CONDITIONAL PROJECTIVE SCENARIO SYSTEMS FOR LONG-HORIZON JOINT FORECASTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Modern time-series foundation models (TSFMs) can estimate value distributions at individual future time steps, but these marginals do not specify which values are likely to occur together. This matters for quantities such as the future maximum, maximum one-step change, or time spent above a threshold, which depend on the complete forecast trajectory. In this study, we introduce the Conditional Projective Scenario System (CPSS), which transforms marginal forecasts from a frozen TSFM (TimesFM-2.5) into 64 weighted complete trajectories. CPSS first represents broad future alternatives and then introduces finer variations while preserving consistency across forecast horizons, so identical truncated prefixes are merged and their probabilities summed. Across eight datasets, CPSS achieves increasing gains over a closely matched 64-trajectory baseline as the forecast horizon grows. On a separate 12-domain, 128-step benchmark, CPSS reduces complete-trajectory error by 6.80% and error on path-dependent quantities by 16.80%. Under present evidence, these results demonstrate that CPSS can extend strong marginal forecasters into coherent scenario generators, showing great promise for more accurate probabilistic reasoning over complete future trajectories.

1 INTRODUCTION

Time-series foundation models (TSFMs) (Woo et al., 2024; Ansari et al., 2025) have substantially improved zero-shot and cross-domain forecasting. When predictive uncertainty is expressed through lead-wise quantiles, these forecasts characterise which values are plausible at each future time without specifying how those values are coupled across the forecast horizon. Two forecasts can therefore have identical timestamp-wise marginals while representing very different temporal behaviour, such as persistence versus reversal. This distinction matters in real-world applications where decisions depend on the complete future path, such as peak demand and ramp events in power systems, hazardous-weather duration, or cumulative exposure and threshold crossings in financial risk. Such quantities generally require a joint predictive distribution over future trajectories rather than timestamp-wise marginals alone.

Representing this joint distribution becomes increasingly challenging as the forecast horizon grows. An H-step trajectory lies in $\mathbb{R}^H$, while practical forecasting systems often require uncertainty to be represented by a compact, finite set of scenarios. Under a fixed scenario budget, each trajectory must account for an increasingly rich set of possible temporal evolutions as H grows. Moreover, uncertainty can emerge at different stages of the forecast where multiple futures may agree initially and diverge only at later lead times. These considerations motivate a compact representation that can progressively introduce new uncertainty as the forecast unfolds.

In this paper, we propose a Conditional Projective Scenario System (CPSS), which transforms the marginal forecasts of a frozen TSFM into a weighted distribution over complete future trajectories. CPSS uses TimesFM-2.5 (Das et al., 2024) as a carrier and represents deviations from its forecast through a hierarchy of suffix-supported refinements. As refinements near the root can modify a large portion of the future, deeper refinements are restricted to progressively later lead times and leave the preceding prefix unchanged. Related scenarios therefore share the same early trajectory before branching into finer alternatives. The hierarchy also distributes probability mass across these branches, so that when a long-horizon forecast is restricted to a shorter horizon, identical prefixes merge and their descendant probabilities are summed. CPSS thus re-organises a fixed scenario budget from coarse to fine while keeping the underlying TSFM frozen.

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.

Experiments demonstrate that this temporal organisation becomes increasingly beneficial as the forecast horizon grows. Across eight standard time-series datasets, we compare CPSS with GlobalBasis-64 (GB64), a closely matched 64-trajectory control in which each scenario is parametrised independently over the full forecast horizon using the same frozen carrier. CPSS exhibits a clear horizon-dependent crossover, performing worse at short horizons but reaching $+8.52$ energy skill points at $H = 720$. On a separate case-disjoint evaluation spanning 12 domains at $H = 128$, CPSS reduces complete-trajectory energy by 6.80% and the continuous ranked probability score (CRPS) of path-dependent quantities by 16.80%. Under a shared-carrier comparison, CPSS also reduces complete-path energy relative to TimeMCL (Cortes et al., 2025) while remaining competitive on timestamp-wise forecasting metrics.

Overall, the key contributions of this paper are:

1. **Hierarchical scenario construction:** We introduce CPSS, a finite weighted distribution over complete future trajectories constructed around a frozen TSFM carrier using hierarchical, suffix-supported refinements.
2. **Exact shared-prefix consistency:** We establish an exact shared-prefix structure across forecast horizons: when a long-horizon forecast is restricted, descendants that have not yet diverged merge into a common coarse scenario, with their probability masses preserved through direct summation.
3. **Long-horizon advantage:** We demonstrate a systematic horizon-dependent crossover relative to the closely matched GB64 control, with the advantage of CPSS increasing as the forecast horizon grows and reaching $+8.52$ energy skill points at $H = 720$.
4. **Improved joint forecasting:** On a separate case-disjoint $H = 128$ benchmark, we demonstrate improvements in complete-path forecasting and probabilistic prediction of path-dependent quantities, while also outperforming TimeMCL under a shared-carrier comparison and remaining competitive on timestamp-wise forecasting metrics.

2 Related Work

2.1 Foundation models and joint trajectory forecasting

TSFMs have substantially improved zero-shot and cross-domain forecasting through architectures based on patched decoders, tokenised predictive distributions, masked encoders, and universal forecasting interfaces (Das et al., 2024; Ansari et al., 2024; Woo et al., 2024; Ansari et al., 2025). More recent models extend these capabilities across forecasting regimes and probabilistic outputs, including TiRex (Podest et al., 2026) and Moirai-2 (Liu et al., 2026). When uncertainty is represented through lead-wise quantiles or marginal predictive distributions, these models provide strong information about individual future times, but do not by themselves specify how uncertainty is coupled across the complete forecast trajectory. CPSS builds on this capability rather than replacing it by keeping the foundation model frozen and learning a structured joint distribution around its forecast.

Compared with marginal forecasting, joint trajectory forecasting seeks a distribution over complete future paths. Existing approaches use multiple-choice trajectory sets (Cortes et al., 2025), weighted finite scenarios (Dai et al., 2026), copula-based dependence models (Ashok et al., 2024), or conditional generative processes (Kollovieh et al., 2025). CPSS is closest to finite-scenario methods, but organises its support hierarchically, with related trajectories sharing identical prefixes and diverging only through progressively later refinements. This ultimately yields a nested coarse-to-fine representation rather than a collection of independently parameterised paths.

2.2 Finite scenarios, cross-horizon structure, and probabilistic scoring

Finite scenario representations have a long history in stochastic programming, where complex or continuous uncertainty distributions are approximated by a smaller weighted set of representative outcomes (Dupačová et al., 2003). Related probabilistic work studies consistency between distributions defined at different resolutions or under marginalisation, including consistent recovery of joint distributions and marginalisation-compatible forecasting (Majid et al., 2025; Yalavarthi et al., 2026). Any distribution over complete H-step trajectories naturally induces a valid shorter-horizon distribution by truncation. CPSS imposes an additional support-level structure by making trajectories that belong to the same branch be exactly identical before their point of divergence. Restricting the forecast to an earlier horizon therefore causes such descendants to merge into a single coarse scenario,

whose probability is the sum of their descendant masses. This resulting hierarchy explicitly links scenarios across horizons while preserving a fixed finite scenario budget.

