ORCHESTRA: WHEN THE BEST TEACHER IS NOT THE BEST GUIDE IN MULTI-TEACHER TIME SERIES DISTILLATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Knowledge distillation (KD) can transfer the complementary strengths of time-series forecasting models to a compact student. However, temporal patterns differ across forecasting tasks, so the best-performing teacher may change with context, and even its guidance may be difficult for the current student to imitate. Choosing or combining multiple teachers does not by itself determine how strongly the student should learn each teacher's temporal signals. To address these challenges, we propose ORCHESTRA, a multi-teacher distillation framework that separates learning into credit, trust, and execution. A reinforcement-learned Teacher-Credit Agent uses the *Adaptive Credit Policy* (ACP) to learn context-dependent teacher credit. The *Multi-Scale Trust Estimator* (MSTE) calibrates point, scale, and first-difference supervision based on student–teacher discrepancies, while the *Sparse Execution Policy* (SEP) converts learned credit into deterministic sparse distillation weights. Extensive experiments on seven forecasting datasets and four prediction lengths show that ORCHESTRA consistently improves the compact TSMixer student, achieves the best overall MSE and MAE among the evaluated direct forecasters, and retains efficient student-only deployment.

1 INTRODUCTION

Time-series forecasting supports decisions in power grids, renewable energy, and weather monitoring (Zhou et al., 2021; Wu et al., 2021), and the series in these domains differ widely in structure. Meteorological indicators influence one another, solar output climbs from zero at sunrise and returns to zero at sunset, electricity load repeats daily and weekly cycles, and the oil temperature of power transformers drifts over months. To capture such structure, deep forecasters build in different inductive biases, that is, different assumptions about which structure in a series matters most. iTransformer treats each variate as a token to model dependencies among variables (Liu et al., 2024); PatchTST splits each series into patches to learn local shapes (Nie et al., 2023); TimeMixer mixes information across temporal scales to combine fine-scale variation with coarse trends (Wang et al., 2024); and ModernTCN uses large convolutional kernels to cover long-range context (Luo & Wang, 2024). An inductive bias helps most when the structure it assumes dominates the data, so no single model performs best across all tasks, as benchmark studies (Qiu et al., 2024) and our teacher pool (Figure 1(b)) both show. These models are therefore complementary, each capturing structure that the others model less well. This raises a natural question: can a single forecaster inherit their complementary strengths?

Knowledge distillation (KD) offers a way to do so: it trains a compact student to reproduce the predictions of stronger models, and was originally proposed to compress an ensemble into a single model (Hinton et al., 2015; Gou et al., 2021). Later work extended the transferred knowledge from outputs to intermediate features and attention maps (Romero et al., 2015; Zagoruyko & Komodakis, 2017), and recent forecasting methods adapt KD to temporal data, transferring multi-scale and multi-period patterns, regime-aware structure, privileged representations, and the knowledge of time-series foundation models into lightweight students (Ni et al., 2026; Hong et al., 2026; Liu et al., 2025; Li et al., 2026). These methods refine how knowledge is transferred from a teacher, but the teacher itself is fixed in advance. Given a pool of complementary forecasters, one must therefore decide which of them to distill from, and the natural answer is the most accurate one for the task. This answer turns out to be unreliable. In our experiments, distilling from the most accurate teacher 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.

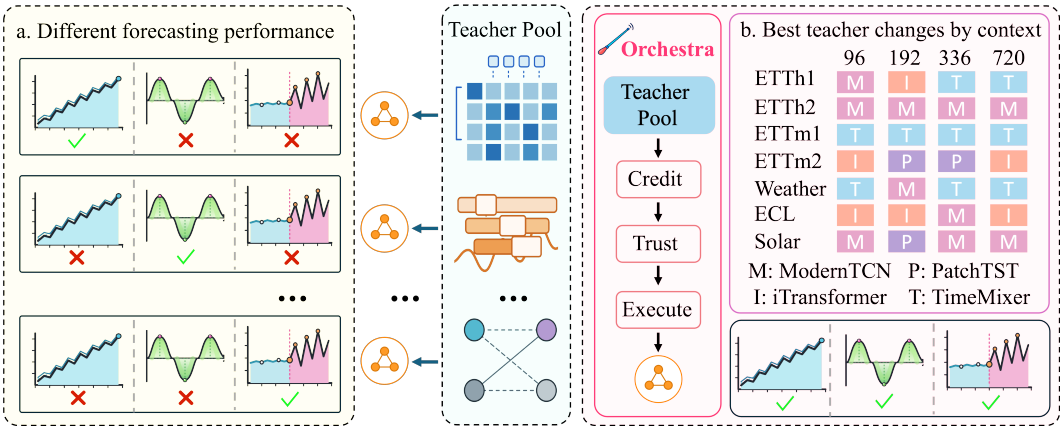


Figure 1: **Complementary forecasters motivate multi-teacher distillation. (a)** Schematic: each forecaster captures some temporal patterns and misses others, so no single model covers all of them. **(b)** The most accurate frozen teacher in each of the 28 dataset–horizon settings. ModernTCN, TimeMixer, iTransformer, and PatchTST lead in 10, 9, 6, and 3 settings, respectively, and the leader changes across horizons even within one dataset, with three different leaders on ETTh1. Right: ORCHESTRA distills the pool into a single student through credit, trust, and execution, and only the student is deployed.

leave the student worse than one that learns from all teachers with equal weights, and even worse than one trained without any teacher (Table 2). The most accurate teacher is not necessarily the most useful one.

If no single teacher can be relied on, the natural next step is to learn from the whole pool. The simplest multi-teacher KD averages the teachers' outputs (You et al., 2017), and more refined schemes weight each teacher by how reliable it appears, for example by its confidence with respect to the ground truth (Zhang et al., 2022) or, in time-series forecasting, by its predictive uncertainty (Dey et al., 2026). Averaging is a reasonable default, and it already outperforms the most accurate single teacher in the cases above, but it treats all teachers as interchangeable and blends away the distinctive strengths that motivated using a pool. Reliability-based weighting lets stronger teachers count more, yet it still ranks teachers by how good they are, which is the criterion that has just failed. Studies of the teacher–student capacity gap explain why this criterion falls short: a student learns best from supervision it can actually follow, so a target far from its current predictions can be harder to absorb than a closer but less accurate one (Cho & Hariharan, 2019; Mirzadeh et al., 2020). The question therefore shifts from *which teacher is better* to *which teacher the current student can learn from*. Because the student keeps changing during training, the answer must be revisited as training proceeds rather than fixed in advance.

Methods that adapt the teachers during training come closer to this goal. They learn instance-level teacher weights (Liu et al., 2020; Zhang et al., 2023), sample a teacher from a learned distribution (Ding et al., 2024), select teachers through reinforcement learning (Yuan et al., 2021; Xu et al., 2024a), or fuse teachers with a learned gate (Song et al., 2026). Yet they all decide how much to learn from each teacher, not from each part of what a teacher predicts. For forecasting, this granularity is too coarse, because a forecast carries several kinds of temporal information: the level of future values, their trend over longer spans, and their step-to-step movement. Forecasting KD already benefits from transferring multi-scale and multi-period patterns in addition to pointwise values (Ni et al., 2026), and the inductive biases that make the teachers complementary also shape these aspects unevenly. A student may therefore follow a teacher's level closely while struggling with its movement. Under a single weight per teacher, the student must take both with the same strength, either absorbing the hard-to-follow movement along with the useful level or giving up both. Transfer should instead be weighed for each teacher and each temporal signal, and updated as the student learns.

We propose ORCHESTRA to make these decisions during training. It splits the transfer from a pool of frozen teachers into three steps: *credit*, *trust*, and *execution*. Credit decides which teachers the student should learn from on the current input. A lightweight policy reads each teacher's validation accuracy together with how the teachers behave on that input, such as how much they disagree and

how sharply their forecasts move, and it is trained with a reward for choosing teachers that the student can currently follow. Its preference therefore changes with both the input and the student. Trust decides how strongly the level, trend, and movement of each teacher are learned, giving more weight to the signals on which the student currently follows that teacher closely. Execution turns the credit into sparse distillation weights, so that teachers with low credit are excluded from the update entirely instead of being blended in with small weights. The teachers, the policy, and the trust estimator are used only during training, and at inference the student runs alone. This design brings three advantages: (i) transfer adapts to the student's learning state rather than to a fixed ranking of teachers, which is exactly where selection by accuracy fails; (ii) the student can take from each teacher only the temporal information it is able to use, instead of accepting or rejecting a teacher's forecast as a whole; and (iii) deployment is unaffected, since ORCHESTRA changes how the student is trained but not what is deployed, so the final forecaster keeps the size and speed of a compact student.

Our contributions are as follows:

- We show that selecting teachers by accuracy is unreliable in multi-teacher time-series distillation, and reformulate the problem around whether the current student can follow each teacher on each temporal signal.
- We propose ORCHESTRA, which decomposes multi-teacher distillation into credit, trust, and execution: a credit policy trained with a reward for teachers the student can follow, a trust estimator that weighs the level, trend, and movement of each teacher, and a sparse execution rule that excludes low-credit teachers from each update.
- On seven benchmarks and four forecast horizons, ORCHESTRA lowers the error of the same compact student on every dataset compared with training without teachers, uniform multi-teacher distillation, and distillation from the most accurate teacher, and this ordering holds in each of three random seeds. Replacing learned credit with a static prior, replacing sparse execution with dense weights, or removing trust each increases the error on every dataset.

2 RELATED WORK

2.1 EXPERIMENTAL SETUP

Forecasting architectures and heterogeneous teachers. Recent forecasters use different inductive biases. Informer and Autoformer explore sparse attention and decomposition; FEDformer and TimesNet emphasize frequency or temporal variation; DLinear provides a linear reference (Zhou et al., 2021; Wu et al., 2021; Zhou et al., 2022; Wu et al., 2023; Zeng et al., 2023). PatchTST learns from channel-independent patches, iTransformer treats variates as tokens, TimeMixer mixes information across temporal scales, and ModernTCN uses wide-receptive-field temporal convolutions (Nie et al., 2023; Liu et al., 2024; Wang et al., 2024; Luo & Wang, 2024). TSMixer is a compact all-MLP backbone with separate temporal and channel mixing (Chen et al., 2023). These complementary designs motivate a teacher *pool*. ORCHESTRA keeps the teachers frozen and learns how their supervision trains the TSMixer student, which alone is deployed.

