
	720	VS	4093	−5.596±0.247	−4.646±0.100	−3.141±0.071	−3.438±0.048	−3.701±0.179	−3.199±0.087	−3.833±0.096
		ES-Chan	1.277	+0.088±0.107	−0.397±0.084	−0.464±0.008	−0.338±0.004	−0.293±0.028	+0.234±0.006	−0.508±0.008
	Avg	VS	1813	−5.808±0.054	−4.589±0.131	−3.261±0.057	−3.395±0.025	−3.721±0.078	−3.191±0.027	−3.876±0.039
		ES-Chan	1.168	+0.050±0.185	−0.589±0.110	−0.638±0.010	−0.480±0.013	−0.426±0.030	+0.099±0.016	−0.650±0.011
ETTm2	96	VS	230	−0.043±0.502	+0.483±0.114	+0.848±0.005	+0.730±0.030	+0.434±0.018	+1.047±0.034	+0.573±0.023
		ES-Chan	0.7839	+0.497±0.205	+0.047±0.031	−0.113±0.006	+0.450±0.005	+0.396±0.040	+0.888±0.040	+0.117±0.036
	192	VS	568.8	+0.071±0.160	+0.457±0.192	+0.785±0.020	+0.626±0.023	+0.378±0.068	+0.945±0.027	+0.528±0.021
		ES-Chan	0.9139	+0.726±0.132	+0.195±0.066	+0.016±0.017	+0.492±0.007	+0.422±0.039	+0.863±0.007	+0.188±0.005
	336	VS	1148	−0.102±0.134	+0.426±0.196	+0.744±0.017	+0.610±0.019	+0.480±0.015	+0.899±0.028	+0.487±0.035
		ES-Chan	1.013	+0.713±0.087	+0.198±0.135	+0.086±0.004	+0.501±0.003	+0.405±0.013	+0.857±0.034	+0.212±0.034
	720	VS	3054	−0.239±0.207	+0.383±0.074	+0.773±0.012	+0.637±0.005	+0.690±0.053	+0.838±0.009	+0.484±0.012
		ES-Chan	1.191	+0.619±0.052	+0.200±0.015	+0.223±0.013	+0.538±0.002	+0.540±0.027	+0.859±0.018	+0.263±0.019
	Avg	VS	1250	−0.078±0.136	+0.437±0.130	+0.788±0.005	+0.651±0.017	+0.496±0.030	+0.932±0.011	+0.518±0.012
		ES-Chan	0.9754	+0.639±0.051	+0.160±0.051	+0.053±0.008	+0.495±0.000	+0.441±0.012	+0.866±0.021	+0.195±0.019
Weather	96	VS	2992	−1.445±0.114	−1.094±0.069	−0.858±0.155	−0.868±0.007	−0.891±0.025	−0.487±0.015	−0.964±0.011
		ES-Chan	1.386	+0.239±0.140	+0.109±0.074	+0.106±0.140	+0.201±0.002	+0.256±0.018	+0.809±0.006	+0.168±0.005
	192	VS	7494	−1.213±0.080	−0.703±0.058	−0.630±0.225	−0.629±0.011	−0.668±0.008	−0.244±0.015	−0.696±0.016
		ES-Chan	1.662	+0.366±0.091	+0.229±0.048	+0.099±0.151	+0.204±0.004	+0.276±0.000	+0.811±0.014	+0.187±0.006
	336	VS	1.51×10^4	−1.214±0.033	−0.702±0.092	−0.557±0.284	−0.533±0.009	−0.632±0.021	−0.208±0.010	−0.642±0.016
		ES-Chan	1.872	+0.280±0.068	+0.201±0.108	+0.110±0.146	+0.218±0.004	+0.299±0.002	+0.795±0.008	+0.210±0.010
	720	VS	3.69×10^4	−1.255±0.153	−0.665±0.059	−0.487±0.299	−0.483±0.002	−0.577±0.012	−0.149±0.003	−0.593±0.005
		ES-Chan	2.127	+0.291±0.138	+0.247±0.045	+0.149±0.161	+0.246±0.002	+0.325±0.003	+0.794±0.004	+0.249±0.003
	Avg	VS	1.56×10^4	−1.282±0.048	−0.791±0.053	−0.633±0.241	−0.628±0.007	−0.692±0.002	−0.272±0.009	−0.724±0.012
		ES-Chan	1.762	+0.294±0.052	+0.197±0.038	+0.116±0.149	+0.217±0.003	+0.289±0.003	+0.802±0.007	+0.204±0.006
1st count		VS		12	0	0	0	0	0	0
		ES-Chan		0	1	9	0	0	0	2
Avg. marginal & temporal		CRPS		−0.054±0.033	−0.015±0.021	+0.000±0.000	+0.000±0.000	+0.000±0.000	+0.344±0.001	+0.000±0.000
		NMAE		−0.047±0.071	−0.001±0.026	+0.000±0.000	+0.000±0.000	+0.000±0.000	+0.131±0.006	+0.000±0.000
		ES-Time		+0.055±0.067	+0.077±0.081	+0.009±0.031	+1.100±0.003	+0.554±0.019	+0.611±0.003	+0.024±0.004

Table 34: **Detailed experiment results for Sundial.** Percent change vs. the uncoupled arm of the same settings; negative is better. ± is the standard deviation over 3 seeds (2026–2028); all settings of a cell are first averaged per seed. Red best, blue second. *Avg. marginal & temporal* averages all cells.