Evaluating these forecasts requires distinguishing joint trajectory quality from timestamp-wise accuracy. Proper scoring rules thefore provide a principled framework for doing so (Gneiting & Raftery, 2007). We use the energy score to evaluate distributions over complete trajectory vectors, while CRPS evaluates both timestamp-wise marginals and scalar functionals of a trajectory. This separation enables us to test whether improved joint structure translates into better prediction of path-dependent quantities without sacrificing the quality of the induced marginal forecasts.

3 Conditional Projective Scenario System

3.1 Carrier-conditioned distribution over complete paths

Given an observed history x, the frozen TimesFM-2.5 carrier produces nine quantile trajectories over a forecast horizon of length H. Let $q_\tau(x) \in \mathbb{R}^H$ denote the trajectory associated with quantile level τ, and let $q_{\tau,t}(x)$ denote its value at forecast step t. We use the median trajectory

$$m(x) = q_{0.5}(x) \in \mathbb{R}^H \tag{1}$$

as the centre of the forecast. We also summarise the uncertainty provided by the carrier using the mean width of the 10th–90th percentile interval,

$$c(x) = \max\left\{ \frac{1}{H} \sum_{t=1}^{H} \left| q_{0.9,t}(x) - q_{0.1,t}(x) \right|, 10^{-3} \right\}. \tag{2}$$

The lower bound of 10^{-3} prevents numerical degeneracy when the carrier uncertainty is very small. With this safeguard, $c(x)$ provides a stable, target-independent scale for the scenario head. The learned refinements introduced below determine both the magnitude of the resulting scenario spread and how that spread evolves across the forecast horizon.

CPSS models the joint predictive distribution using a fixed set of $L = 64$ complete future trajectories. For history x, we write

$$Q_x^H = \sum_{\ell=1}^{L} \omega_\ell(x)\, \delta_{Z_\ell(x)}, \qquad \omega_\ell(x) \geq 0, \qquad \sum_{\ell=1}^{L} \omega_\ell(x) = 1. \tag{3}$$

Here, $Z_\ell(x) \in \mathbb{R}^H$ is the ℓ-th complete future trajectory and $\omega_\ell(x)$ is its probability mass. The notation $\delta_{Z_\ell(x)}$ denotes a Dirac point mass concentrated at $Z_\ell(x)$. Thus, Q_x^H is a normalised discrete distribution over complete future paths. We next describe how the trajectories $Z_\ell(x)$ are constructed around the carrier forecast.

3.2 Suffix-supported hierarchical refinement

Rather than parameterising each of the L trajectories independently over the entire forecast horizon, CPSS organises them as the leaves of a binary tree. Each node introduces a refinement to the trajectory inherited from its parent. Refinements near the root may influence a large portion of the future, while refinements deeper in the tree begin progressively later and modify only the remaining suffix. As a result, related scenarios share an identical prefix before separating into finer alternatives.

Let $\tau_v \in \{0, \ldots, H-1\}$ denote the forecast index after which the refinement associated with node v becomes active. For a temporal basis index k, we define a temporal basis function:

$$\phi_{\tau_v,k}(t) = \mathbf{1}\{t > \tau_v\} \kappa_{\tau_v,k} \cos\left(\frac{\pi k (t - \tau_v - 1/2)}{H - \tau_v} \right), \qquad t \in \{1, \ldots, H\}, \tag{4}$$

where $\mathbf{1}\{\cdot\}$ is the indicator function and $\kappa_{\tau_v,k}$ is a fixed scaling constant for the basis component. The indicator makes $\phi_{\tau_v,k}(t)$ exactly zero for all $t \leq \tau_v$. A refinement introduced at node v therefore cannot alter the prefix that precedes its activation point.

For each node v, the scenario head predicts bounded coefficients $\alpha_{v,k}(x)$ for a fixed collection of temporal basis functions. The refinement contributed by that node is:

$$u_v(x) = \sum_{k \in \mathcal{K}} \alpha_{v,k}(x)\, \phi_{\tau_v,k}, \tag{5}$$

where $\mathcal{K}$ denotes the set of basis indices used by CPSS. In the headline configuration, each refinement level uses four cosine components. Combining several basis components allows the model to represent

different temporal patterns, while suffix support ensures that refinements introduced later in the tree leave earlier forecast values unchanged.

A complete trajectory is obtained by following a path from the root to a terminal leaf and accumulating the refinements encountered along that path:

$$Z_\ell(x) = m(x) + c(x) \sum_{v \preceq \ell} u_v(x). \tag{6}$$

The relation $v \preceq \ell$ means that node v lies on the root-to-leaf path terminating at leaf ℓ. The carrier median $m(x)$ therefore provides the common central trajectory, while $c(x)$ scales the learned residual structure. Since every descendant can modify only its designated suffix, two related leaves remain exactly identical until the point at which their corresponding branches become active and diverge.

The suffix structure and the tree hierarchy therfore serve complementary roles: suffix support determines when additional variation may enter the forecast, while hierarchical branching determines which scenarios share a common earlier trajectory and how a coarse possibility divides into finer alternatives. Together, these components organise the fixed scenario budget from coarse to fine as the forecast horizon unfolds.

3.3 Probability masses and exact restriction

We now assign probability mass to its branches (Figure 1). Each internal node v predicts a branch probability $p_v(x) = \sigma\big(\eta_v(x)\big)$, where $\eta_v(x) \in \mathbb{R}$ is the branch logit predicted by the scenario head and $\sigma(r) = 1/(1+\exp(-r))$ is the logistic sigmoid. For terminal leaf ℓ, let $b_{\ell v} \in \{0,1\}$ indicate which child of node v is followed along the root-to-leaf path. The probability mass assigned to leaf ℓ is then:

$$\omega_\ell(x) = \prod_{v \prec \ell} p_v(x)^{b_{\ell v}} \left[1 - p_v(x)\right]^{1 - b_{\ell v}}, \tag{7}$$

where $v \prec \ell$ denotes the internal nodes encountered before reaching terminal leaf ℓ. At every internal node, the incoming probability mass is divided between its two children. Consequently, the masses of all terminal leaves sum to one.

We next consider the same predictive distribution at a shorter horizon $h < H$. For this, we define the restriction operator:

$$R_{h:H} : \mathbb{R}^H \to \mathbb{R}^h, \qquad R_{h:H}(z_1, \ldots, z_H) = (z_1, \ldots, z_h), \tag{8}$$

which retains only the first h forecast steps. Applying this operator to every trajectory induces the shorter-horizon distribution:

$$Q_x^h = (R_{h:H})_{\#} Q_x^H, \tag{9}$$

where $(R_{h:H})_{\#}$ denotes the pushforward of the distribution under the restriction map.

For a generic set of complete trajectories, restricting the horizon simply truncates every path. CPSS provides additional structure because descendants whose later refinements have not yet become active are exactly identical over the first h steps. For a restricted trajectory $z \in \mathbb{R}^h$, we define its prefix class as:

$$\mathcal{C}_h(z) = \{\ell : R_{h:H} Z_\ell(x) = z\}. \tag{10}$$

All terminal leaves in $\mathcal{C}_h(z)$ therefore correspond to the same trajectory at horizon h. Their masses on coalescence can be combined into:

$$\omega_h(z) = \sum_{\ell \in \mathcal{C}_h(z)} \omega_\ell(x). \tag{11}$$

Restricting a long-horizon CPSS forecast therefore produces a smaller set of distinct coarse scenarios whenever later branches have not yet diverged. Their probability masses are obtained directly by summing the masses of the corresponding descendants.