Single-teacher distillation for time-series forecasting. Most forecasting KD methods begin with a preselected teacher and enrich the temporal supervision received by the student. TimeDistill transfers multi-scale and multi-period patterns; Beyond Point Predictions combines manifold alignment with regime-aware supervision; SDKD aligns frequency-sensitive representations; TimeKD uses privileged representations; and DistilTS balances horizon difficulty while aligning foundation-model representations (Ni et al., 2026; Hong et al., 2026; Li et al., 2025; Liu et al., 2025; Li et al., 2026). These methods refine what knowledge a given teacher transfers, but do not address how a pool of heterogeneous direct forecasters should contribute to each student update.

Multi-teacher and selective distillation for time series. A smaller, task-specific line considers more than one teacher. Reinforced KD constructs a negatively correlated teacher ensemble and uses DDQN for teacher selection in time-series regression; MCTSN combines two teachers through meta-contrastive adaptation and gated fusion; ThermoStill transfers guidance from several time-series foundation models to an HVAC dynamics student; Guard routes among foundation-model teachers with an uncertainty gate for scientific forecasting; and TIPS distils bias-specialized teachers for financial forecasting (Xu et al., 2024a; Song et al., 2026; Liang et al., 2026; Dey et al., 2026; Den

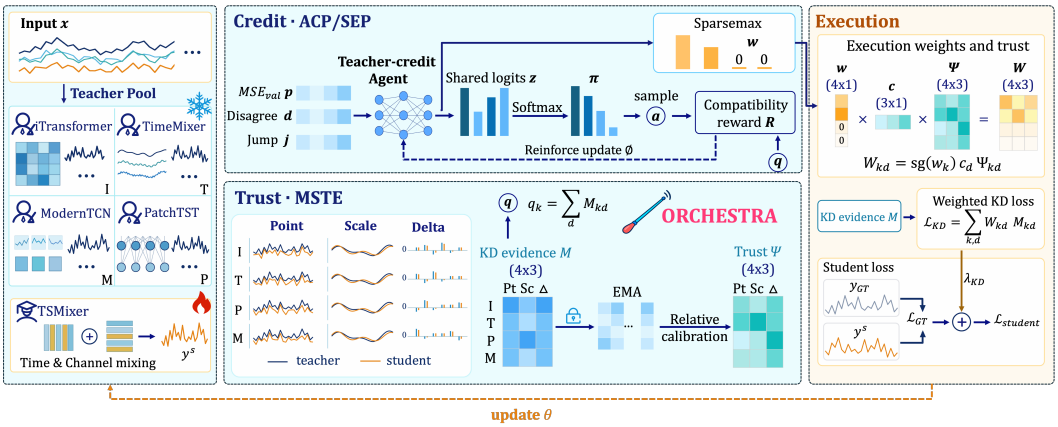


Figure 2: **ORCHESTRA training pipeline.** Four frozen teachers and the current TSMixer student yield the 4×3 KD-evidence matrix $\mathbf{M}$. **Credit:** the Teacher-Credit Agent maps validation-MSE, disagreement, and temporal-jump features to shared logits $\mathbf{z}$; ACP samples a for the REINFORCE credit update, whereas Sparsemax of the same logits yields deterministic execution weights $\mathbf{w}$. **Trust:** MSTE transforms only $\mathbf{M}$ into the relative-trust multiplier $\bar{\Psi}$; $q_k = \sum_d M_{k,d}$ bypasses MSTE and scores ACP compatibility. **Execution:** $\mathbf{W}$ combines $\mathrm{sg}(\mathbf{w})$, fixed signal coefficients $\mathbf{c}$, and $\bar{\Psi}$ to weight KD evidence. The dashed paths and locks denote parameter-specific updates and stop-gradient operations.

et al., 2026). Adjacent time-series work studies adaptive ensemble distillation and reinforcement-guided selection under domain shift (Campos et al., 2023; Xu et al., 2024b). However, these approaches still adapt mainly at the teacher or sample level: selecting a source, fusing teacher outputs, or gating examples. ORCHESTRA instead separates source credit from teacher–signal transfer, calibrating each teacher's temporal evidence before sparsely executing the loss over a fixed heterogeneous pool of direct forecasters.

3 ORCHESTRA

3.1 PROBLEM SETUP AND TEMPORAL EVIDENCE

Figure 2 provides an overview of the three training-time paths. For a mini-batch of input histories $\mathbf{x} \in \mathbb{R}^{B\times L\times C}$ and targets $\mathbf{y} \in \mathbb{R}^{B\times H\times C}$, let $\hat{\mathbf{y}}^s = f_\theta(\mathbf{x})$ be the student forecast. A frozen pool of $K = 4$ heterogeneous teachers produces $\{\hat{\mathbf{y}}^k = g_k(\mathbf{x})\}_{k=1}^K$. The supervised forecast loss is

$$\mathcal{L}_{\mathrm{GT}} = \frac{1}{BCH} \|\hat{\mathbf{y}}^s - \mathbf{y}\|_2^2 . \tag{1}$$

The distillation evidence contains three temporal *signals*; we reserve "channel" for the physical variates indexed by c. Let P_ℓ denote ℓ repeated temporal average-pooling operations, each with kernel and stride two, and let $\Delta\mathbf{u}_{b,t,c} = \mathbf{u}_{b,t+1,c} - \mathbf{u}_{b,t,c}$. For every frozen teacher we construct a $K \times 3$ discrepancy matrix on the current mini-batch,

$$M_{k,\mathrm{pt}} = \mathrm{MSE}(\hat{\mathbf{y}}^s, \hat{\mathbf{y}}^k), \tag{2}$$

$$M_{k,\mathrm{sc}} = \frac{1}{3}\sum_{\ell=1}^{3} \mathrm{MSE}\big(P_\ell(\hat{\mathbf{y}}^s), P_\ell(\hat{\mathbf{y}}^k)\big) , \tag{3}$$

$$M_{k,\Delta} = \mathrm{MSE}\big(\Delta\hat{\mathbf{y}}^s, \Delta\hat{\mathbf{y}}^k\big) , \qquad \mathbf{M} \in \mathbb{R}^{K\times 3}. \tag{4}$$

The point channel preserves forecast level, the scale channel tests agreement after successively coarsening the horizon, and the delta channel tests one-step temporal movement. Thus the three signals answer different questions about a teacher's compatibility with the *current* student; they are not three redundant copies of pointwise MSE. Their fixed signal coefficients are $\mathbf{c} = (c_{\mathrm{pt}}, c_{\mathrm{sc}}, c_\Delta) = (0.5, 0.25, 0.25)$.

3.2 TEACHER-CREDIT AGENT AND CONTEXTUAL CREDIT

The TCA maps a deliberately low-dimensional teacher-behaviour state to shared logits. Its state has three four-dimensional blocks,

$$\mathbf{s} = [\mathbf{p}; \mathbf{d}; \mathbf{j}] \in \mathbb{R}^{12}, \qquad \mathbf{z} = h_\phi(\mathbf{s}) \in \mathbb{R}^K, \tag{5}$$

The prior $p_k = e_k/(K^{-1}\sum_j e_j + \epsilon)$ is the frozen teacher-k element-weighted MSE against ground truth over the *complete validation split*; it is computed once before training and is used only in ACP. It is an error descriptor, not a manually imposed routing weight. To make the two dynamic blocks comparable across non-stationary windows, define $\bar{\mathbf{y}} = K^{-1}\sum_k \hat{\mathbf{y}}^k$ and

$$\tilde{d}_k = \frac{\operatorname{mean}\left[(\hat{\mathbf{y}}^k - \bar{\mathbf{y}})^2\right]}{\operatorname{Var}(\bar{\mathbf{y}}) + \epsilon}, \qquad \tilde{j}_k = \frac{\operatorname{mean}\left[\left|\Delta\hat{\mathbf{y}}^k\right|\right]}{\operatorname{Std}(\hat{\mathbf{y}}^k) + \epsilon}. \tag{6}$$

The fixed prior is mean-normalized as stated above. Each *dynamic* K-vector is first mean-normalized and then passed through $\mathcal{N}(v)_k = (v_k - \min_j v_j)/(\max_j v_j - \min_j v_j)$, returning the zero vector when the range is numerically zero; $\mathbf{d} = \mathcal{N}(\tilde{\mathbf{d}})$ and $\mathbf{j} = \mathcal{N}(\tilde{\mathbf{j}})$. This within-batch min–max normalization preserves the relative disagreement and volatility ordering while removing window-to-window magnitude changes. The implementation uses a LayerNorm–MLP with widths $12 \to 64 \to 64 \to 4$. Crucially, $\mathbf{M}$ and the MSTE statistics are excluded from this state, so validation quality and teacher-output geometry inform credit, whereas student–teacher KD evidence informs temporal trust.

At epoch e of a run with E training epochs, the shared logits use the annealed temperature $\tau_e = \tau_0 \max\{\tau_{\min}, 1 - \gamma e/E\}$. This gives a diffuse early policy and sharper credit and execution distributions later in training.

Adaptive Credit Policy (ACP). ACP is a context-conditioned stochastic teacher-credit policy. Following the score-function estimator of Williams (1992), it samples

$$a \sim \pi_\phi(\cdot \mid \mathbf{s}) = \operatorname{Categorical}(\operatorname{softmax}(\mathbf{z}/\tau_e)). \tag{7}$$

After a one-epoch policy warm-up, ACP is updated every U mini-batches from the current training batch. Let $q_k = \sum_{d\in\mathcal{D}} M_{k,d}$ denote teacher k's *unweighted* total student–teacher discrepancy on that batch. The complete vector $\{q_k\}_{k=1}^K$ is available whenever ACP is updated because $\mathbf{M}$ has already been constructed for the student loss. ACP therefore uses complete current-batch compatibility evidence to update credit. Here $\mathbf{M}$ is excluded from the ACP state and is used only to score the sampled credit decision. The action-quality component normalizes the sampled teacher within the contemporaneous pool:

$$u(a) = \begin{cases} \dfrac{q_{\text{worst}} - q_a}{q_{\text{worst}} - q_{\text{best}}}, & q_{\text{worst}} - q_{\text{best}} > \epsilon, \\ 0.5, & \text{otherwise}, \end{cases} \qquad q_{\text{best}} = \min_k q_k, \quad q_{\text{worst}} = \max_k q_k. \tag{8}$$

Thus $u(a) \in [0,1]$ rewards a sampled teacher that is currently easier for the student to match; it is a compatibility score rather than an estimate of future target-error reduction. Figure 2 depicts $\mathbf{q}$ as a separate return path to the compatibility reward: it is formed directly from $\mathbf{M}$ and never enters MSTE. The same shared logits independently form the deterministic companion $\mathbf{w} = \operatorname{sparsemax}(\mathbf{z}/\tau_e)$, which is combined with temporal trust only in Equation 17. The implementation forms, under stop-gradient, the scalar

$$R(a, \mathbf{s}) = \alpha\big(u(a) - 0.5\big) + \beta\ \operatorname{sg}\left[\frac{\mathcal{H}(\mathbf{w})}{\log K}\right], \tag{9}$$

where $\mathcal{H}(\mathbf{w}) = -\sum_k w_k \log(w_k + 10^{-8})$. In the reported routing configuration, $(\alpha, \beta) = (0.5, 0.3)$. The first term supplies action-dependent compatibility credit. The second is a detached, action-independent SEP-entropy offset recorded in the scalar reward. Conditional on $\mathbf{s}$, it assigns the same value to all sampled actions and does not regularize SEP or determine its sparsity.