Dataset	h	Metric (%)	Uncoupled	GeNCoup-Dyn (ours)	GeNCoup-Corr (ours)	COBASE	Schaake	SimSch.	GCA	IC
ECL	96	VS	6.80×10^{5}	+ 0.132±0.042	+ 3.526±0.029	+ 0.709±0.022	+ 1.469±0.001	+ 2.574±0.044	+ 4.324±0.036	+ 3.741±0.043
		ES-Chan	5.506	+ 0.123±0.062	+ 0.843±0.059	+ 0.015±0.004	+ 0.342±0.001	+ 0.572±0.006	+ 1.630±0.027	+ 0.851±0.029
	192	VS	1.40×10^{6}	+ 0.142±0.036	+ 3.792±0.018	+ 0.797±0.012	+ 1.617±0.001	+ 2.805±0.040	+ 4.625±0.005	+ 4.040±0.013
		ES-Chan	5.65	+ 0.118±0.028	+ 0.830±0.029	+ 0.013±0.003	+ 0.363±0.001	+ 0.577±0.008	+ 1.648±0.006	+ 0.887±0.020
	336	VS	2.62×10^{6}	+ 0.136±0.033	+ 3.980±0.024	+ 0.929±0.053	+ 1.696±0.001	+ 2.937±0.071	+ 4.836±0.007	+ 4.262±0.014
		ES-Chan	5.986	+ 0.111±0.039	+ 0.781±0.057	+ 0.020±0.002	+ 0.364±0.001	+ 0.603±0.011	+ 1.649±0.004	+ 0.923±0.022
	720	VS	6.40×10^{6}	+ 0.207±0.052	+ 4.325±0.011	+ 1.199±0.017	+ 1.773±0.001	+ 3.204±0.038	+ 5.187±0.008	+ 4.613±0.011
		ES-Chan	6.624	+ 0.179±0.034	+ 0.844±0.073	+ 0.048±0.004	+ 0.331±0.000	+ 0.591±0.022	+ 1.632±0.003	+ 0.914±0.010
	Avg	VS	2.78×10^{6}	+ 0.154±0.012	+ 3.906±0.006	+ 0.908±0.017	+ 1.639±0.000	+ 2.880±0.018	+ 4.743±0.010	+ 4.164±0.020
		ES-Chan	5.942	+ 0.133±0.023	+ 0.825±0.014	+ 0.024±0.003	+ 0.350±0.000	+ 0.586±0.008	+ 1.640±0.004	+ 0.894±0.018
Traffic	96	VS	8.69×10^{6}	+ 0.011±0.128	+ 2.466±0.038	+ 0.940±0.087	+ 1.936±0.001	+ 4.037±0.081	+ 3.368±0.023	–
		ES-Chan	14.9	+ 0.057±0.051	+ 0.833±0.060	− 0.046±0.031	+ 0.187±0.002	+ 1.055±0.041	+ 1.734±0.016	–
	192	VS	1.67×10^{7}	− 0.029±0.066	+ 2.727±0.076	+ 0.941±0.030	+ 2.116±0.001	+ 4.333±0.118	+ 3.600±0.018	–
		ES-Chan	14.46	+ 0.042±0.076	+ 0.893±0.062	− 0.051±0.008	+ 0.273±0.000	+ 1.196±0.053	+ 1.775±0.017	–
	336	VS	2.99×10^{7}	− 0.033±0.030	+ 2.906±0.104	+ 0.905±0.037	+ 2.230±0.001	+ 4.519±0.102	+ 3.709±0.019	–
		ES-Chan	14.61	+ 0.023±0.082	+ 1.030±0.089	− 0.054±0.009	+ 0.298±0.001	+ 1.223±0.045	+ 1.782±0.010	–
	720	VS	6.73×10^{7}	− 0.096±0.036	+ 2.712±0.041	+ 0.950±0.092	+ 2.209±0.001	+ 4.578±0.034	+ 3.631±0.016	–
		ES-Chan	15.14	+ 0.034±0.070	+ 0.823±0.054	− 0.051±0.021	+ 0.301±0.001	+ 1.240±0.008	+ 1.676±0.005	–