3.4 Training objective

The predictive distribution in Equation 3 is defined over complete trajectories, so a timestamp-wise training loss alone cannot evaluate the dependence structure across forecast steps. We therefore use the energy score as the primary complete-path objective. For a predictive distribution μ over $\mathbb{R}^h$, an observed future $y \in \mathbb{R}^h$, and the scaled Euclidean distance $d_h(u,v) = h^{-1/2}\|u - v\|_2$, the energy score (ES) is defined as:

$$\mathrm{ES}(\mu, y) = \int d_h(z, y)\, \mathrm{d}\mu(z) - \frac{1}{2} \iint d_h(z, z')\, \mathrm{d}\mu(z)\, \mathrm{d}\mu(z'). \tag{12}$$

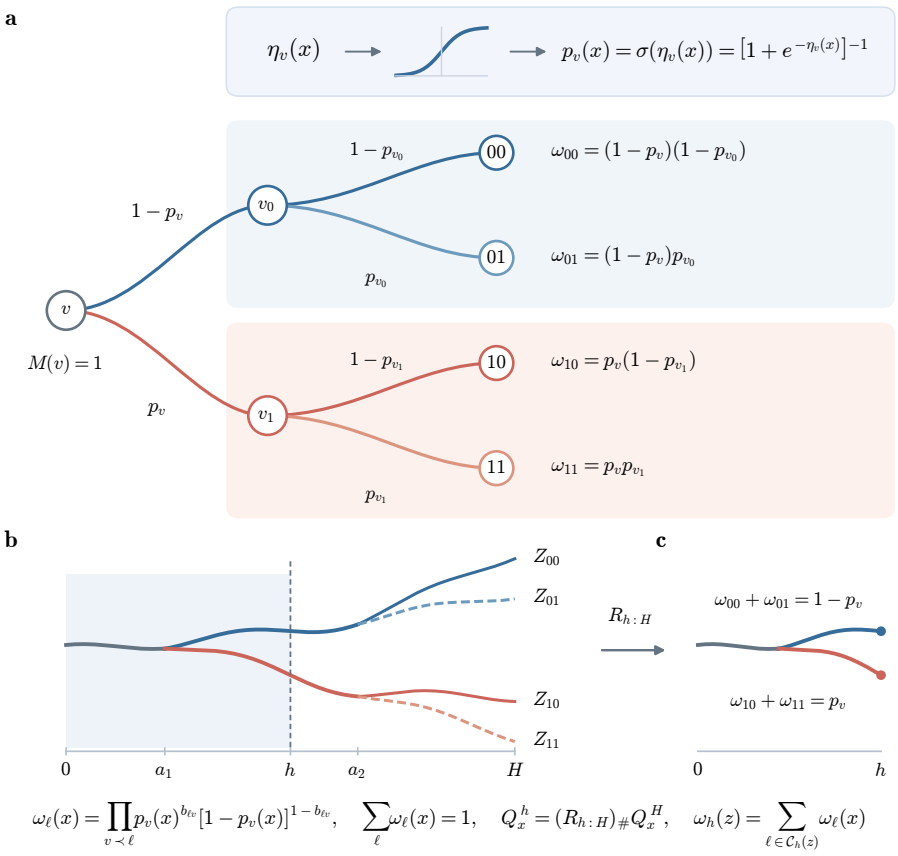

Figure 1: Illustrative sigmoid branching and exact prefix merging mechanism in CPSS using four schematic leaves. (a) Each node converts its history-conditioned logit into $p_v(x) = \sigma(\eta_v(x))$, assigning fractions $1 - p_v$ and p_v of its incoming mass to its children. Binary leaf indices identify the branch sequence, and terminal masses multiply the corresponding conditional probabilities, with dependence on x suppressed in the tree. (b) Suffix refinements at a_1 and a_2 produce nested shared prefixes. Shading marks the prefix retained at $a_1 < h < a_2$. (c) Restriction merges each pair of identical prefixes and sums its masses. Colours link descendants to their trajectories, and line widths do not encode probabilities.

The first term on the right, measures the distance between the predicted trajectories and the realised future. The second term accounts for dispersion within the predictive distribution. Lower ES scores indicate better probabilistic forecasts.

For a finite CPSS distribution, these integrals reduce to weighted sums. Let $Z_i^{(h)} = R_{h:H} Z_i(x)$ denote the i-th trajectory restricted to horizon h. We define:

$$(\mathbf{d}_y)_i = d_h\big(Z_i^{(h)}, y\big), \qquad D_{ij} = d_h\big(Z_i^{(h)}, Z_j^{(h)}\big). \tag{13}$$

where $\mathbf{d}_y$ contains distances from each scenario to the realised future and $\mathbf{D}$ contains pairwise distances between scenarios. With $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_L)^\top$, the energy score transitions into:

$$\mathrm{ES}(Q_x^h, y) = \boldsymbol{\omega}^\top \mathbf{d}_y - \frac{1}{2} \boldsymbol{\omega}^\top \mathbf{D} \boldsymbol{\omega}. \tag{14}$$

While complete-path quality is the primary objective, the induced timestamp-wise marginals should also remain well calibrated. We therefore include the coordinate-averaged CRPS:

$$\mathrm{CRPS}_{\mathrm{marg}}(\mu, y) = \frac{1}{h} \sum_{t=1}^{h} \left[\int |z_t - y_t| \, \mathrm{d}\mu(z) - \frac{1}{2} \iint |z_t - z_t'| \, \mathrm{d}\mu(z) \, \mathrm{d}\mu(z') \right]. \tag{15}$$

This evaluates the scalar CRPS at each forecast step and averages the result across all h coordinates. To prevent the predictive mass from collapsing onto only a few scenarios, we additionally regularise the entropy of the restricted distributions. For a forecast horizon h, let $\mathcal{S}_h(x)$ denote the set of distinct

scenarios after restriction and coalescence, and let $\omega_h(z)$ denote the probability mass assigned to scenario $z \in \mathcal{S}_h(x)$. We define the support-normalised entropy as

$$\mathcal{E}_h(x) = -\frac{1}{\log|\mathcal{S}_h(x)|} \sum_{z\in\mathcal{S}_h(x)} \omega_h(z)\log\omega_h(z), \qquad \bar{\mathcal{E}}(x) = \frac{1}{|\mathcal{H}|}\sum_{h\in\mathcal{H}} \mathcal{E}_h(x), \tag{16}$$

where $\mathcal{H}$ is the set of forecast horizons used during training and $|\mathcal{H}|$ is its cardinality. The quantity $\bar{\mathcal{E}}(x)$ therefore measures the mean normalised entropy across the training horizons.

The final training objective finally combines complete-path energy, marginal CRPS, and an entropy barrier:

$$\mathcal{L} = \frac{1}{2|\mathcal{H}|}\sum_{h\in\mathcal{H}} [\mathrm{ES}_h + 0.25\ \mathrm{CRPS}_{\mathrm{marg},h}] + 0.1\left[0.35 - \bar{\mathcal{E}}\right]_+^2, \tag{17}$$

where $[r]_+ = \max(r, 0)$ denotes the positive-part operator. The entropy penalty is active only when $\bar{\mathcal{E}} < 0.35$, encouraging probability mass to remain distributed across the available scenarios.