A momentum-ρ_b EMA baseline and a clipped advantage stabilize the score estimator:

$$A_t = \operatorname{clip}(R_t - b_{t-1}, -c_A, c_A), \qquad b_t = \rho_b b_{t-1} + (1-\rho_b)R_t, \qquad b_{-1} = 0. \tag{10}$$

The agent loss is

$$\mathcal{L}_{\text{ACP}} = -A \log \pi_\phi(a \mid \mathbf{s}) - \eta \mathcal{H}(\pi_\phi(\cdot \mid \mathbf{s})), \tag{11}$$

where $(\rho_b, c_A, \eta) = (0.99, 0.1, 10^{-5})$ for the reported routing configuration. The second term is the direct entropy regularizer for the *stochastic ACP policy*; it encourages exploration in ACP and makes no direct claim about the deterministic SEP support.

3.3 TRUST-CALIBRATED SPARSE EXECUTION

A source-level credit decision does not determine whether the current student can smoothly absorb every temporal signal from a selected teacher. For example, a teacher can agree on forecast level but disagree on temporal movement. MSTE therefore receives only the $K \times 3$ KD evidence in Equation 4; in particular, it receives neither the validation-MSE ACP prior nor ACP probabilities. The lower branch of Figure 2 makes this boundary explicit: it transforms the full $\mathbf{M}$ into teacher–signal trust, whereas the separately displayed $q_k = \sum_d M_{k,d}$ bypasses MSTE and scores only the ACP reward. Let t index mini-batches. MSTE is a detached state estimator: it updates from $\mathrm{sg}(\mathbf{M})$ and returns a fixed trust multiplier for the student loss,

$$\bar{M}^{(0)} = \mathrm{sg}\Big(M^{(0)}\Big), \qquad \bar{M}^{(t)} = 0.8\bar{M}^{(t-1)} + 0.2\,\mathrm{sg}\Big(M^{(t)}\Big) \quad (t \geq 1). \tag{12}$$

For each temporal signal d, it standardizes across teachers and converts smaller student–teacher discrepancy into larger relative trust, using $T_r = 0.5$ in the reported configuration,

$$\zeta_{k,d}^{(t)} = \frac{\bar{M}_{k,d}^{(t)} - K^{-1}\sum_j \bar{M}_{j,d}^{(t)}}{\mathrm{Std}_j(\bar{M}_{j,d}^{(t)}) + \epsilon}, \qquad G_{k,d}^{(t)} = K \cdot \mathrm{softmax}_k\Big(-\zeta_{k,d}^{(t)}/T_r\Big). \tag{13}$$

The factor K changes a simplex probability (column mean $1/K$) into a relative multiplier (column mean 1 before the next step). MSTE therefore calibrates relative student–teacher compatibility over temporal signals rather than absolute teacher reliability or ground-truth error. It then preserves a neutral mean for every teacher across the three signals,

$$\widetilde{\Psi}_{k,d}^{(t)} = \frac{G_{k,d}^{(t)}}{|\mathcal{D}|^{-1}\sum_{d' \in \mathcal{D}} G_{k,d'}^{(t)} + \epsilon}. \tag{14}$$

To avoid amplifying small numerical differences, we use an empirical low-separation guardrail rather than a statistical tie test. Let $r^{(t)} = \max_k \mathrm{mean}_d\, \bar{M}_{k,d}^{(t)} - \min_k \mathrm{mean}_d\, \bar{M}_{k,d}^{(t)}$ and $\rho^{(t)} = \min\{1, r^{(t)}/\gamma\}$, with $\gamma = 0.01$ in the reported configuration. The implementation replaces $\widetilde{\Psi}^{(t)}$ by $\rho^{(t)}\widetilde{\Psi}^{(t)} + (1-\rho^{(t)})\mathbf{1}$ before output smoothing. Finally, trust itself is separately smoothed and bounded:

$$\Psi^{(t)} = 0.5\Psi^{(t-1)} + 0.5\widetilde{\Psi}^{(t)}, \qquad \Psi^{(-1)} = \mathbf{1}, \qquad \bar{\Psi}^{(t)} = \mathrm{clip}\Big(\Psi^{(t)}, 0.5, 2.0\Big). \tag{15}$$

Before clipping, Equation 14 gives each teacher row mean one; the clamp is only a numerical guardrail and can slightly alter that exact mean. Hence $\bar{\Psi}_{k,d}$ is a positive relative trust multiplier, not a probability.

Sparse Execution Policy (SEP). Using sparsemax (Martins & Astudillo, 2016), SEP maps the shared logits to the deterministic execution vector

$$\mathbf{w} = \mathrm{sparsemax}(\mathbf{z}/\tau_e) = \arg\min_{\mathbf{q} \in \Delta^{K-1}} \|\mathbf{q} - \mathbf{z}/\tau_e\|_2^2. \tag{16}$$

The KKT solution has $w_k = [z_k/\tau_e - \nu]_+$, with ν selected such that $\sum_k w_k = 1$. Consequently, teachers below the simplex threshold receive exactly zero execution mass, while the remaining support forms a deterministic ensemble. This is sparse *loss support* rather than a hard sampled route: the sampled ACP action is used only for credit, and SEP always executes $\mathbf{w}$. All teacher forecasts are nevertheless evaluated during training to construct $\mathbf{M}$ and $\mathbf{q}$; SEP does not claim teacher-forward savings. ACP and SEP share $\mathbf{z}$, so policy learning can change future execution supports, but the two paths remain distinct in the updates they receive.

Trust-calibrated sparse execution enters the student branch multiplicatively, without constructing a blended teacher forecast. The right branch of Figure 2 denotes this per-batch effective-weight matrix by $\mathbf{W}$:

$$W_{k,d} \equiv w_{k,d}^{\mathrm{eff}} = \mathrm{sg}(w_k)\, c_d\, \bar{\Psi}_{k,d}, \tag{17}$$

and the KD objective is the evidence-weighted sum

$$\mathcal{L}_{\mathrm{KD}} = \sum_{k=1}^{K}\sum_{d \in \mathcal{D}} W_{k,d} M_{k,d}, \qquad \mathcal{L}_{\mathrm{student}} = \mathcal{L}_{\mathrm{GT}} + \lambda_{\mathrm{KD}}\mathcal{L}_{\mathrm{KD}}. \tag{18}$$

Here $\mathbf{W}$ is an ephemeral per-batch KD-weight matrix rather than a learned policy, and $\mathrm{sg}(\cdot)$ denotes stop-gradient. The coefficients need not sum to one: $\mathbf{w}$ selects teachers, $\mathbf{c}$ fixes the relative importance of temporal signals, and $\bar{\Psi}$ adjusts teacher–signal strengths. The student updates θ only through $\mathcal{L}_{\text{student}}$, whereas the agent updates ϕ through $\mathcal{L}_{\text{ACP}}$ with cleared gradients. Appendix S2 derives the score-function identity, reward bounds, and sparse-support properties; these establish the update semantics, not target-error optimality.

4 Experiments

We evaluate ORCHESTRA's forecasting accuracy, component contributions, and deployment efficiency. The complete per-setting scores and supporting diagnostics are deferred to the appendix.

4.1 Experimental setup

Datasets. We evaluate ORCHESTRA on seven multivariate forecasting benchmarks: ETTh1, ETTh2, ETTm1, ETTm2, Weather, ECL, and Solar. The ETT datasets contain seven electricity-transformer variables sampled hourly or every 15 minutes (Zhou et al., 2021). Weather contains 21 meteorological variables, ECL records hourly electricity consumption from 321 clients, and Solar contains 137 solar-energy variables.

Baselines. We first compare ORCHESTRA with the four frozen teacher forecasters: iTransformer, TimeMixer, PatchTST, and ModernTCN. We then compare distillation strategies using the same TSMixer student. NoKD removes teacher supervision. Uniform KD uses equal weights for all teachers. Best-Teacher-KD distils from one preselected teacher for each dataset across its four prediction lengths. This fixed-source control does not select a different teacher after observing the outcome at each prediction length.

Implementation details. We use a fixed look-back window of 720 and evaluate prediction lengths $H \in \{96, 192, 336, 720\}$ with MSE and MAE. Each dataset-level score averages the four prediction lengths, and the final average weights the seven datasets equally. All datasets use chronological train/validation/test splits and are standardized with training-split statistics. The primary experiments used NVIDIA A100 GPUs. The four teachers remain frozen throughout distillation, while the compact TSMixer student is trained with a learning rate of 5×10^{-5}, zero weight decay, and early stopping with patience five. ACP updates begin after one warm-up epoch and occur every 50 mini-batches. The component ablations use this 28-setting protocol. SEP and MSTE are training-only; the PL=96 efficiency study profiles the TSMixer student alone. Appendix S6 provides dataset statistics, split sizes, training and routing configurations, and complete per-setting results.

4.2 Comparison with the frozen teacher pool

Table 1 compares the deployed ORCHESTRA student with its four frozen teachers. Each entry averages four horizons, testing whether the training-time pool yields one student that remains competitive across forecasting regimes. All rows are direct forecasters, without test-time ensembling or routing; matched fixed-source controls appear in Table 2.

Table 1: **Direct-model comparison averaged over four forecasting horizons.** Lower is better. The best and second-best direct-model values in each row are shown in **bold red** and underlined blue, respectively. Table 2 reports the fixed TSMixer distillation controls; complete 28-setting results are in Appendix S3.1.