	Avg	VS	3.06×10^{7}	− 0.037±0.029	+ 2.703±0.010	+ 0.934±0.011	+ 2.123±0.001	+ 4.367±0.076	+ 3.577±0.005	–
		ES-Chan	14.78	+ 0.039±0.008	+ 0.895±0.006	− 0.051±0.004	+ 0.265±0.001	+ 1.179±0.034	+ 1.742±0.005	–
Solar	96	VS	2.94×10^{5}	−30.596±0.751	−36.060±0.149	−31.265±0.104	−11.213±0.013	−29.970±0.564	−34.506±0.136	−34.922±0.104
		ES-Chan	5.825	− 5.448±0.249	− 5.853±0.083	− 5.854±0.013	− 1.255±0.002	− 5.136±0.042	− 5.098±0.021	− 5.792±0.042
	192	VS	6.37×10^{5}	−32.157±1.610	−34.885±0.274	−31.531±0.226	−10.998±0.012	−26.267±0.104	−33.456±0.113	−33.830±0.156
		ES-Chan	6.243	− 5.216±0.269	− 5.260±0.115	− 5.330±0.001	− 1.191±0.001	− 4.218±0.047	− 4.629±0.017	− 5.264±0.005
	336	VS	1.08×10^{6}	−31.120±0.284	−32.621±0.182	−29.225±0.169	−10.111±0.010	−25.075±0.230	−31.424±0.043	−31.754±0.052
		ES-Chan	6.08	− 4.695±0.183	− 4.748±0.085	− 4.821±0.002	− 1.060±0.002	− 3.967±0.055	− 4.107±0.001	− 4.682±0.011
	720	VS	2.33×10^{6}	−33.294±0.668	−34.282±0.066	−31.242±0.275	−10.812±0.007	−25.658±0.515	−33.164±0.036	−33.444±0.049
		ES-Chan	5.971	− 4.609±0.058	− 4.618±0.174	− 4.823±0.004	− 1.184±0.002	− 3.918±0.090	− 4.042±0.012	− 4.662±0.013
	Avg	VS	1.08×10^{6}	−31.792±0.724	−34.462±0.161	−30.816±0.155	−10.783±0.009	−26.743±0.058	−33.138±0.030	−33.487±0.048
		ES-Chan	6.03	− 4.992±0.072	− 5.120±0.091	− 5.207±0.004	− 1.172±0.001	− 4.310±0.024	− 4.469±0.004	− 5.100±0.008
1st count		VS		8	4	0	0	0	0	0
		ES-Chan		0	0	12	0	0	0	0
Avg. marginal & temporal		CRPS		− 0.046±0.031	+ 0.011±0.127	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.429±0.000	+ 0.000±0.000
		NMAE		− 0.057±0.038	+ 0.010±0.090	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.239±0.001	+ 0.000±0.000
		ES-Time		− 0.008±0.029	− 0.019±0.021	+ 0.020±0.001	+ 0.256±0.001	+ 0.275±0.024	+ 0.686±0.001	+ 0.097±0.002

Table 35: **Experiment results for Aurora, averaged over horizons.** Percent change vs. the uncoupled arm of the same settings; negative is better. ± is the standard deviation over 3 seeds (2026–2028); all settings of a cell are first averaged per seed. Red best, blue second. *Avg. marginal & temporal* averages all cells.