3.5 Matched control

To evaluate the complete representation of CPSS, we introduce a closely matched global-basis control, GlobalBasis-64 (GB64). GB64 uses the same frozen TimesFM-2.5 carrier, the same carrier median $m(x)$ and uncertainty scale $c(x)$, and the same budget of $L = 64$ weighted trajectories. The key difference lies in how these trajectories are constructed: rather than introducing variation progressively through suffix-supported refinements, GB64 parameterises each trajectory independently over the full forecast horizon using a global cosine basis:

$$\widetilde{Q}_x^H = \sum_{j=1}^{L} \widetilde{\omega}_j(x)\,\delta_{m(x)+c(x)\sum_{k=0}^{7}\gamma_{j,k}(x)\phi_{0,k}}, \qquad \widetilde{Q}_x^h = (R_{h:H})_{\#}\widetilde{Q}_x^H. \tag{18}$$

Here, j indexes the GB64 trajectories, $\widetilde{\omega}_j(x)$ denotes the learned probability mass assigned to trajectory j, and $\gamma_{j,k}(x)$ is its learned coefficient for the k-th full-horizon basis function $\phi_{0,k}$. The restriction operator $R_{h:H}$ is the same as that used for CPSS and retains the first h steps of each trajectory.

GB64 provides a controlled comparison between two representations of the same 64-scenario budget. CPSS shares structure across related trajectories and introduces variation progressively at later lead times, whereas GB64 parameterises each trajectory independently over the full horizon. The models are also closely matched in size, with $640,059$ trainable parameters for CPSS and $642,624$ for GB64. GB64 therefore serves as the primary structural ablation, retaining the same frozen carrier and scenario budget while replacing CPSS's hierarchical suffix-supported construction with a flat global-basis representation. This comparison isolates the benefit of CPSS's structured coarse-to-fine trajectory representation, while controlling for model capacity and scenario count.

4 Evaluation Design

4.1 Forecast horizons and held-out panels

We evaluate CPSS across a range of forecast horizons on eight standard time-series benchmarks: ETTh1, ETTh2, ETTm1, ETTm2, Electricity, Traffic, Weather, and Exchange (Zhou et al., 2021; Wu et al., 2021). We use H to denote the number of future time steps in a forecast. Since the datasets are sampled at different frequencies, the same value of H corresponds to different physical durations. For example, $H = 720$ spans 30 days for the hourly datasets, 7.5 days for the 15-minute ETT datasets, 5 days for Weather, and 720 days for Exchange. Each experiment is repeated with three training seeds.

The loss scales differ substantially across datasets, so raw scores cannot be averaged directly. We therefore compute a bounded symmetric skill score separately for each dataset and then average the resulting scores across all eight datasets. Let B_d and C_d denote the mean losses of control (GB64) and CPSS, respectively, on dataset d. The aggregate skill score is:

$$S = \left(\frac{1}{8}\sum_{d=1}^{8} \frac{(B_d - C_d)}{\frac{|B_d|+|C_d|}{2} + 10^{-12}}\right) \times 100. \tag{19}$$

Multiplying with 100 sets the reporting scale in skill points and division by 8 is used to average across the eight dataset skills equally. The constant 10^{-12} prevents numerical instability when both

losses are very small. Positive values indicate lower loss for CPSS, while negative values favour the control. We report S in skill points rather than as a % improvement and compute paired uncertainty intervals within each dataset.

We complement the horizon-sweep evaluation with two separate $H = 128$ panels spanning 12 domains. A 96-case panel is used for complete-path evaluation, path-dependent quantities, and the controlled comparison with TimeMCL, with two training seeds. A separate 192-case panel is used for timestamp-wise marginal evaluation. Additional empirical results are reported in Appendix C.

4.2 Scoring complete future distributions

A probabilistic trajectory forecast can fail at several levels. Its complete trajectories may poorly represent the realised future, distributions of path-dependent quantities may be inaccurate, or the induced marginals may perform poorly at individual forecast steps. We therefore evaluate CPSS from all three perspectives.

1. **Complete-path quality:** We first evaluate the joint distribution over complete trajectories using the energy score defined in Equation 12. For a realised future $y \in \mathbb{R}^H$, the energy score compares y with the full weighted set of predicted trajectories while also accounting for dispersion among them. This directly evaluates the quality of the complete predictive distribution rather than only its individual marginals.
2. **Path-dependent quantities:** Many real-world decisions depend on properties of an entire trajectory rather than on a single forecast step. Let $f : \mathbb{R}^H \to \mathbb{R}$ denote a scalar functional of a trajectory. Applying f to each trajectory in the predictive distribution Q_x^H induces the distribution $f_{\#}Q_x^H$, where $f_{\#}$ denotes the pushforward under f. We consider four functionals: the future maximum, temporal mean, largest absolute one-step change, and fraction of the forecast horizon spent above a pre-specified high threshold. Each induced scalar distribution is evaluated using CRPS.
3. **Timestamp-wise quality:** Finally, improvements in joint trajectory structure should not come at the expense of the individual forecast marginals. We therefore evaluate the induced timestamp-wise distributions using marginal CRPS, pinball loss, calibration error, coverage error, and the mean absolute error (MAE) of the predictive median.

Together, these three evaluation levels assess the complete future distribution, quantities derived from complete trajectories, and forecast quality at individual lead times.

5 Results

5.1 Effect of forecast horizon

CPSS is designed to organise a fixed scenario budget more effectively as the forecast horizon increases. At short horizons, 64 independently parameterised GB64 trajectories can already represent a relatively limited space of future evolutions. As the horizon grows, the same number of trajectories must capture a wider range of possible temporal behaviours. CPSS addresses this setting by introducing variation progressively through its hierarchical suffix-supported structure rather than parameterising every trajectory independently over the full horizon.

The results (in Figure 2a) show a clear horizon-dependent crossover. Using the dataset-balanced skill score defined in Equation 19, CPSS obtains -5.62 skill points at $H = 12$, -2.84 at $H = 24$, and -0.53 at $H = 48$, indicating an advantage for GB64 at short horizons. The ordering reverses at $H = 72$, where CPSS reaches $+1.42$ skill points. Its advantage then increases to $+3.35$ at $H = 96$, $+6.52$ at $H = 192$, $+7.52$ at $H = 336$, and $+8.52$ at $H = 720$.

This pattern is also reflected in the fitted horizon trend (Figure 2b). Across the eight datasets, the aggregate slope is $+2.571$ skill points for each doubling of the forecast horizon, with interval $[2.464, 2.870]$. All three training seeds produce positive slopes. From $H = 96$, CPSS achieves positive skill on all eight datasets. Although the magnitude of the improvement varies across datasets, the overall pattern is consistent, with CPSS becoming increasingly favourable relative to GB64 as the forecast horizon grows. Appendix B reports the complete dataset-by-horizon results.