	ORCHESTRA		iTransformer		TimeMixer		PatchTST		ModernTCN	
Dataset	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	**0.4321**	**0.4501**	0.4678	0.4764	0.4442	0.4531	0.4933	0.4873	0.4686	0.4651
ETTh2	**0.3501**	**0.4002**	0.3982	0.4264	0.3789	0.4126	0.4372	0.4458	0.3574	0.4028
ETTm1	**0.3546**	**0.3869**	0.3724	0.4018	0.3642	0.3886	0.4161	0.4323	0.3893	0.4093
ETTm2	**0.2586**	**0.3226**	0.2760	0.3365	0.2794	0.3321	0.2846	0.3416	0.2976	0.3480
Weather	**0.2239**	**0.2631**	0.2563	0.2898	0.2302	0.2688	0.2388	0.2786	0.2386	0.2774
ECL	**0.1553**	**0.2571**	0.1618	0.2573	0.1652	0.2590	0.1647	0.2660	0.1667	0.2621
Solar	**0.1871**	0.2509	0.2090	0.2699	0.2157	0.2681	0.2046	0.2543	0.1900	**0.2459**
Avg.	**0.2802**	**0.3330**	0.3059	0.3512	0.2968	0.3403	0.3199	0.3580	0.3012	0.3444

ORCHESTRA has the lowest MSE on all seven datasets and the lowest MAE on six, with an equal-dataset average of $0.2802/0.3330$. Relative to the strongest frozen teacher on this average,

Table 2: **Comparison of distillation controls.** All methods use the same TSMixer student. Dataset-average MSE and MAE are reported; lower is better. Within each dataset–metric pair, the best and second-best values are shown in **bold red** and underlined blue, respectively. The final two columns report ORCHESTRA's relative reduction from NoKD.

	ORCHESTRA		NoKD		Uniform KD		Best-Teacher KD		ORCHESTRA vs. NoKD	
Dataset	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	ΔMSE	ΔMAE
ETTh1	**0.4321**	**0.4501**	0.4707	0.4741	0.4602	0.4639	0.4737	0.4758	−8.20%	−5.06%
ETTh2	**0.3501**	**0.4002**	0.3765	0.4196	0.3674	0.4104	0.3525	0.4007	−7.01%	−4.62%
ETTm1	**0.3546**	**0.3869**	0.3699	0.3949	0.3623	0.3927	0.3757	0.3968	−4.14%	−2.03%
ETTm2	**0.2586**	**0.3226**	0.2933	0.3463	0.2769	0.3413	0.2621	0.3253	−11.83%	−6.84%
Weather	**0.2239**	**0.2631**	0.2376	0.2767	0.2315	0.2755	0.2314	0.2734	−5.77%	−4.92%
ECL	**0.1553**	**0.2571**	0.1747	0.2788	0.1625	0.2659	0.1578	0.2579	−11.10%	−7.78%
Solar	**0.1871**	0.2509	0.1979	0.2615	0.1962	0.2602	0.1921	**0.2459**	−5.46%	−4.05%
Avg.	**0.2802**	**0.3330**	0.3029	0.3503	0.2938	0.3443	0.2922	0.3413	−7.49%	−4.94%

TimeMixer, it lowers MSE by 5.6% and MAE by 2.1%. Solar MAE is the sole exception: ModernTCN records .2459 versus .2509 for ORCHESTRA.

The closest frozen teacher also changes by dataset: TimeMixer on ETTh1, ETTm1, and Weather; iTransformer on ETTm2 and ECL; and ModernTCN on ETTh2 and Solar. Equal-dataset aggregation prevents a large-channel benchmark from dominating this comparison. The pattern reflects heterogeneous biases in the pool and motivates the adaptive-allocation controls below.

4.3 ADAPTIVE MULTI-TEACHER DISTILLATION

Table 2 compares teacher-allocation strategies with the same TSMixer student, point/scale/delta KD formulation, and task-level KD strength. NoKD removes teacher supervision, Uniform KD weights all teachers equally, and Best-Teacher-KD uses one fixed teacher per dataset across horizons. Although both fixed controls improve over NoKD, the selected teacher does not guarantee the best student: Uniform KD and Best-Teacher-KD obtain average MSEs of 0.2938 and 0.2922, while ORCHESTRA reaches 0.2802; their corresponding MAEs are 0.3443, 0.3413, and 0.3330. ORCHESTRA improves on both controls across all seven datasets and outperforms Best-Teacher-KD on 24 of 28 horizon-level MSE settings. This supports the observation that teacher accuracy alone is insufficient, because distillation also depends on student–teacher compatibility and the temporal signals being transferred. Table 3 analyzes the resulting credit, execution, and MSTE components.

4.4 ABLATION STUDY OF THE ORCHESTRA DESIGN

All variants in Table 3 retain multi-teacher KD. Under the 28-setting protocol, we ablate ACP, SEP, and MSTE, then remove each temporal KD signal.

Table 3: **Component and signal ablations.** All variants retain multi-teacher KD. ✓/✗ indicate retained/removed components or signals. The best and second-best MSE/MAE values within each block are shown in **bold red** and underlined blue; Δ columns show percentage change from Full.

	Modules			*Average*		*Δ vs. Full*	
Variant	ACP	SEP	MSTE	MSE	MAE	MSE	MAE
Static prior	✗	✓	✓	0.2848	0.3391	+1.64%	+1.83%
ACP + Softmax	✓	✗	✓	0.2827	0.3358	+0.89%	+0.84%
w/o MSTE	✓	✓	✗	0.2838	0.3374	+1.28%	+1.32%
ORCHESTRA (Full)	✓	✓	✓	**0.2802**	**0.3330**	–	–
	KD signals			*Average*		*Δ vs. Full*	
Variant	Point	Scale	Delta	MSE	MAE	MSE	MAE
Full (all KD)	✓	✓	✓	**0.2802**	**0.3330**	–	–
w/o Point KD	✗	✓	✓	0.2830	0.3349	+1.00%	+0.57%
w/o Scale KD	✓	✗	✓	0.2814	0.3361	+0.43%	+0.93%
w/o Delta KD	✓	✓	✗	0.2834	0.3386	+1.14%	+1.68%

Removing point, scale, or delta supervision raises average MSE/MAE by $1.00/0.57\%$, $0.43/0.93\%$, and $1.14/1.68\%$, respectively. Because the remaining coefficients are not renormalized, each removal also reduces total KD mass. Appendix S4 reports per-dataset results, including local improvements under some removals.

Observed routing behavior. In the recorded PL=96 run, SEP retains 3.11 of four teachers on average; exactly three are active in 89.0% of 2,194 batches, whereas ETTh1 retains all four. These diagnostics complement the controlled ablations in Table 3; Appendix S5.2 gives the full routing results.

4.5 Deployment efficiency and accuracy–cost trade-offs

Deployment uses only the distilled TSMixer student; the teachers, ACP, and MSTE are absent at inference. Figure 3 profiles the PL=96 deployed path on one A100-SXM4-40GB after 30 warm-up and 200 timed iterations. Panel (a) compares latency, replay MSE, and parameter count on high-channel ECL, while panel (b) traces p50 latency from ETT to ECL. Replay MSE is reported only for this profiling run; Table 1 and Appendix S3.1 give the canonical four-horizon results.

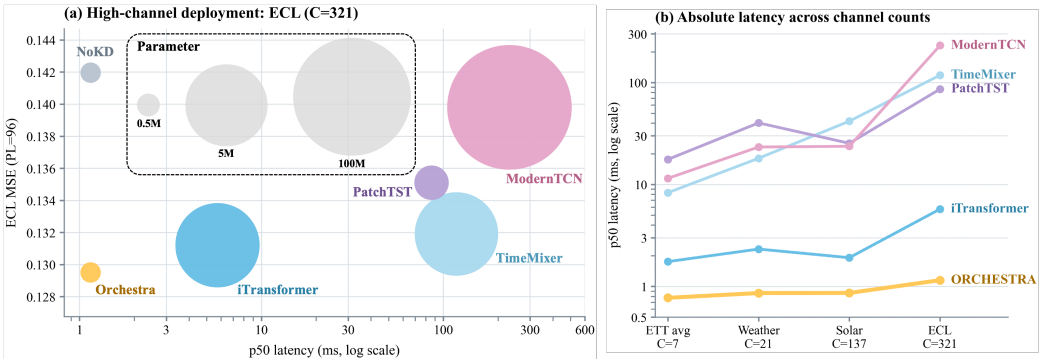


Figure 3: **Deployment efficiency across channel counts.** **(a)** On ECL ($C = 321$), models are located by p50 latency and replay MSE; bubble area denotes parameter count. **(b)** p50 latency across ETT ($C = 7$, averaged over four variants), Weather ($C = 21$), Solar ($C = 137$), and ECL ($C = 321$). ORCHESTRA is the deployed student in both panels.

At ETTh1 ($C = 7$), the student uses 0.257M parameters and 0.0031G MACs, reaches 56,382 samples/s, and records .3730 replay MSE. On ECL ($C = 321$), it uses 0.339M parameters and 0.1405G MACs, compared with 1.373–129.311M parameters and 1.9938–25.6355G MACs for the teachers; it reaches 13,804 samples/s at .1295 replay MSE. Its p50 latency stays below 1 ms through Solar and is 1.151 ms on ECL. These are student-only profiles; Appendix Table 14 reports the complete values.

5 Conclusion

Multi-teacher time-series distillation must determine both which sources are informative in a context and how their temporal evidence should shape the current student. ORCHESTRA addresses this problem through contextual credit, student-relative trust, and sparse execution. ACP learns teacher credit from teacher behaviour and current compatibility; MSTE calibrates point, scale, and first-difference evidence; and SEP transforms the shared credit representation into deterministic sparse KD weights. Across seven datasets and four forecasting horizons, ORCHESTRA achieves the best aggregate MSE and MAE among the evaluated direct forecasters, improves the TSMixer student over NoKD, Uniform KD, and Best-Teacher-KD, and retains student-only inference. Our method provides a new perspective on multi-teacher time-series distillation by treating a heterogeneous teacher pool as structured temporal evidence whose source, content, and strength adapt to the student's learning state.

REFERENCES

David Campos, Miao Zhang, Bin Yang, Tung Kieu, Chenjuan Guo, and Christian S. Jensen. Lightts: Lightweight time series classification with adaptive ensemble distillation. *Proceedings of the ACM on Management of Data*, 1(2):171:1–171:27, 2023. doi: 10.1145/3589316. URL https://doi.org/10.1145/3589316.

Si-An Chen, Chun-Liang Li, Nathanael C. Yoder, Sercan Ö. Arik, and Tomas Pfister. Tsmixer: An all-mlp architecture for time series forecasting. *Transactions on Machine Learning Research*, 2023.