Dataset	Metric (%)	GeNCoup-Dyn (ours)	GeNCoup-Corr (ours)	COBASE	Schaake	SimSch.	GCA	IC
ETTh1	VS	− 1.260±0.083	− 0.383±0.025	+ 0.008±0.045	+ 0.255±0.007	− 0.176±0.027	+ 0.302±0.014	− 0.259±0.015
	ES-Chan	+ 0.558±0.030	+ 0.549±0.019	+ 0.293±0.271	+ 0.899±0.003	+ 0.609±0.006	+ 1.119±0.007	+ 0.378±0.004
ETTh2	VS	− 0.079±0.025	+ 0.421±0.072	+ 0.362±0.342	+ 0.553±0.021	+ 0.501±0.043	+ 0.874±0.015	+ 0.498±0.020
	ES-Chan	+ 0.584±0.142	+ 0.542±0.014	+ 0.405±0.400	+ 0.925±0.020	+ 0.918±0.020	+ 1.262±0.032	+ 0.593±0.034
ETTm1	VS	− 2.170±0.111	− 1.280±0.024	− 0.708±0.250	− 0.426±0.015	− 0.905±0.079	− 0.353±0.023	− 0.946±0.025
	ES-Chan	+ 0.617±0.056	+ 0.333±0.021	+ 0.142±0.114	+ 0.569±0.003	+ 0.502±0.013	+ 0.956±0.003	+ 0.232±0.001
ETTm2	VS	+ 0.130±0.020	+ 0.400±0.043	+ 0.620±0.002	+ 0.570±0.004	+ 0.454±0.017	+ 0.820±0.004	+ 0.441±0.008
	ES-Chan	+ 0.656±0.054	+ 0.522±0.032	+ 0.511±0.010	+ 0.857±0.002	+ 0.726±0.007	+ 1.274±0.016	+ 0.513±0.021
Weather	VS	− 0.710±0.020	− 0.319±0.032	− 0.288±0.007	− 0.228±0.004	− 0.248±0.004	+ 0.143±0.004	− 0.281±0.002
	ES-Chan	+ 0.328±0.028	+ 0.521±0.015	+ 0.452±0.005	+ 0.374±0.004	+ 0.467±0.001	+ 1.257±0.002	+ 0.464±0.003
ECL	VS	+ 0.008±0.008	+ 3.583±0.006	+ 0.992±0.013	+ 1.352±0.000	+ 2.368±0.021	+ 4.301±0.009	+ 3.686±0.017
	ES-Chan	+ 0.115±0.013	+ 1.357±0.012	+ 0.329±0.005	+ 0.487±0.002	+ 0.764±0.017	+ 2.306±0.003	+ 1.337±0.010
Traffic	VS	− 0.062±0.004	+ 3.109±0.039	+ 2.497±0.007	+ 2.240±0.000	+ 4.364±0.072	+ 3.806±0.006	–
	ES-Chan	+ 0.135±0.003	+ 1.734±0.029	+ 1.058±0.003	+ 0.345±0.001	+ 1.353±0.019	+ 2.520±0.001	–
Solar	VS	−17.428±0.718	−20.095±0.063	−17.962±0.027	− 2.722±0.001	−14.730±0.265	−19.067±0.018	−18.668±0.014
	ES-Chan	+ 0.734±0.135	+ 1.671±0.093	+ 1.080±0.013	+ 0.541±0.001	− 0.349±0.042	+ 2.206±0.008	+ 1.395±0.007
1st count	VS	7	1	0	0	0	0	0
	ES-Chan	3	0	4	0	1	0	0
Avg. marginal & temporal	CRPS	+ 0.009±0.009	+ 0.020±0.002	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.229±0.001	+ 0.000±0.000
	NMAE	+ 0.000±0.014	+ 0.016±0.009	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.068±0.001	+ 0.000±0.000
	ES-Time	+ 0.009±0.010	+ 0.005±0.014	− 0.575±0.113	+ 0.981±0.003	+ 0.478±0.021	+ 0.514±0.005	− 0.708±0.006

Table 36: **Detailed experiment results for Aurora.** Percent change vs. the uncoupled arm of the same settings; negative is better. ± is the standard deviation over 3 seeds (2026–2028); all settings of a cell are first averaged per seed. Red best, blue second. *Avg. marginal & temporal* averages all cells.