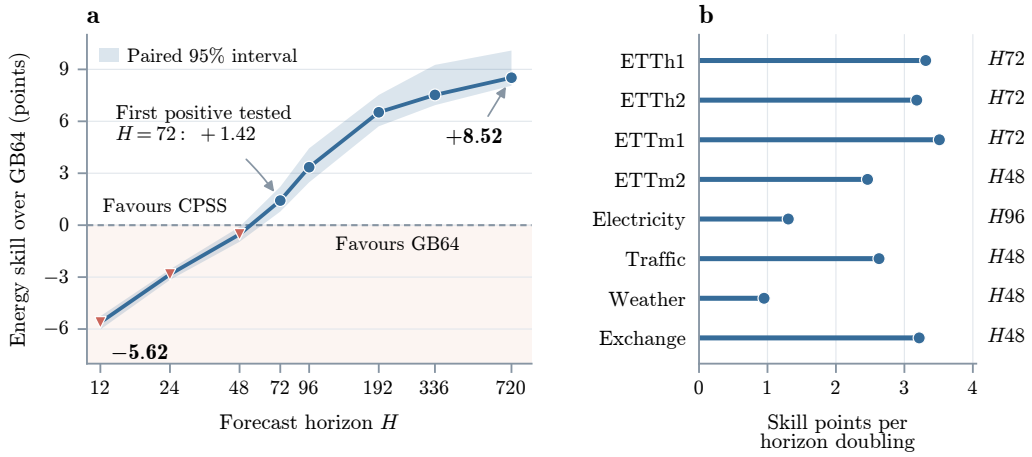

Figure 2: CPSS performance relative to GB64 as the forecast horizon increases. Panel (a) reports the dataset-balanced bounded-symmetric energy skill where positive values favour CPSS, and the shaded region denotes the paired 95% interval. Panel (b) shows the fitted skill change (in skill points per horizon doubling) labeled with the first forecast horizon achieving positive skill for each dataset.

Table 1: CPSS relative to the controlled TimeMCL carrier adapter on the 96-case, 12-domain $H = 128$ panel, with seeds 11 and 22. Values show ordinary relative loss reductions, and brackets show the archived paired 95% bootstrap intervals from resampling domains and cases within domains while holding seeds fixed.

Metric	CPSS loss reduction (%)	Paired 95% interval
Energy	+6.69	[1.96, 9.67]
Four-functional CRPS	+11.18	[2.32, 19.14]
Variogram	−5.78	[−14.49, −0.14]

5.2 Comparison with TimeMCL

The comparison with TimeMCL (Table 1) uses the case-disjoint $H = 128$ panel containing 96 forecast windows from 12 domains. Under the shared-carrier protocol, CPSS reduces complete-path Energy by 6.69% relative to TimeMCL, with a paired 95% interval of [1.96, 9.67]. CPSS also reduces the composite CRPS over the path-dependent quantities by 11.18%.

Table 2: Path-dependent probabilistic forecasting on the case-disjoint $H = 128$ panel. Values report the relative CRPS improvement of CPSS over GB64 with 95% spreads in square brackets. The added loss columns average cases and both seeds within each domain, then weight the 12 domains equally. Maximum, mean and ramp CRPS use trajectories centred by the history median and scaled by its interquartile range, while duration is a fraction. The composite averages four baseline-normalised losses, giving GB64 a reference value of one. Relative reductions are $100(1 - C/B)$, computed from unrounded losses.

Path quantity	GB64 loss	CPSS loss	CPSS CRPS improvement
Future maximum	1.3924	1.1296	+18.87% [11.92, 26.83]
Temporal mean	0.3080	0.2662	+13.57% [6.11, 18.46]
Maximum one-step ramp	1.2473	0.9475	**+24.03%** [14.84, 38.34]
High duration	0.0678	0.0605	+10.74% [−17.10, 22.61]
Equal-functional composite	1.0000	0.8320	+16.80% [10.62, 22.36]

5.3 Path-dependent forecast quality

On the case-disjoint $H = 128$ panel, CPSS improves the equally weighted path-functional CRPS composite over GB64 by 16.80%, with interval [10.62, 22.36]. The improvement is consistent across all four functionals, reaching 18.87% for the future maximum, 13.57% for the temporal mean, 24.03% for the maximum one-step ramp, and 10.74% for high duration.

5.4 MARGINAL FORECASTING QUALITY

Table 3 evaluates the timestamp-wise marginals induced by each method. CPSS achieves a pinball loss of 0.2614, median MAE of 0.6040, calibration L_1 error of 0.0441, and 80% coverage error of 0.0284. The pinball loss and median MAE are lower than those of TiRex-2 (Podest et al., 2026), Chronos-2 (Ansari et al., 2025), Moirai-2-small (Liu et al., 2026), Toto2-22M (Khwaja et al., 2026), and FlowState r1.1 (Graf et al., 2025) on this panel. The frozen TimesFM-2.5 carrier remains slightly stronger on the marginal metrics, with a pinball loss of 0.2528 and median MAE of 0.5919. Performance close to the frozen carrier indicates that the gains in complete-path forecasting are achieved while largely preserving marginal forecast quality.

Table 3: Marginal forecasting on the 12-domain $H = 128$ confirmation panel. Scores use the separate 192-case marginal-confirmation panel, with lower values indicating better performance. Bold values denote the lowest score among the displayed methods.

Method	Pinball ↓	Median MAE ↓	Calibration L_1 ↓	80% coverage error ↓
TiRex-2	0.3215	0.7431	0.0647	0.0818
Chronos-2	0.3394	0.7997	0.0692	0.0776
Moirai-2-small	0.3420	0.7814	0.0770	0.1007
Toto2-22M	0.3566	0.8250	0.0767	0.0851
FlowState r1.1	0.3687	0.8457	0.0937	0.1924
GB64	0.2660	**0.5945**	0.0722	0.1983
CPSS	**0.2614**	0.6040	**0.0441**	**0.0284**

CPSS outperforms GB64 in pinball loss (0.2614 vs. 0.2660), calibration L_1 error (0.0441 vs. 0.0722), and 80% coverage error (0.0284 vs. 0.1983), while GB64 achieves a slightly lower median MAE (0.5945 vs. 0.6040). These results provide strong evidence that CPSS improves complete-path forecasting while retaining strong marginal forecast quality.

6 DISCUSSION AND CONCLUSION

The horizon sweep (in Figure 2) reveals a clear change in the relative performance of CPSS and GB64. At short horizons, the independently parameterised global-basis trajectories of GB64 remain competitive, but CPSS consistent with it's inherent design becomes increasingly favourable as the forecast horizon increses. GB64 models each complete trajectory over the full horizon, whereas CPSS introduces additional variation progressively through hierarchical suffix-supported refinements. With the scenario budget fixed at 64 trajectories, this organisation becomes increasingly useful as the space of possible future evolutions expands. Beyond fixed-horizon accuracy, CPSS provides a cross-horizon structure in which scenarios that have not yet diverged retain identical prefixes, allowing a single forecast to define a nested family of coarse-to-fine scenario distributions.

Overall, CPSS transforms the marginal forecasts of a frozen TSFM into a weighted distribution over complete future trajectories. On the $H = 128$ evaluation, CPSS reduces complete-trajectory energy by 6.80% and the CRPS of path-dependent quantities by 16.80% relative to GB64. The improvements span the future maximum, temporal mean, maximum one-step change, and time spent above a high threshold. Under the same evaluation protocol, CPSS also reduces complete-trajectory energy by 6.69% relative to TimeMCL.