Jang Hyun Cho and Bharath Hariharan. On the efficacy of knowledge distillation. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4793–4801. IEEE, 2019. doi: 10.1109/ICCV.2019.00489. URL https://doi.org/10.1109/ICCV.2019.00489.

Yu-Chen Den, Kuan-Yu Chen, Kendro Vincent, and Tien-Hao Chang. Integrating inductive biases in transformers via distillation for financial time series forecasting. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2*, pp. 803–814, 2026. doi: 10.1145/3770855.3817749. URL https://doi.org/10.1145/3770855.3817749.

Rupasree Dey, Abdul Matin, Nathan Orwick, Yao Zhang, Shrideep Pallickara, and Sangmi Lee Pallickara. When to trust, how to distill: Multi-foundation model guidance for lightweight, robust scientific time series forecasting. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2*, pp. 10807–10818, 2026. doi: 10.1145/3770855.3819018. URL https://doi.org/10.1145/3770855.3819018.

Zixiang Ding, Guoqing Jiang, Shuai Zhang, Lin Guo, and Wei Lin. How to trade off the quantity and capacity of teacher ensemble: Learning categorical distribution to stochastically employ a teacher for distillation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 17915–17923, 2024. doi: 10.1609/aaai.v38i16.29746. URL https://ojs.aaai.org/index.php/AAAI/article/view/29746.

Jianping Gou, Baosheng Yu, Stephen John Maybank, and Dacheng Tao. Knowledge distillation: A survey. *International Journal of Computer Vision*, 129(6):1789–1819, 2021. doi: 10.1007/s11263-021-01453-z. URL https://doi.org/10.1007/s11263-021-01453-z.

Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015.

Junyao Hong, Zesheng Lai, Xinyi Xiao, Suyang Zhou, Aodong Shen, and Youyong Kong. Beyond point predictions: Manifold expansion and dual alignment for robust time series distillation. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=CYh8wTyRkZ.

Yuqi Li, Chuanguang Yang, Hansheng Zeng, Zeyu Dong, Zhulin An, Yongjun Xu, Yingli Tian, and Hao Wu. Frequency-aligned knowledge distillation for lightweight spatiotemporal forecasting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 7262–7272, 2025. doi: 10.1109/ICCV51701.2025.00682. URL https://openaccess.thecvf.com/content/ICCV2025/html/Li_Frequency-Aligned_Knowledge_Distillation_for_Lightweight_Spatiotemporal_Forecasting_ICCV_2025_paper.html.

Yuqi Li, Kuiye Ding, Chuanguang Yang, Szu-Yu Chen, and Yingli Tian. Distilling time series foundation models for efficient forecasting. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 4631–4635, 2026. doi: 10.1109/ICASSP55912.2026.11460474. URL https://arxiv.org/abs/2601.12785.

Rui Liang, Yang Deng, Yaohui Liu, Dafang Zhao, Ozan Baris Mulayim, Ittetsu Taniguchi, and Dan Wang. Thermostill: Distilling time series foundation model into thermal dynamics model for HVAC model predictive control. In *Proceedings of the 17th ACM International Conference on Future and Sustainable Energy Systems*, pp. 636–651, 2026. doi: 10.1145/3744255.3811745. URL https://doi.org/10.1145/3744255.3811745.

Chenxi Liu, Hao Miao, Qianxiong Xu, Shaowen Zhou, Cheng Long, Yan Zhao, Ziyue Li, and Rui Zhao. Efficient multivariate time series forecasting via calibrated language models with privileged knowledge distillation. In *2025 IEEE 41st International Conference on Data Engineering (ICDE)*, pp. 3165–3178, 2025. doi: 10.1109/ICDE65448.2025.00237. URL https://doi.org/10.1109/ICDE65448.2025.00237.

Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. itransformer: Inverted transformers are effective for time series forecasting. In *The Twelfth International Conference on Learning Representations*, 2024.

Yuang Liu, Wei Zhang, and Jun Wang. Adaptive multi-teacher multi-level knowledge distillation. *Neurocomputing*, 415:106–113, 2020. doi: 10.1016/j.neucom.2020.07.048. URL https://doi.org/10.1016/j.neucom.2020.07.048.

Donghao Luo and Xue Wang. Moderntcn: A modern pure convolution structure for general time series analysis. In *The Twelfth International Conference on Learning Representations*, 2024.

Andre F. T. Martins and Ramon F. Astudillo. From softmax to sparsemax: A sparse model of attention and multi-label classification. In *Proceedings of the 33rd International Conference on Machine Learning*, pp. 1614–1623, 2016.

Seyed Iman Mirzadeh, Mehrdad Farajtabar, Ang Li, Nir Levine, Akihiro Matsukawa, and Hassan Ghasemzadeh. Improved knowledge distillation via teacher assistant. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 5191–5198, 2020. doi: 10.1609/aaai.v34i04.5963. URL https://doi.org/10.1609/aaai.v34i04.5963.

Juntong Ni, Zewen Liu, Shiyu Wang, Ming Jin, and Wei Jin. Timedistill: Efficient long-term time series forecasting with MLP via cross-architecture distillation. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.1*, pp. 1113–1124, 2026. doi: 10.1145/3770854.3780218. URL https://doi.org/10.1145/3770854.3780218.

Yuqi Nie, Nam H. Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. In *The Eleventh International Conference on Learning Representations*, 2023.

Xiangfei Qiu, Jilin Hu, Lekui Zhou, Xingjian Wu, Junyang Du, Buang Zhang, Chenjuan Guo, Aoying Zhou, Christian S. Jensen, Zhenli Sheng, and Bin Yang. Tfb: Towards comprehensive and fair benchmarking of time series forecasting methods. *Proceedings of the VLDB Endowment*, 17 (9):2363–2377, 2024. doi: 10.14778/3665844.3665863. URL https://www.vldb.org/pvldb/vol17/p2363-hu.pdf.

Adriana Romero, Nicolas Ballas, Samira Ebrahimi Kahou, Antoine Chassang, Carlo Gatta, and Yoshua Bengio. Fitnets: Hints for thin deep nets. In *International Conference on Learning Representations*, 2015. URL https://mlanthology.org/iclr/2015/romero2015iclr-fitnets/.

Weimin Song, Qianqian Ren, Xingfeng Lv, and Jinbao Li. Meta-contrastive task adaptation for time series forecasting: A structured dual-teacher framework. *Information Fusion*, 132:104238, 2026. doi: 10.1016/j.inffus.2026.104238. URL https://doi.org/10.1016/j.inffus.2026.104238.

Shiyu Wang, Haixu Wu, Xiaoming Shi, Tengge Hu, Huakun Luo, Lintao Ma, James Y. Zhang, and Jun Zhou. Timemixer: Decomposable multiscale mixing for time series forecasting. In *The Twelfth International Conference on Learning Representations*, 2024.

Ronald J. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, 8(3–4):229–256, 1992.

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, 2021.

Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. Timesnet: Temporal 2d-variation modeling for general time series analysis. In *The Eleventh International Conference on Learning Representations*, 2023.

Qing Xu, Keyu Wu, Min Wu, Kezhi Mao, Xiaoli Li, and Zhenghua Chen. Reinforced knowledge distillation for time series regression. *IEEE Transactions on Artificial Intelligence*, 5(6):3184–3194, 2024a. doi: 10.1109/TAI.2023.3341854. URL https://doi.org/10.1109/TAI.2023.3341854.

Qing Xu, Min Wu, Xiaoli Li, Kezhi Mao, and Zhenghua Chen. Reinforced cross-domain knowledge distillation on time series data. In *Advances in Neural Information Processing Systems*, volume 37, pp. 47217–47241, 2024b. URL https://papers.nips.cc/paper_files/paper/2024/hash/5424f6b74f475ad738b54888d609283a-Abstract-Conference.html.

Shan You, Chang Xu, Chao Xu, and Dacheng Tao. Learning from multiple teacher networks. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2017. doi: 10.1145/3097983.3098135. URL https://www.kdd.org/kdd2017/papers/view/learning-from-multiple-teacher-networks.

Fei Yuan, Linjun Shou, Jian Pei, Wutao Lin, Ming Gong, Yan Fu, and Daxin Jiang. Reinforced multi-teacher selection for knowledge distillation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 14284–14291, 2021. doi: 10.1609/aaai.v35i16.17680. URL https://ojs.aaai.org/index.php/AAAI/article/view/17680.

Sergey Zagoruyko and Nikos Komodakis. Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer. In *International Conference on Learning Representations*, 2017. URL https://mlanthology.org/iclr/2017/zagoruyko2017iclr-paying/.

Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series forecasting? In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 11121–11128, 2023.

Hailin Zhang, Defang Chen, and Can Wang. Confidence-aware multi-teacher knowledge distillation. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 4498–4502, 2022. doi: 10.1109/ICASSP43922.2022.9747534. URL https://doi.org/10.1109/ICASSP43922.2022.9747534.

Hailin Zhang, Defang Chen, and Can Wang. Adaptive multi-teacher knowledge distillation with meta-learning. In *IEEE International Conference on Multimedia and Expo*, pp. 1943–1948, 2023. doi: 10.1109/ICME55011.2023.00333. URL https://doi.org/10.1109/ICME55011.2023.00333.

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, 2021.

Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun, and Rong Jin. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *Proceedings of the 39th International Conference on Machine Learning*, pp. 27268–27286, 2022.

Appendix

S1 AI USE STATEMENTS

No generative AI tools were used to formulate the research questions or hypotheses, develop the methodology, design or conduct the experiments, generate or process data, analyze the results, or prepare the figures and tables reported in this paper. The scientific claims and conclusions were developed and verified by the authors, who take full responsibility for the content of the submission. A generative AI assistant was used solely to help draft the wording of this disclosure statement; the authors reviewed and finalized that wording.

S2 FORMAL ANALYSIS OF CREDIT AND EXECUTION

This appendix formalizes the optimization role of ACP and the deterministic execution role of SEP. The analysis treats the current mini-batch evidence as fixed during an agent update, exactly as in the implementation. It establishes score-function semantics and bounded update signals; it does not assert convergence to a target-error-optimal teacher policy or statistical superiority over a full-information optimization rule.

S2.1 UNCLIPPED SCORE-FUNCTION CREDIT ASSIGNMENT

For a fixed state $\mathbf{s}$, define the context-conditioned stochastic credit objective $J(\phi; \mathbf{s}) = \mathbb{E}_{a\sim\pi_\phi}[\alpha u(a)]$. The implementation uses the previously recorded EMA value b_{t-1} as an action-independent baseline; it is fixed before the current action is sampled and can reduce estimator variance without altering its expectation.