Dataset	h	Metric (%)	Uncoupled	GeNCoup-Dyn (ours)	GeNCoup-Corr (ours)	COBASE	Schaake	SimSch.	GCA	IC
ETTh1	96	VS	431.7	−1.237±0.193	−0.345±0.048	−0.043±0.089	+0.217±0.049	−0.211±0.091	+0.294±0.032	−0.215±0.055
		ES-Chan	1.146	+0.440±0.042	+0.500±0.025	+0.218±0.228	+0.781±0.002	+0.533±0.008	+0.965±0.007	+0.299±0.016
	192	VS	992.1	−1.308±0.263	−0.324±0.044	+0.024±0.048	+0.331±0.009	−0.044±0.018	+0.253±0.049	−0.207±0.059
		ES-Chan	1.251	+0.567±0.038	+0.506±0.047	+0.263±0.245	+0.828±0.005	+0.574±0.002	+1.057±0.008	+0.337±0.014
	336	VS	1846	−1.256±0.210	−0.378±0.088	+0.030±0.037	+0.260±0.015	−0.163±0.034	+0.300±0.035	−0.250±0.029
		ES-Chan	1.307	+0.545±0.083	+0.548±0.036	+0.313±0.281	+0.925±0.004	+0.641±0.017	+1.157±0.012	+0.401±0.005
	720	VS	4069	−1.238±0.112	−0.485±0.033	+0.023±0.035	+0.215±0.004	−0.287±0.038	+0.361±0.009	−0.366±0.012
		ES-Chan	1.323	+0.679±0.055	+0.643±0.035	+0.378±0.328	+1.062±0.004	+0.688±0.005	+1.298±0.009	+0.476±0.008
	Avg	VS	1835	−1.260±0.083	−0.383±0.025	+0.008±0.045	+0.255±0.007	−0.176±0.027	+0.302±0.014	−0.259±0.015
		ES-Chan	1.257	+0.558±0.030	+0.549±0.019	+0.293±0.271	+0.899±0.003	+0.609±0.006	+1.119±0.007	+0.378±0.004
ETTh2	96	VS	346.4	−0.181±0.277	+0.550±0.267	+0.458±0.473	+0.746±0.061	+0.577±0.165	+1.204±0.053	+0.730±0.050
		ES-Chan	1.063	+0.639±0.237	+0.497±0.034	+0.342±0.443	+0.921±0.069	+0.872±0.050	+1.290±0.075	+0.563±0.062
	192	VS	792.8	−0.331±0.466	+0.369±0.048	+0.336±0.334	+0.545±0.032	+0.635±0.006	+0.893±0.024	+0.526±0.027
		ES-Chan	1.186	+0.463±0.355	+0.480±0.011	+0.394±0.385	+0.917±0.022	+1.043±0.048	+1.273±0.045	+0.583±0.051
	336	VS	1508	+0.013±0.098	+0.377±0.061	+0.324±0.286	+0.454±0.011	+0.449±0.014	+0.700±0.006	+0.383±0.016
		ES-Chan	1.294	+0.580±0.010	+0.569±0.029	+0.415±0.369	+0.904±0.004	+0.940±0.013	+1.218±0.020	+0.589±0.027
	720	VS	3168	+0.181±0.147	+0.390±0.025	+0.329±0.276	+0.468±0.017	+0.344±0.020	+0.700±0.018	+0.356±0.013
		ES-Chan	1.308	+0.654±0.100	+0.623±0.038	+0.469±0.404	+0.958±0.005	+0.816±0.006	+1.265±0.006	+0.635±0.003
	Avg	VS	1454	−0.079±0.025	+0.421±0.072	+0.362±0.342	+0.553±0.021	+0.501±0.043	+0.874±0.015	+0.498±0.020
		ES-Chan	1.213	+0.584±0.142	+0.542±0.014	+0.405±0.400	+0.925±0.020	+0.918±0.020	+1.262±0.032	+0.593±0.034
1st count		VS		8	0	0	0	0	0	0
		ES-Chan		0	0	8	0	0	0	0
Avg. marginal & temporal		CRPS		+0.017±0.016	+0.007±0.005	+0.000±0.000	+0.000±0.000	+0.000±0.000	+0.204±0.002	+0.000±0.000
		NMAE		+0.004±0.007	+0.000±0.004	+0.000±0.000	+0.000±0.000	+0.000±0.000	+0.041±0.000	+0.000±0.000
		ES-Time		+0.020±0.027	+0.001±0.014	−0.258±0.223	+1.646±0.005	+0.632±0.052	+0.706±0.017	−0.381±0.013

Table 37: **Detailed experiment results for Aurora.** Percent change vs. the uncoupled arm of the same settings; negative is better. ± is the standard deviation over 3 seeds (2026–2028); all settings of a cell are first averaged per seed. Red best, blue second. *Avg. marginal & temporal* averages all cells.