The present study considers 64 univariate scenarios with fixed refinement locations. Future work could extend this setting to adaptive refinement locations, alternative scenario budgets, multivariate trajectory distributions, and downstream sequential decision-making, providing a broader test of how well the hierarchical construction generalises. Within the scope of the current study, the results nevertheless show strong promise for hierarchical scenario representations as a scalable approach to long-horizon joint forecasting, with potential real-world applications in domains where decisions depend on coherent future trajectories rather than isolated time-step predictions.

AI-USE STATEMENT

The authors used LLMs (GPT-5.6 Sol and GPT-6 Astra) for code assistance, evidence extraction, manuscript restructuring, and prose review. The authors checked scientific claims, numerical values, citations, figures, and final wording against preserved experimental artefacts.

ETHICS STATEMENT

This work uses existing time-series benchmark datasets and does not involve human participants, personal data collection, or interventions. CPSS is intended as a methodological contribution to probabilistic forecasting; any deployment in high-stakes settings such as energy, weather, or financial decision-making would require domain-specific validation, calibration, and appropriate human oversight.

REPRODUCIBILITY STATEMENT

Appendix A specifies the datasets, forecast horizons, training procedure, aggregation rules, and uncertainty calculations used in the experiments. The model construction and training objective are given explicitly in Sections 3 and 4. Code and experiment configurations will be released with the final publication.

REFERENCES

Taha Aksu, Gerald Woo, Juncheng Liu, Xu Liu, Chenghao Liu, Silvio Savarese, Caiming Xiong, and Doyen Sahoo. GIFT-Eval: A benchmark for general time series forecasting model evaluation. *arXiv preprint arXiv:2410.10393*, 2024. doi: 10.48550/arXiv.2410.10393. URL https://doi.org/10.48550/arXiv.2410.10393. Dataset archive: https://huggingface.co/datasets/Salesforce/GiftEvalPretrain.

Abdul Fatir Ansari, Lorenzo Stella, Caner Turkmen, Xiyuan Zhang, Pedro Mercado, Huibin Shen, Oleksandr Shchur, Syama Sundar Rangapuram, Sebastian Pineda Arango, Shubham Kapoor, Jasper Zschiegner, Danielle C. Maddix, Hao Wang, Michael W. Mahoney, Kari Torkkola, Andrew Gordon Wilson, Michael Bohlke-Schneider, and Yuyang Wang. Chronos: Learning the language of time series. *Transactions on Machine Learning Research*, 2024. URL https://arxiv.org/abs/2403.07815.

Abdul Fatir Ansari, Oleksandr Shchur, Jaris Küken, Andreas Auer, Boran Han, Pedro Mercado, Syama Sundar Rangapuram, Huibin Shen, Lorenzo Stella, Xiyuan Zhang, Mononito Goswami, Shubham Kapoor, Danielle C. Maddix, Pablo Guerron, Tony Hu, Junming Yin, Nick Erickson, Prateek Mutalik Desai, Hao Wang, Huzefa Rangwala, George Karypis, Yuyang Wang, and Michael Bohlke-Schneider. Chronos-2: From univariate to universal forecasting. *arXiv preprint arXiv:2510.15821*, 2025. doi: 10.48550/arXiv.2510.15821. URL https://doi.org/10.48550/arXiv.2510.15821.

Arjun Ashok, Étienne Marcotte, Valentina Zantedeschi, Nicolas Chapados, and Alexandre Drouin. TACTiS-2: Better, faster, simpler attentional copulas for multivariate time series. In *International Conference on Learning Representations*, volume 2024, pp. 22897–22924, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/63796148c99205adb0fcac069cc714d4-Abstract-Conference.html.

Maximilian Beck, Korbinian Pöppel, Markus Spanring, Andreas Auer, Oleksandra Prudnikova, Michael Kopp, Günter Klambauer, Johannes Brandstetter, and Sepp Hochreiter. xLSTM: Extended long short-term memory. In *Advances in Neural Information Processing Systems*, volume 37, pp. 107547–107603. Curran Associates, Inc., 2024. doi: 10.52202/079017-3417. URL https://doi.org/10.52202/079017-3417.

Tianqi Chen and Carlos Guestrin. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794. ACM, 2016. doi: 10.1145/2939672.2939785. URL https://doi.org/10.1145/2939672.2939785.

Adrien Cortes, Remi Rehm, and Victor Letzelter. Winner-takes-all for multivariate probabilistic time series forecasting. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 11288–11312. PMLR, 2025. URL https://proceedings.mlr.press/v267/cortes25b.html.

Xilin Dai, Zhijian Xu, Wanxu Cai, and Qiang Xu. From samples to scenarios: A new paradigm for probabilistic forecasting. In *International Conference on Learning Representations*, volume 2026, pp. 140721–140750,

2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/e44337573fcac83f219c8effa4ebf90d-Abstract-Conference.html.

Abhimanyu Das, Weihao Kong, Rajat Sen, and Yichen Zhou. A decoder-only foundation model for time-series forecasting. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 10148–10167. PMLR, 2024. URL https://proceedings.mlr.press/v235/das24c.html.

Jitka Dupačová, Nicole Gröwe-Kuska, and Werner Römisch. Scenario reduction in stochastic programming. *Mathematical Programming*, 95(3):493–511, 2003. doi: 10.1007/s10107-002-0331-0. URL https://doi.org/10.1007/s10107-002-0331-0.

Tilmann Gneiting and Adrian E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007. doi: 10.1198/016214506000001437. URL https://doi.org/10.1198/016214506000001437.

Lars Graf, Thomas Ortner, Stanisław Woźniak, and Angeliki Pantazi. FlowState: Sampling-rate-equivariant time-series forecasting. *arXiv preprint arXiv:2508.05287*, 2025. doi: 10.48550/arXiv.2508.05287. URL https://arxiv.org/abs/2508.05287v3. Version 3, 2026. Evaluated checkpoint: r1.1, https://huggingface.co/ibm-research/flowstate.

Hassan Ismail Fawaz, Benjamin Lucas, Germain Forestier, Charlotte Pelletier, Daniel F. Schmidt, Jonathan Weber, Geoffrey I. Webb, Lhassane Idoumghar, Pierre-Alain Muller, and François Petitjean. InceptionTime: Finding AlexNet for time series classification. *Data Mining and Knowledge Discovery*, 34(6):1936–1962, 2020. doi: 10.1007/s10618-020-00710-y. URL https://doi.org/10.1007/s10618-020-00710-y.

Emaad Khwaja, Chris Lettieri, Gerald Woo, Eden Belouadah, Marc Cenac, Guillaume Jarry, Enguerrand Paquin, Xunyi Zhao, Viktoriya Zhukov, Othmane Abou-Amal, Chenghao Liu, Ameet Talwalkar, and David Asker. Toto 2.0: Time series forecasting enters the scaling era. *arXiv preprint arXiv:2605.20119*, 2026. doi: 10.48550/arXiv.2605.20119. URL https://doi.org/10.48550/arXiv.2605.20119. Evaluated model: Toto-2.0-22m.