Proposition 1 (Credit–execution separation)**.** *For the action-independent, pre-sampling baseline b_{t-1} and fixed current-batch evidence, the* unclipped *score-function term satisfies*

$$\nabla_\phi J(\phi; \mathbf{s}) = \mathbb{E}_{a\sim\pi_\phi}\left[\big(\alpha u(a) - b_{t-1}\big)\nabla_\phi \log \pi_\phi(a \mid \mathbf{s})\right]. \tag{19}$$

Proof. The score-function identity gives

$$\nabla_\phi J(\phi; \mathbf{s}) = \mathbb{E}_a[\alpha u(a)\nabla_\phi \log \pi_\phi(a \mid \mathbf{s})].$$

Because b_{t-1} does not depend on the current action a,

$$\mathbb{E}_a[b_{t-1}\nabla_\phi \log \pi_\phi(a \mid \mathbf{s})] = b_{t-1}\nabla_\phi \sum_a \pi_\phi(a \mid \mathbf{s}) = \mathbf{0},$$

which yields Equation 19. □

S2.2 BOUNDED REWARDS AND ACTION-INDEPENDENT OFFSETS

Lemma 1 (Compatibility range and SEP-entropy neutrality)**.** *When $q_{\text{worst}} - q_{\text{best}} > \epsilon$, the utility in Equation 8 obeys $0 \le u(a) \le 1$. Let $\mathcal{H}_0(\mathbf{w}) = -\sum_{k:w_k>0} w_k \log w_k$ be the ideal entropy of the SEP simplex vector. Then $0 \le \mathcal{H}_0(\mathbf{w}) \le \log K$ and, up to the implementation's 10^{-8} numerical safeguard, the reward in Equation 9 lies in*

$$-\frac{\alpha}{2} \;\le\; R(a, \mathbf{s}) \;\le\; \frac{\alpha}{2} + \beta. \tag{20}$$

Moreover, for $C(\mathbf{s}) = \beta\ \mathrm{sg}[\mathcal{H}(\mathbf{w})/\log K]$,

$$\mathbb{E}_{a\sim\pi_\phi}[C(\mathbf{s})\nabla_\phi \log \pi_\phi(a \mid \mathbf{s})] = \mathbf{0}. \tag{21}$$

Proof. The definition of q_{best} and q_{worst} bounds the numerator in Equation 8 between zero and its denominator. Any simplex vector has entropy between zero and $\log K$, which gives Equation 20. Finally, $C(\mathbf{s})$ is fixed conditional on $\mathbf{s}$, so Equation 21 follows from $C(\mathbf{s})\nabla_\phi \sum_a \pi_\phi(a \mid \mathbf{s}) = \mathbf{0}$. □

For the reported $(\alpha, \beta) = (0.5, 0.3)$, the idealized reward interval is $[-0.25, 0.55]$ before baseline centering and advantage clipping. The entropy offset can affect finite-sample, clipped updates through the recorded scalar reward, but it is neither an action-discriminating credit term nor a direct gradient through SEP.

S2.3 SPARSE EXECUTION GEOMETRY

Lemma 2 (Sparsemax support)**.** *Let $\mathbf{v} = \mathbf{z}/\tau_e$ and $\mathbf{w} = \mathrm{sparsemax}(\mathbf{v})$. There exists a threshold ν such that*

$$w_k = [v_k - \nu]_+, \qquad \sum_{k=1}^{K} w_k = 1. \tag{22}$$

Consequently, $\mathbf{w}$ lies on the probability simplex, at least one teacher has positive execution mass, and a teacher is excluded from the forward KD update exactly when $v_k \le \nu$.

Proof. Equation 16 is the Euclidean projection of $\mathbf{v}$ onto the simplex. The KKT stationarity and complementary-slackness conditions give the thresholded form in Equation 22; the simplex equality constraint selects ν. A simplex vector cannot have all entries equal to zero, and the thresholded form gives the stated support characterization. □

Advantage clipping deliberately sacrifices the exact unbiasedness in Proposition 19 for bounded agent updates. Separately, the student receives gradients only through $\mathcal{L}_{\text{student}}$, while $\mathbf{w}$ is stop-gradient in Equation 17. The formal results therefore establish a credit–execution separation, not a guarantee that the sampled action is globally optimal for target forecasting error.

S3 Complete Experimental Results and Stability

S3.1 Complete Per-Setting Teacher Comparison

Table 4 reports all 28 dataset–horizon settings underlying Table 1. Each model has paired MSE/MAE columns. The table includes settings in which Orchestra is not lowest; the derived Best-T oracle is omitted because it is not an additional forecaster.

Table 4: **Complete per-setting direct-model comparison.** Results use the primary evaluation, and lower is better. Each model has paired MSE/MAE columns. Within each row and metric, the best and second-best direct-forecaster values are shown in **bold red** and underlined blue, respectively.

		Orchestra		iTransformer		TimeMixer		PatchTST		ModernTCN	
Dataset	PL	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	**.3730**	**.4045**	.3922	.4233	.3981	.4232	.4276	.4388	.3889	.4120
	192	**.4249**	**.4391**	.4281	.4476	.4444	.4538	.4706	.4722	.4470	.4524
	336	.4494	.4539	.4613	.4727	**.4408**	**.4501**	.5445	.5199	.4822	.4684
	720	**.4812**	.5030	.5895	.5623	.4936	**.4852**	.5306	.5182	.5563	.5275
ETTh2	96	**.2786**	**.3447**	.3032	.3641	.3025	.3647	.3315	.3824	.2878	.3495
	192	**.3383**	**.3825**	.4099	.4227	.3556	.3925	.3961	.4185	.3416	.3874
	336	**.3686**	.4243	.4403	.4502	.3816	**.4154**	.4756	.4645	.3762	.4186
	720	**.4149**	**.4494**	.4394	.4686	.4760	.4776	.5457	.5178	.4238	.4558
ETTm1	96	**.2919**	**.3512**	.3185	.3667	.3094	.3575	.3141	.3742	.3230	.3680
	192	**.3299**	**.3765**	.3470	.3879	.3415	.3766	.4188	.4269	.3722	.3968
	336	**.3681**	**.3931**	.3875	.4133	.3751	.3945	.4365	.4556	.4056	.4229
	720	**.4283**	.4268	.4366	.4392	.4308	**.4256**	.4951	.4725	.4566	.4497
ETTm2	96	**.1665**	**.2602**	.1802	.2735	.1862	.2684	.1863	.2835	.1848	.2760
	192	**.2211**	**.3002**	.2428	.3158	.2436	.3084	.2412	.3133	.2599	.3243
	336	**.2770**	**.3419**	.2987	.3523	.3021	.3473	.2977	.3468	.3200	.3642
	720	**.3696**	**.3881**	.3822	.4045	.3859	.4044	.4133	.4229	.4258	.4275
Weather	96	.1460	.2037	.1743	.2279	**.1454**	**.1994**	.1568	.2130	.1519	.2077
	192	**.1925**	**.2358**	.2215	.2638	.1998	.2491	.2030	.2539	.1984	.2514
	336	**.2408**	**.2771**	.2804	.3091	.2487	.2859	.2667	.3014	.2566	.2940
	720	**.3162**	**.3360**	.3490	.3582	.3267	.3406	.3286	.3459	.3474	.3563
ECL	96	**.1295**	.2312	.1312	.2274	.1319	**.2272**	.1351	.2387	.1398	.2387
	192	**.1450**	**.2442**	.1519	.2471	.1573	.2489	.1524	.2545	.1526	.2495
	336	**.1592**	**.2611**	.1702	.2675	.1701	.2642	.1681	.2703	.1679	.2643
	720	**.1876**	.2920	.1939	**.2872**	.2015	.2956	.2033	.3006	.2063	.2959
Solar	96	.1685	.2399	.2003	.2475	.1936	.2621	.1743	**.2316**	**.1635**	.2395
	192	.1864	.2489	.2038	.2694	.2080	.2875	**.1827**	**.2482**	.1910	.2485
	336	**.1895**	.2533	.2147	.2818	.2334	.2694	.2309	.2659	.1981	**.2444**
	720	**.2041**	.2614	.2172	.2810	.2277	.2534	.2306	.2715	.2075	**.2515**

S3.2 COMPLETE PER-SETTING DISTILLATION CONTROLS

Table 5 expands the dataset-average control comparison in Table 2 to all 28 settings. It isolates adaptive multi-teacher distillation while holding the TSMixer student fixed: NoKD removes KD, Uniform KD keeps all teachers at equal weight, and Best-Teacher-KD distills from one preselected teacher per dataset. ORCHESTRA is lower-MSE than Uniform KD in all 28 settings and than Best-Teacher-KD in 24 settings; it ties the latter on ECL/336 and is higher on three ETTh2/ETTm2 settings. These exceptions are retained to make the scope of the aggregate claim inspectable.

Table 5: **Complete adaptive-control comparison.** All entries use the same TSMixer student from the primary evaluation, and lower MSE is better. Within each row, the best and second-best controls are shown in **bold red** and underlined blue, respectively; tied best values share the red mark. The table provides horizon-level detail for Table 2.

Dataset	PL	NoKD	Uniform KD	Best-Teacher KD	ORCHESTRA
ETTh1	96	.4019	.3963	.3969	**.3730**
	192	.4418	.4442	.4597	**.4249**
	336	.4933	.4671	.5040	**.4494**
	720	.5457	.5330	.5342	**.4812**
ETTh2	96	.3146	.2959	**.2748**	.2786
	192	.3642	.3634	**.3371**	.3383
	336	.3927	.3880	.3729	**.3686**
	720	.4344	.4222	.4252	**.4149**
ETTm1	96	.3027	.2940	.3006	**.2919**
	192	.3554	.3489	.3710	**.3299**
	336	.3822	.3730	.3712	**.3681**
	720	.4392	.4333	.4599	**.4283**
ETTm2	96	.1783	.1737	.1715	**.1665**
	192	.2422	.2431	.2285	**.2211**
	336	.3370	.2958	.2830	**.2770**
	720	.4158	.3950	**.3653**	.3696
Weather	96	.1569	.1478	.1513	**.1460**
	192	.2016	.1959	.1973	**.1925**
	336	.2600	.2570	.2474	**.2408**
	720	.3319	.3252	.3294	**.3162**
ECL	96	.1420	.1361	.1336	**.1295**
	192	.1594	.1501	.1465	**.1450**
	336	.1759	.1641	**.1592**	**.1592**
	720	.2215	.1995	.1920	**.1876**
Solar	96	.1801	.1766	.1773	**.1685**
	192	.1950	.1954	.1902	**.1864**
	336	.2040	.1987	.1946	**.1895**
	720	.2123	.2140	.2062	**.2041**

S3.3 STABILITY ACROSS THREE RANDOM SEEDS

The main direct-model table uses one common primary evaluation so that ORCHESTRA and the four frozen teachers are compared under matched conditions. Table 6 quantifies primary-student variation across three random seeds, and Table 7 tests whether the corresponding four-control equal-dataset-average MSE ordering persists across those runs. For each random-seed run and dataset, we first average the four prediction lengths and then assign equal weight to the seven dataset scores. These tables do not estimate teacher, mechanism-ablation, or routing-policy variation across random seeds.