Dataset	h	Metric (%)	Uncoupled	GeNCoup-Dyn (ours)	GeNCoup-Corr (ours)	COBASE	Schaake	SimSch.	GCA	IC
ETTm1	96	VS	429.6	−2.513±0.292	−1.943±0.139	−1.311±0.424	−1.319±0.047	−1.848±0.155	−1.110±0.099	−1.704±0.098
		ES-Chan	1.095	+0.567±0.038	+0.324±0.107	+0.041±0.158	+0.564±0.006	+0.499±0.028	+0.977±0.020	+0.169±0.003
	192	VS	1166	−2.153±0.084	−1.243±0.023	−0.673±0.365	−0.436±0.013	−0.858±0.092	−0.366±0.055	−0.917±0.058
		ES-Chan	1.413	+0.531±0.062	+0.305±0.038	+0.101±0.143	+0.528±0.006	+0.454±0.031	+0.910±0.016	+0.216±0.015
	336	VS	2339	−1.997±0.321	−0.935±0.049	−0.370±0.232	+0.048±0.008	−0.414±0.029	+0.050±0.023	−0.555±0.027
		ES-Chan	1.573	+0.631±0.095	+0.330±0.035	+0.140±0.154	+0.558±0.002	+0.505±0.005	+0.942±0.002	+0.249±0.002
	720	VS	5279	−2.019±0.183	−0.997±0.006	−0.478±0.022	+0.002±0.007	−0.500±0.065	+0.016±0.029	−0.608±0.035
		ES-Chan	1.633	+0.740±0.053	+0.372±0.018	+0.285±0.002	+0.626±0.001	+0.551±0.011	+0.997±0.004	+0.294±0.006
	Avg	VS	2304	−2.170±0.111	−1.280±0.024	−0.708±0.250	−0.426±0.015	−0.905±0.079	−0.353±0.023	−0.946±0.025
		ES-Chan	1.429	+0.617±0.056	+0.333±0.021	+0.142±0.114	+0.569±0.003	+0.502±0.013	+0.956±0.003	+0.232±0.001
ETTm2	96	VS	221.2	+0.135±0.087	+0.476±0.047	+0.723±0.007	+0.671±0.024	+0.415±0.013	+1.000±0.011	+0.574±0.010
		ES-Chan	0.7944	+0.792±0.128	+0.511±0.046	+0.465±0.006	+0.904±0.007	+0.683±0.026	+1.370±0.044	+0.516±0.043
	192	VS	573.3	+0.079±0.008	+0.399±0.103	+0.669±0.010	+0.615±0.006	+0.418±0.025	+0.854±0.009	+0.467±0.021
		ES-Chan	0.9523	+0.548±0.106	+0.501±0.066	+0.486±0.017	+0.866±0.012	+0.644±0.018	+1.271±0.002	+0.514±0.017
	336	VS	1195	+0.135±0.013	+0.447±0.023	+0.598±0.012	+0.552±0.009	+0.457±0.030	+0.787±0.005	+0.408±0.014
		ES-Chan	1.085	+0.720±0.161	+0.573±0.006	+0.519±0.006	+0.838±0.002	+0.668±0.009	+1.246±0.023	+0.507±0.033
	720	VS	3187	+0.170±0.025	+0.280±0.045	+0.491±0.002	+0.443±0.001	+0.527±0.019	+0.639±0.002	+0.314±0.004
		ES-Chan	1.278	+0.562±0.053	+0.502±0.037	+0.575±0.012	+0.819±0.003	+0.910±0.016	+1.209±0.008	+0.515±0.012
	Avg	VS	1294	+0.130±0.020	+0.400±0.043	+0.620±0.002	+0.570±0.004	+0.454±0.017	+0.820±0.004	+0.441±0.008
		ES-Chan	1.027	+0.656±0.054	+0.522±0.032	+0.511±0.010	+0.857±0.002	+0.726±0.007	+1.274±0.016	+0.513±0.021
Weather	96	VS	3374	−1.002±0.095	−0.388±0.099	−0.360±0.015	−0.297±0.012	−0.309±0.017	+0.096±0.015	−0.380±0.011
		ES-Chan	1.543	+0.236±0.059	+0.544±0.051	+0.468±0.012	+0.366±0.012	+0.456±0.023	+1.300±0.008	+0.456±0.008
	192	VS	7917	−0.785±0.140	−0.409±0.041	−0.329±0.006	−0.274±0.010	−0.278±0.001	+0.075±0.005	−0.322±0.002
		ES-Chan	1.756	+0.310±0.165	+0.496±0.029	+0.461±0.010	+0.391±0.009	+0.485±0.009	+1.289±0.012	+0.474±0.016