Marcel Kollovieh, Marten Lienen, David Lüdke, Leo Schwinn, and Stephan Günnemann. Flow matching with Gaussian process priors for probabilistic time series forecasting. In *International Conference on Learning Representations*, volume 2025, pp. 95435–95457, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/ee1a1ecc92f35702b5c29dad3dc909ea-Abstract-Conference.html.

Chenghao Liu, Taha Aksu, Juncheng Liu, Xu Liu, Hanshu Yan, Quang Pham, Silvio Savarese, Doyen Sahoo, Caiming Xiong, and Junnan Li. Moirai 2.0: When less is more for time series forecasting. *arXiv preprint arXiv:2511.11698*, 2026. doi: 10.48550/arXiv.2511.11698. URL https://arxiv.org/abs/2511.11698v3. Version 3, 3 February 2026; first submitted in 2025.

Mahbod Majid, Rattana Pukdee, Vishwajeet Agrawal, Burak Varıcı, and Pradeep Kumar Ravikumar. On the consistent recovery of joint distributions from conditionals. In *Proceedings of the 28th International Conference on Artificial Intelligence and Statistics*, volume 258 of *Proceedings of Machine Learning Research*, pp. 4708–4716. PMLR, 2025. URL https://proceedings.mlr.press/v258/majid25a.html.

Patrick Podest, Marco Pichler, Elias Bürger, Levente Zólyomi, Bernhard Voggenberger, Wilhelm Berghammer, Daniel Klotz, Sebastian Böck, Günter Klambauer, and Sepp Hochreiter. TiRex-2: Generalizing TiRex to multivariate data and streaming. *arXiv preprint arXiv:2607.01204*, 2026. doi: 10.48550/arXiv.2607.01204. URL https://doi.org/10.48550/arXiv.2607.01204.

Gerald Woo, Chenghao Liu, Akshat Kumar, Caiming Xiong, Silvio Savarese, and Doyen Sahoo. Unified training of universal time series forecasting transformers. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 53140–53164. PMLR, 2024. URL https://proceedings.mlr.press/v235/woo24a.html.

Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In *Advances in Neural Information Processing Systems*, volume 34, pp. 22419–22430. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper/2021/hash/bcc0d400288793e8bdcd7c19a8ac0c2b-Abstract.html.

Vijaya Krishna Yalavarthi, Randolf Scholz, Christian Klötergens, Kiran Madhusudhanan, Stefan Born, and Lars Schmidt-Thieme. Reliable probabilistic forecasting of irregular time series through marginalization-consistent flows. In *International Conference on Learning Representations*, volume 2026, pp. 85403–85427, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/899c6a43b9976e1077522fe5a39cafa3-Abstract-Conference.html.

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 *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 11106–11115. Association for the Advancement of Artificial Intelligence (AAAI), 2021. doi: 10.1609/aaai.v35i12.17325. URL https://doi.org/10.1609/aaai.v35i12.17325.

APPENDIX A EXPERIMENTAL DETAILS

A.1 DATA AND EVALUATION PANELS

We use one panel for horizon-sweep panel and two separate H128 panels for confirmation. The horizon sweep contains 1,480 context-512 cases from ETTh1, ETTh2, ETTm1, ETTm2 (Zhou et al., 2021), Electricity, Traffic, Weather and Exchange (Wu et al., 2021), with seeds 11, 22 and 33. Eight forecast origins per dataset are crossed with the selected channels in Table A.1. Origins lie within the source test split and exclude previously used origin identifiers. H12–H96 restrict the H96 forecasts; H192, H336 and H720 use separately trained terminal heads. Horizons count native time steps: H720 spans 30 days for hourly series, 7.5 days for ETTm, five days for Weather and 720 days for Exchange.

The H128 joint, path-functional and TimeMCL comparisons use 96 cases, eight per domain, with seeds 11 and 22. The marginal benchmark uses a separate 192-case panel, with 16 cases per domain. Both panels use contexts 128, 256 or 512 and contain Australian Electricity Demand, Bitcoin with missing values, CDC FluView ILINet, COVID-19 Energy, GFC17 Electricity Load, NN5 Daily with missing values, Oikolab Weather, Pedestrian Counts, Residential Load Power, Tourism Monthly, Traffic Hourly and Wind Power from the GIFT-Eval Pretrain collection (Aksu et al., 2024). Their case identifiers (domain, series, channel, forecast origin and context) are disjoint, although windows can overlap in time.

Table A.1: Sampling configuration for the long-horizon benchmark. All forecasts use context length 512 and eight forecast origins per selected channel.

Dataset	Sampling interval	Channels (total/used)	Cases
ETTh1	1 hour	7/7	56
ETTh2	1 hour	7/7	56
ETTm1	15 minutes	7/7	56
ETTm2	15 minutes	7/7	56
Electricity	1 hour	321/64	512
Traffic	1 hour	862/64	512
Weather	10 minutes	21/21	168
Exchange	1 day	8/8	64

A.2 MODEL AND TRAINING DETAILS

The horizon-sweep and H128 experiments use their respective training configurations. Within each comparison, competing heads receive the same carrier information and optimisation schedule. Only the scenario head is trained while TimesFM-2.5 remains frozen. Table A.2 summarises the model and optimisation settings.

The horizon-sweep loss is Equation 17 averaged over training history–future pairs. The TimeMCL adapter retains its relaxed winner-takes-all objective.

For H128 functional scores, paths and outcomes are centred by the history median b_x and scaled by $s_x = \max\{\mathrm{IQR}(x), 10^{-6}\max(1, |b_x|)\}$. The functionals are maximum, mean, maximum absolute one-step ramp and the fraction above 1 in these normalised units. The composite averages their CRPS values after dividing each by its GB64 equal-domain mean loss. Variogram scoring uses power $1/2$ and lags 1, 4, 16 and 32, weighted by inverse lag.

Table A.2: Scenario-head configurations. Locations and training restrictions are multiplied by $H/128$ and rounded half upward for terminal horizon H.

Setting	Configuration
Training data	GiftEvalPretrain: 36 domains for the horizon sweep; 37 for H128
Training contexts	Horizon sweep: 96, 128, 256, 512; H128: 128, 256, 512
Updates and batches	1,000 updates; 16 windows per batch; final checkpoint evaluated
Optimiser	AdamW; $(\beta_1, \beta_2) = (0.9, 0.95)$; weight decay 0.01; gradient-norm clipping at 1
Learning rate	3×10^{-4} peak, 50-update warm-up, cosine decay to 3×10^{-5}
Conditioner inputs	64 resampled history values and 32 values from each carrier quantile $q_{0.1}$, $q_{0.5}$ and $q_{0.9}$: 160 inputs
Conditioner network	Two width-512 affine layers, each with SiLU and layer normalisation
CPSS support and basis	64 leaves, 127 nodes; cosine frequencies $k = 0, 1, 2, 3$ per node with tanh-bounded coefficients
Suffix starts	$(0, 4, 8, 16, 32, 64, 96)\ H/128$
Training restrictions	$(16, 32, 64, 96, 128)\ H/128$
Basis normalisation	$\kappa_{a,0} = (H-a)^{-1/2}$; $\kappa_{a,k} = \sqrt{2/(H-a)}$ for $k > 0$
GB64 decoder	Eight full-horizon cosine components per path; flat softmax masses
Trainable parameters	CPSS: 640,059 and GB64: 642,624

Horizon-sweep losses average cases and seeds within dataset before applying Equation 19 and weighting the eight datasets equally. H128 reductions are $100(1 - \overline{C}/\overline{B})$, with risks averaged within domain and then equally across domains. H128 intervals use 10,000 paired resamples of domains followed by cases within domains; horizon-sweep intervals resample paired cases within dataset. All keep the training seeds as fixed.