Table 6: **Stability of ORCHESTRA across three random seeds and seven datasets.** Each random-seed result averages four prediction lengths. The Avg. row is the mean ± sample standard deviation over the three random-seed runs; lower is better. This table concerns the primary student only.

	MSE			MAE		
Dataset	Run 1	Run 2	Run 3	Run 1	Run 2	Run 3
ETTh1	.4321	.4363	.4313	.4501	.4502	.4469
ETTh2	.3501	.3476	.3531	.4002	.3939	.3978
ETTm1	.3546	.3561	.3532	.3869	.3817	.3836
ETTm2	.2586	.2559	.2597	.3226	.3220	.3187
Weather	.2239	.2258	.2203	.2631	.2644	.2648
ECL	.1553	.1548	.1557	.2571	.2521	.2527
Solar	.1871	.1876	.1871	.2509	.2525	.2488
Avg.	**.2803 ± .0003**			**.3315 ± .0013**		

Note. For each random-seed run, scores are first averaged across horizons and then equally across datasets. Teacher comparisons, mechanism ablations, and routing diagnostics remain primary-evaluation analyses.

Table 7: **Matched distillation controls over three random seeds.** Values use the same horizon-then-dataset aggregation as Table 6; lower MSE is better. The best and second-best control in every run and in the average are shown in **bold red** and underlined blue, respectively. Uniform KD has a verified MSE export only.

Method	Three random-seed runs	Avg. MSE
NoKD	.3029 / .2969 / .2962	.2987 ± .0037
Uniform KD	.2938 / .2963 / .2927	.2943 ± .0018
Best-Teacher-KD	.2922 / .2940 / .2917	.2926 ± .0012
ORCHESTRA	**.2802 / .2806 / .2801**	**.2803 ± .0003**

S4 COMPONENT ANALYSIS

S4.1 SOURCE ALLOCATION AND TEMPORAL SUPERVISION

Table 8 reports the source-allocation controls before equal-dataset aggregation. Full is lower-MSE than Uniform KD on every dataset. Fixed prior, Softmax RL, and MSTE removal provide the intermediate comparisons. These results come from the primary evaluation and do not estimate multi-seed variation or isolate a no-RL control.

Table 8: **Per-dataset source-allocation controls.** Each entry is the four-horizon average MSE from the primary evaluation; lower is better. NoKD has no teacher supervision; Best-Teacher KD retains one fixed source; the remaining KD variants use the four-teacher pool. Within each row, the best and second-best values are shown in **bold red** and underlined blue, respectively. Full combines ACP, SEP, and MSTE.

Dataset	NoKD	Uniform	Best Teacher	Static prior	Softmax RL	w/o MSTE	Full
ETTh1	.4707	.4602	.4737	.4400	.4365	.4382	**.4321**
ETTh2	.3765	.3674	.3525	.3561	.3533	.3547	**.3501**
ETTm1	.3699	.3623	.3757	.3603	.3574	.3589	**.3546**
ETTm2	.2933	.2769	.2621	.2629	.2610	.2620	**.2586**
Weather	.2376	.2315	.2314	.2269	.2255	.2261	**.2239**
ECL	.1747	.1625	.1578	.1575	.1566	.1570	**.1553**
Solar	.1979	.1962	.1921	.1901	.1888	.1895	**.1871**
Avg.	.3029	.2938	.2922	.2848	.2827	.2838	**.2802**

Table 9 shows the aggregate effect of retaining point, scale, and delta supervision, but it does not display MSTE's teacher–signal trust. Figure 4 reports the final-epoch relative trust coefficient $\Psi_{k,d}$

Table 9: **Per-dataset KD-channel ablation.** Each entry is the four-horizon average MSE from the primary evaluation; lower is better. Removing a channel sets its original coefficient to zero without renormalizing the remaining KD signals. Within each row, the best and second-best values are shown in **bold red** and underlined blue, respectively; tied best values share the red mark.

Dataset	Full	w/o point	w/o scale	w/o delta
ETTh1	**.4321**	.4433	.4406	.4354
ETTh2	**.3501**	.3551	.3511	.3523
ETTm1	**.3546**	.3610	.3607	.3599
ETTm2	.2586	**.2583**	.2625	**.2583**
Weather	.2239	**.2212**	.2221	.2213
ECL	.1553	.1560	**.1548**	.1550
Solar	**.1871**	.1888	.1891	.1876
Avg.	**.2802**	.2830	.2814	.2834

for every teacher–signal pair at PL=96. The values are normalized by the contemporaneous teacher-pool reference, allowing comparison across signals. Ψ is not a probability, a per-window route, or an absolute teacher-quality score. Figure 4 is descriptive; the controlled w/o MSTE result provides the causal comparison.



Figure 4: **MSTE relative-trust arrangements (PL=96; primary evaluation, final epoch).** Each panel reports the $\Psi_{k,d}$ coefficient for four frozen teachers across point, scale, and first-difference KD signals. The common colour scale is centred at $\Psi = 1$, the teacher-pool reference: cool cells are relatively downweighted and warm cells relatively upweighted for that signal. Ψ is a positive relative KD-trust coefficient, not a probability or an absolute teacher-performance score. These seven final-epoch matrices are descriptive diagnostics; the causal evidence for MSTE remains the controlled ablation in Tables 3 and 9.

S4.2 Backbone probes

TSMixer is the primary student. Table 10 records DLinear and MLP only as architecture probes under the same framework; it does not claim matched NoKD-to-KD gains for these alternative students. The equal-dataset results are consistent with using TSMixer as the reported deployment

student, while the other probes demonstrate that the implementation is not tied to a single student class.

Table 10: **Student-backbone probes.** Each value is a four-horizon dataset average from the primary evaluation; lower is better. Within each row and metric, the best and second-best values are shown in **bold red** and underlined blue, respectively. The primary TSMixer has 0.257–0.339M parameters depending on dataset channel count; DLinear and MLP are framework instantiations rather than matched NoKD-to-KD comparisons.

	TSMixer		DLinear (138K)		MLP (52K)	
Dataset	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	.4321	.4501	**.4147**	**.4354**	.4582	.4685
ETTh2	.3501	.4002	**.3460**	**.3949**	.3813	.4220
ETTm1	**.3546**	.3869	.3605	**.3827**	.3781	.4021
ETTm2	.2586	.3226	**.2512**	**.3192**	.2906	.3542
Weather	.2239	**.2631**	.2348	.2782	**.2223**	.2746
ECL	**.1553**	**.2571**	.1631	.2590	.1805	.2835
Solar	**.1871**	**.2509**	.2178	.2745	.2120	.2695
Avg.	**.2802**	**.3330**	.2840	.3348	.3033	.3535

S5 Routing and Sensitivity Diagnostics

S5.1 KD-strength and temperature sensitivity

Figure 5 shows all 112 measured entries in the $7 \times 4 \times 4$ λ_{KD}–temperature grid. Each panel is a PL=96, ten-epoch primary-evaluation run. Responses vary across datasets, consistent with the per-task KD strength used in the primary protocol; this diagnostic therefore does not establish a universal setting or multi-run robustness.

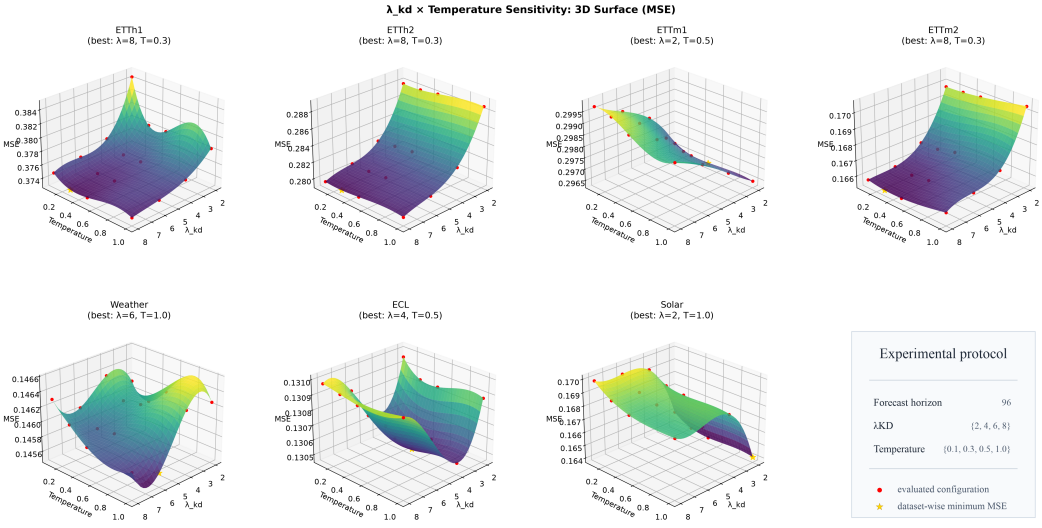


Figure 5: **Measured λ_{KD}–temperature response surfaces.** Red dots are observed grid points, gold stars mark the lowest measured value per dataset, and the cubic surface is used only for visualization.

S5.2 Credit and sparse-execution diagnostics

Figure 6 distinguishes the two routing outputs on the PL=96 per-batch test replay. Panel (a) records the ACP action sampled on each batch, whereas panel (b) aggregates the deterministic SEP coefficient that enters the evidence-weighted forward KD objective on the same replay. Their distributions differ because they serve distinct operations. Table 11 confirms sparse forward allocations: mean

support is 3.11 of four teachers across 2,194 replay batches, with exact three-teacher support on 89.0% of batches. ETTh1 retains dense four-teacher execution on its PL=96 replay. These diagnostics describe the specified credit–execution factorization and do not estimate per-window routes or multi-run routing stability.