	336	VS	1.58×10^4	−0.594±0.058	−0.276±0.028	−0.257±0.007	−0.189±0.002	−0.226±0.004	+0.180±0.005	−0.236±0.008
		ES-Chan	1.974	+0.399±0.052	+0.530±0.044	+0.448±0.010	+0.381±0.006	+0.473±0.006	+1.259±0.004	+0.472±0.011
	720	VS	3.86×10^4	−0.457±0.024	−0.201±0.047	−0.205±0.004	−0.151±0.002	−0.181±0.002	+0.222±0.004	−0.185±0.002
		ES-Chan	2.242	+0.365±0.017	+0.513±0.036	+0.432±0.003	+0.357±0.005	+0.454±0.004	+1.180±0.006	+0.455±0.004
	Avg	VS	1.64×10^4	−0.710±0.020	−0.319±0.032	−0.288±0.007	−0.228±0.004	−0.248±0.004	+0.143±0.004	−0.281±0.002
		ES-Chan	1.879	+0.328±0.028	+0.521±0.015	+0.452±0.005	+0.374±0.004	+0.467±0.001	+1.257±0.002	+0.464±0.003
1st count		VS		12	0	0	0	0	0	0
		ES-Chan		2	1	6	2	0	0	1
Avg. marginal & temporal		CRPS		−0.026±0.021	−0.003±0.020	+0.000±0.000	+0.000±0.000	+0.000±0.000	+0.179±0.000	+0.000±0.000
		NMAE		−0.026±0.032	−0.001±0.018	+0.000±0.000	+0.000±0.000	+0.000±0.000	+0.045±0.003	+0.000±0.000
		ES-Time		+0.010±0.002	+0.013±0.032	−0.728±0.154	+1.478±0.005	+0.697±0.011	+0.338±0.002	−0.814±0.004

Table 38: **Detailed experiment results for Aurora.** Percent change vs. the uncoupled arm of the same settings; negative is better. ± is the standard deviation over 3 seeds (2026–2028); all settings of a cell are first averaged per seed. Red best, blue second. *Avg. marginal & temporal* averages all cells.

Dataset	h	Metric (%)	Uncoupled	GeNCoup-Dyn (ours)	GeNCoup-Corr (ours)	COBASE	Schaake	SimSch.	GCA	IC
ECL	96	VS	8.13×10^{5}	− 0.011±0.066	+ 3.245±0.042	+ 0.794±0.025	+ 1.216±0.001	+ 2.139±0.044	+ 3.956±0.040	+ 3.356±0.045
		ES-Chan	6.486	+ 0.102±0.050	+ 1.252±0.064	+ 0.336±0.005	+ 0.468±0.004	+ 0.743±0.001	+ 2.256±0.032	+ 1.278±0.051
	192	VS	1.63×10^{6}	+ 0.006±0.029	+ 3.523±0.015	+ 0.873±0.003	+ 1.333±0.001	+ 2.312±0.038	+ 4.197±0.014	+ 3.593±0.005
		ES-Chan	6.46	+ 0.088±0.029	+ 1.378±0.025	+ 0.297±0.008	+ 0.492±0.002	+ 0.747±0.004	+ 2.292±0.016	+ 1.324±0.013
	336	VS	3.03×10^{6}	+ 0.021±0.020	+ 3.664±0.009	+ 1.016±0.049	+ 1.398±0.001	+ 2.417±0.056	+ 4.392±0.011	+ 3.770±0.017
		ES-Chan	6.77	+ 0.131±0.039	+ 1.386±0.021	+ 0.314±0.005	+ 0.494±0.002	+ 0.777±0.032	+ 2.323±0.002	+ 1.353±0.003
	720	VS	7.30×10^{6}	+ 0.014±0.022	+ 3.902±0.013	+ 1.285±0.021	+ 1.461±0.000	+ 2.605±0.039	+ 4.658±0.004	+ 4.024±0.005
		ES-Chan	7.423	+ 0.138±0.010	+ 1.413±0.014	+ 0.368±0.008	+ 0.495±0.001	+ 0.787±0.049	+ 2.352±0.004	+ 1.393±0.006
	Avg	VS	3.19×10^{6}	+ 0.008±0.008	+ 3.583±0.006	+ 0.992±0.013	+ 1.352±0.000	+ 2.368±0.021	+ 4.301±0.009	+ 3.686±0.017
		ES-Chan	6.785	+ 0.115±0.013	+ 1.357±0.012	+ 0.329±0.005	+ 0.487±0.002	+ 0.764±0.017	+ 2.306±0.003	+ 1.337±0.010