APPENDIX B HORIZON RESULTS

B.1 COMPLETE FORECAST EXAMPLE

Figure B.1 shows a representative H720 ETTm1 forecast selected using only the observed history and forecast statistics.

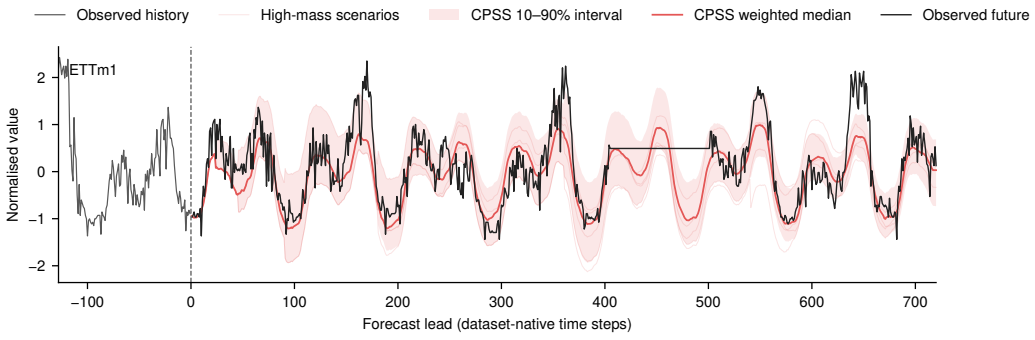

Figure B.1: A target-blind complete H720 CPSS forecast from ETTm1. The plot shows the final 128 observed context values, five high-mass trajectories, the weighted 10–90% interval, the CPSS weighted marginal median, and the subsequently revealed future. The case was selected without target access and illustrates recurring curves, changing path spread, and ill-timed peaks.

B.2 DATASET-WISE RESULTS

Table B.1 reports the complete CPSS-versus-GB64 energy sequence. Positive values favour CPSS. Every dataset has a positive fitted slope per horizon doubling and a positive long-minus-short

Table B.1: Dataset-wise bounded-symmetric energy skill of CPSS over GB64.

Dataset	H12	H24	H48	H72	H96	H192	H336	H720
ETTh1	-5.56	-3.85	-1.26	+1.60	+3.51	+7.76	+10.43	+12.06
ETTh2	-5.92	-2.95	-1.14	+0.55	+3.32	+8.36	+10.78	+10.61
ETTm1	-6.18	-3.04	-2.20	+0.22	+3.22	+8.09	+10.91	+12.92
ETTm2	-4.36	-1.44	+0.59	+1.98	+3.12	+5.81	+7.73	+10.19
Electricity	-7.18	-3.85	-2.00	-0.36	+0.84	+4.32	≈ 0	≈ 0
Traffic	-4.47	-1.96	+0.53	+3.02	+6.08	+8.44	+9.69	+9.07
Weather	-6.86	-2.88	≈ 0	≈ 0	≈ 0	≈ 0	≈ 0	≈ 0
Exchange	-4.40	-2.76	+1.25	+4.37	+6.70	+9.39	+10.65	+13.29
Equal-dataset mean	-5.62	-2.84	-0.53	+1.42	+3.35	+6.52	+7.52	+8.52

difference. A leave-one-dataset-out analysis retains a positive aggregate slope under every omission. Weather and late Electricity effects remain numerically small, therefore breadth reflects directional trend rather than a uniformly large gain.

B.3 Aggregate uncertainty intervals

Table B.2: CPSS energy skill over GB64 (out of eight datasets) with paired within-dataset case-bootstrap 95% intervals, holding seeds fixed. Positive values favour CPSS.

Horizon	Skill (points)	95% interval
H12	-5.617	$[-6.026, -5.257]$
H24	-2.842	$[-3.163, -2.557]$
H48	-0.529	$[-0.960, -0.095]$
H72	$+1.422$	$[0.785, 2.221]$
H96	$+3.349$	$[2.475, 4.441]$
H192	$+6.522$	$[5.711, 7.523]$
H336	$+7.524$	$[6.931, 9.259]$
H720	$+8.518$	$[8.055, 10.088]$

Table B.2 gives the aggregate skill and paired 95% intervals. H72 is the first tested horizon with positive aggregate skill. H96 is the first with positive point estimates on all eight datasets.

Appendix C Additional Empirical Results

C.1 TimeMCL at H128

Table C.1: CPSS relative to the TimeMCL carrier adapter on the 96-case, 12-domain $H = 128$ panel. Values are relative loss reductions; positive values favour CPSS.

Metric	CPSS loss reduction (%)	95% interval
Energy	$+6.69$	$[1.96, 9.67]$
Four-functional CRPS	$+11.18$	$[2.32, 19.14]$

On the 96-case H128 panel, CPSS reduces energy by 6.69% (paired 95% interval $[1.96, 9.67]$) and path-functional CRPS by 11.18% relative to the TimeMCL carrier adapter. Both methods receive the same carrier-derived inputs and produce 64 complete trajectories (Table C.1).

C.2 COMPLEMENTARY EVALUATIONS

As an additional point-forecast comparison, we evaluate the weighted marginal median of CPSS at $H = 720$ (shown in Table C.2). It reduces dataset-balanced MAE relative to seasonal ARIMA, LightGBM, XGBoost (Chen & Guestrin, 2016), xLSTM (Beck et al., 2024) and an InceptionTime-style regressor (Ismail Fawaz et al., 2020) by 15.07%, 32.06%, 32.02%, 10.91% and 20.16%, respectively. This indicates that modelling a distribution over complete trajectories does not come at the expense of competitive deterministic forecasting.

Table C.2: H720 point-forecast comparison using the CPSS weighted marginal median. Values are dataset-balanced relative MAE reductions; positive values favour CPSS.

Comparator	MAE reduction (%)	95% interval
Seasonal ARIMA	+15.07	$[13.40, 23.07]$
LightGBM	+32.06	$[29.27, 40.17]$
XGBoost	+32.02	$[29.15, 40.39]$
xLSTM	+10.91	$[8.40, 14.73]$
InceptionTime-style regressor	+20.16	$[18.37, 28.55]$

The complete predictive distribution also improves the prospective decision benchmark relative to GB64. Across allocation, exceedance-duration, peak and ramp decisions at $H = 192$, $H = 336$ and $H = 720$, aggregate regret decreases from 4.401 for GB64 to 4.107 for CPSS, corresponding to a 6.69% reduction with 95% interval $[2.64, 39.89]$.

CPSS also retains substantial scenario diversity at $H = 720$. Its mean effective support is 43.21 of 64 scenarios, compared with 37.61 for GB64; normalised weight entropy is 0.904 versus 0.870, and weighted pairwise path distance is 0.833 versus 0.338. These diagnostics therefore indicate that CPSS uses its finite scenario set to represent distinct possible futures rather than concentrating most probability mass on a small number of near-identical trajectories.