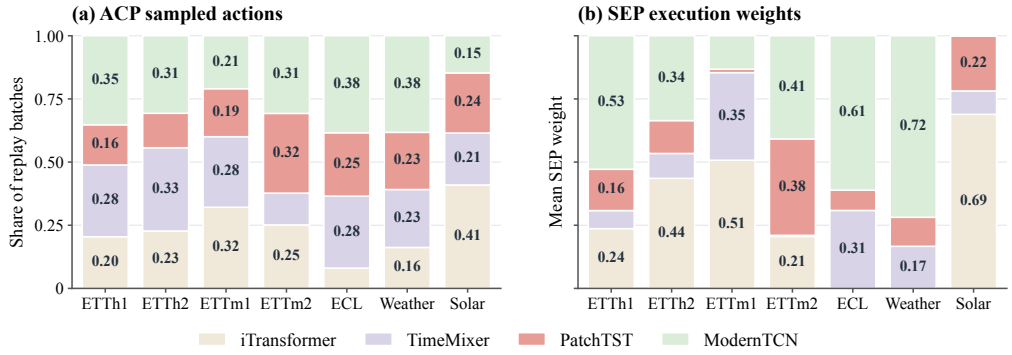


Figure 6: **Credit–execution diagnostic. (a)** Batch shares of sampled ACP actions. **(b)** Mean deterministic SEP execution weights on the same PL=96 primary test replay. Each stacked bar sums to one; the panels show distinct routing roles rather than window-level quality or multi-run stability.

Table 11: **Observed sparse-execution support (PL=96 test replay; primary evaluation).** $\|\mathbf{w}\|_0$ counts the teachers with nonzero SEP coefficient on a replay batch. Percentages are within-dataset batch shares, not window-level quantities.

Dataset	Batches	$\mathbb{E}[\|\mathbf{w}\|_0]$	2 teachers	3 teachers	4 teachers
ETTh1	88	4.00	0.0%	0.0%	100.0%
ETTh2	88	3.98	0.0%	2.3%	97.7%
ETTm1	358	3.09	0.0%	90.5%	9.5%
ETTm2	358	3.06	0.0%	94.1%	5.9%
Weather	327	3.01	0.0%	99.1%	0.9%
ECL	323	3.00	0.0%	100.0%	0.0%
Solar	652	3.01	0.2%	98.5%	1.4%
All	2,194	3.11	0.05%	89.0%	11.0%

S6 Reproducibility and Deployment Details

S6.1 Data and Training Configuration

Table 12: **Dataset protocol.** All datasets use input length 720 and prediction lengths $\{96, 192, 336, 720\}$. ETT uses chronological 12/4/4-month splits; Weather, ECL, and Solar use chronological 0.7/0.1/0.2 splits.

Dataset	Vars.	Freq.	Train	Val.	Test
ETTh1	7	1h	8640	2880	2880
ETTh2	7	1h	8640	2880	2880
ETTm1	7	15min	34560	11520	11520
ETTm2	7	15min	34560	11520	11520
Weather	21	10min	36887	5270	10539
ECL	321	1h	18412	2632	5260
Solar	137	10min	36792	5256	10512

Table 13: **Training and routing configuration.** Global settings and the documented task-varying settings of the reported experiments. The KD coefficient varies by task; the preserved result archive does not reconstruct a 28-row selection manifest, so it is not represented as a single global hyperparameter.

Item	Setting
Student / frozen teachers	TSMixer / iTransformer, TimeMixer, PatchTST, ModernTCN
Input / horizons	720 / $\{96, 192, 336, 720\}$
Agent state / MLP	12 dimensions / $12 \to 64 \to 64 \to 4$
ACP prior / dynamic features	complete validation teacher MSE / within-batch min–max disagreement and jump
ACP / SEP	Categorical(softmax($\mathbf{z}/\tau_e$)) / sparsemax($\mathbf{z}/\tau_e$)
Routing temperature / channel weights	$\tau_0 = 1.0 \to \tau_{\min} = 0.3$, $\gamma = .5$ / $(.5, .25, .25)$ for point/scale/delta
MSTE evidence / trust	$\bar{M}$: .8/.2 EMA; Ψ: .5/.5 EMA; $T_r = .5$; $\gamma = .01$; clamp $[.5, 2]$
ACP reward / policy regularization	$.5(u(a) - .5) + .3\ \mathrm{sg}[\mathcal{H}(\mathbf{w})/\log 4]$ / ACP-policy entropy $\eta = 10^{-5}$
ACP update / baseline / advantage	after one warm-up epoch, every 50 mini-batches / momentum .99 / $\mathrm{clip}(R - b, [-.1, .1])$
Learning rate / regularization / schedule	$5{\times}10^{-5}$ / zero weight decay / type3
Early stopping / normalization	patience 5 / training-split standard scaler

S6.2 Training procedure

Algorithm 1 ORCHESTRA training.

Require: Training mini-batches; complete validation loader; frozen teachers $\{g_k\}_{k=1}^{K}$; student f_θ; credit agent h_ϕ.

1: Before training, compute element-weighted teacher-to-ground-truth MSE over the complete validation split; mean-normalize it to obtain the fixed ACP prior $\mathbf{p}$.
2: For each mini-batch, compute $\hat{\mathbf{y}}^s = f_\theta(\mathbf{x})$ and $\{\hat{\mathbf{y}}^k = g_k(\mathbf{x})\}_{k=1}^{K}$; construct the point/scale/delta evidence matrix $\mathbf{M}$.
3: Update MSTE from $\mathbf{M}$; form $\mathbf{s} = [\mathbf{p}; \mathbf{d}; \mathbf{j}]$ and logits $\mathbf{z} = h_\phi(\mathbf{s})$.
4: Sample the ACP credit action $a \sim \mathrm{Categorical}(\mathrm{softmax}(\mathbf{z}/\tau_e))$; obtain deterministic SEP weights $\mathbf{w} = \mathrm{sparsemax}(\mathbf{z}/\tau_e)$.
5: Construct $w_{k,d}^{\mathrm{eff}}$ with Equation 17; update θ with $\mathcal{L}_{\mathrm{GT}} + \lambda_{\mathrm{KD}}\mathcal{L}_{\mathrm{KD}}$.
6: After the first epoch, every U mini-batches compute the detached current-batch reward in Equations 8–9; update the EMA baseline and form the clipped advantage.
7: At these scheduled ACP updates, clear the agent gradient accumulator and update ϕ with Equation 11.
8: **return** updated (θ, ϕ).

S6.3 Deployment profile

Measurement scope. Parameter counts and MACs vary across datasets because channel counts and the resulting model configurations differ. Accordingly, the comparisons in Table 14 are intended within each dataset, rather than as a common cross-dataset hardware ranking. For ORCHESTRA, the timing path contains only the deployed TSMixer student: frozen-teacher forwards, ACP, and MSTE are training-time components. DLinear and MLP are included as lightweight backbone probes, not as fully matched KD baselines. Replay MSE helps locate a profiled point, while the four-horizon results in Section 4 remain the canonical forecasting evaluation.

Table 14: **Complete deployment profile at PL=96.** Params (millions), single-sample MACs (billions), throughput (samples/s), and replay MSE are measured on one A100-SXM4-40GB with PyTorch 2.7.0/CUDA 12.6, using 30 warm-up and 200 timed iterations. Throughput uses batch 32 for ETT/Weather and 16 for ECL/Solar. ORCHESTRA denotes the deployed student; DLinear and MLP are backbone probes. Replay MSE is a profiling coordinate rather than the canonical four-horizon result.

Dataset	Model	Params (M)	MACs (G)	Thpt. (samples/s)	Replay MSE
ETTh1	ORCHESTRA	0.257	0.0031	56,382	.3730
	DLinear	0.138	0.0010	127,348	.3689
	MLP	0.052	0.0004	99,452	.4000
	iTransformer	1.001	0.0111	24,399	.3922
	TimeMixer	3.202	0.3779	3,340	.3981
	PatchTST	7.585	2.0760	3,227	.4276
	ModernTCN	1.250	0.0244	25,333	.3889
ETTh2	ORCHESTRA	0.257	0.0031	36,435	.2786
	DLinear	0.138	0.0010	95,038	.2733
	MLP	0.052	0.0004	68,728	.3025
	iTransformer	0.304	0.0034	19,583	.3032
	TimeMixer	3.202	0.3779	4,196	.3025
	PatchTST	13.890	6.1557	1,393	.3315
	ModernTCN	1.250	0.0244	24,747	.2878
ETTm1	ORCHESTRA	0.257	0.0031	34,397	.2919
	DLinear	0.138	0.0010	90,784	.3147
	MLP	0.052	0.0004	72,674	.3194
	iTransformer	0.304	0.0034	17,357	.3185
	TimeMixer	3.202	0.3779	4,403	.3094
	PatchTST	13.890	6.1557	1,347	.3141
	ModernTCN	2.789	0.2968	1,416	.3230
ETTm2	ORCHESTRA	0.257	0.0031	36,117	.1665
	DLinear	0.138	0.0010	94,811	.1634
	MLP	0.052	0.0004	69,637	.1742
	iTransformer	0.304	0.0034	18,282	.1802
	TimeMixer	3.204	0.7558	3,225	.1862
	PatchTST	13.890	6.1557	1,508	.1863
	ModernTCN	2.789	0.2968	1,457	.1848
Weather	ORCHESTRA	0.261	0.0092	36,927	.1460
	DLinear	0.138	0.0029	95,365	.1659
	MLP	0.052	0.0011	70,063	.1512
	iTransformer	5.153	0.1303	13,688	.1743
	TimeMixer	4.738	1.6771	1,750	.1455
	PatchTST	10.730	12.3308	786	.1568
	ModernTCN	3.090	0.3677	1,346	.1519
ECL	ORCHESTRA	0.339	0.1405	13,804	.1295
	DLinear	0.138	0.0444	29,676	.1348
	MLP	0.054	0.0168	35,234	.1491
	iTransformer	5.153	1.9938	2,780	.1312
	TimeMixer	4.740	25.6355	134	.1319
	PatchTST	1.373	9.3188	184	.1351
	ModernTCN	129.311	23.3709	68	.1398
Solar	ORCHESTRA	0.291	0.0600	18,456	.1685
	DLinear	0.138	0.0189	48,471	.1910
	MLP	0.053	0.0072	34,713	.1906
	iTransformer	3.575	0.5432	8,332	.2003
	TimeMixer	18.193	18.5510	342	.1936
	PatchTST	1.307	3.1691	622	.1743
	ModernTCN	5.243	0.8870	669	.1635