Traffic	96	VS	1.05×10^{7}	− 0.011±0.035	+ 2.965±0.059	+ 2.511±0.071	+ 2.167±0.001	+ 4.136±0.064	+ 3.680±0.022	–
		ES-Chan	17.35	+ 0.164±0.038	+ 1.694±0.003	+ 1.118±0.039	+ 0.347±0.002	+ 1.296±0.022	+ 2.483±0.014	–
	192	VS	1.96×10^{7}	− 0.078±0.032	+ 2.932±0.028	+ 2.235±0.024	+ 2.117±0.000	+ 4.166±0.111	+ 3.635±0.015	–
		ES-Chan	16.38	+ 0.100±0.038	+ 1.659±0.030	+ 0.954±0.011	+ 0.349±0.002	+ 1.377±0.026	+ 2.492±0.015	–
	336	VS	3.50×10^{7}	− 0.063±0.002	+ 3.120±0.042	+ 2.424±0.031	+ 2.246±0.000	+ 4.407±0.114	+ 3.804±0.019	–
		ES-Chan	16.56	+ 0.129±0.006	+ 1.721±0.044	+ 1.011±0.020	+ 0.337±0.001	+ 1.352±0.034	+ 2.518±0.003	–
	720	VS	7.84×10^{7}	− 0.095±0.006	+ 3.418±0.048	+ 2.818±0.073	+ 2.429±0.001	+ 4.746±0.024	+ 4.104±0.015	–
		ES-Chan	17.1	+ 0.145±0.019	+ 1.861±0.055	+ 1.151±0.039	+ 0.347±0.001	+ 1.385±0.005	+ 2.589±0.010	–
	Avg	VS	3.59×10^{7}	− 0.062±0.004	+ 3.109±0.039	+ 2.497±0.007	+ 2.240±0.000	+ 4.364±0.072	+ 3.806±0.006	–
		ES-Chan	16.85	+ 0.135±0.003	+ 1.734±0.029	+ 1.058±0.003	+ 0.345±0.001	+ 1.353±0.019	+ 2.520±0.001	–
Solar	96	VS	2.55×10^{5}	−16.014±2.064	−21.247±0.208	−19.037±0.056	− 4.520±0.007	−16.998±0.121	−20.117±0.068	−19.674±0.168
		ES-Chan	6.368	+ 0.668±0.369	+ 2.045±0.255	+ 1.262±0.050	+ 0.617±0.006	− 0.314±0.065	+ 2.541±0.040	+ 1.606±0.064
	192	VS	5.41×10^{5}	−18.670±1.297	−21.444±0.138	−18.929±0.052	− 2.909±0.002	−14.719±0.627	−20.296±0.073	−19.811±0.085
		ES-Chan	6.636	+ 0.511±0.174	+ 1.335±0.134	+ 0.817±0.030	+ 0.442±0.003	− 0.455±0.065	+ 1.953±0.050	+ 1.154±0.041
	336	VS	8.73×10^{5}	−16.213±0.298	−17.729±0.058	−15.782±0.033	− 1.500±0.006	−12.551±0.328	−16.772±0.027	−16.457±0.039
		ES-Chan	6.702	+ 0.803±0.219	+ 1.600±0.010	+ 1.071±0.038	+ 0.564±0.006	− 0.251±0.058	+ 2.136±0.007	+ 1.378±0.017
	720	VS	1.90×10^{6}	−18.815±0.699	−19.959±0.030	−18.098±0.055	− 1.958±0.005	−14.653±0.248	−19.083±0.038	−18.731±0.030
		ES-Chan	6.801	+ 0.952±0.157	+ 1.704±0.062	+ 1.172±0.019	+ 0.542±0.000	− 0.375±0.057	+ 2.194±0.019	+ 1.443±0.020
	Avg	VS	8.91×10^{5}	−17.428±0.718	−20.095±0.063	−17.962±0.027	− 2.722±0.001	−14.730±0.265	−19.067±0.018	−18.668±0.014
		ES-Chan	6.627	+ 0.734±0.135	+ 1.671±0.093	+ 1.080±0.013	+ 0.541±0.001	− 0.349±0.042	+ 2.206±0.008	+ 1.395±0.007
1st count		VS		8	4	0	0	0	0	0
		ES-Chan		8	0	0	0	4	0	0
Avg. marginal & temporal		CRPS		+ 0.040±0.044	+ 0.052±0.026	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.296±0.001	+ 0.000±0.000
		NMAE		+ 0.024±0.033	+ 0.043±0.007	+ 0.000±0.000	+ 0.000±0.000	+ 0.000±0.000	+ 0.109±0.000	+ 0.000±0.000
		ES-Time		+ 0.002±0.021	+ 0.000±0.024	− 0.634±0.001	+ 0.041±0.001	+ 0.157±0.014	+ 0.562±0.002	− 0.877±0.003