CORRECT THEN FORECAST: OBSERVER STATE-SPACE MODELS FOR TIME SERIES FORECASTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Time series forecasting requires extrapolating the dynamics of an observed process beyond the last available measurement. Yet recurrent forecasting models typically treat observations as inputs that directly control their latent dynamics. It leads to a regime change when these observations become unavailable at prediction time. Following a state-estimation perspective, we introduce **Observer State-Space Models** (OSSMs), a class of recurrent models that **interprets the observed input time series as measurements of an underlying autonomous dynamical system**. OSSMs explicitly separate latent-state propagation from measurement assimilation: a single transition governs the dynamics across both context and forecasting intervals, while available observations **correct** the estimated state through an **observer**. This formulation naturally exposes classical control-theoretic properties, including observability and convergence of the state estimation error. We further show that conventional and recent SSMs can be recovered as particular instances of our OSSM framework, thereby providing a unified interpretation of their recurrent dynamics and revealing modeling inconsistencies. We perform experiments across several benchmarks showing that OSSM achieves substantial improvements while maintaining the same parameter count and training setup as the corresponding SSM baseline. These results support a simple principle for recurrent forecasting: observations should correct the estimated latent state, rather than control the dynamics used to propagate it.

1 INTRODUCTION

In recent years, data-driven methods, and deep learning in particular, have achieved remarkable success across a wide range of learning tasks and data modalities. Among them, sequence modeling occupies a distinctive position at the intersection of time series analysis, natural language processing, dynamical systems, and machine learning (Elman, 1990; Zeng et al., 2023; Tan et al., 2024; Brachet et al., 2026). With the rise of deep learning, language became a major driving force behind advances in sequence modeling. Recurrent neural networks (RNNs) (Elman, 1990; Jordan, 1997), Long-short term memory (LSTM) (Hochreiter & Schmidhuber, 1997), gated architectures (Chung et al., 2014), attention mechanisms (Bahdanau et al., 2016), Transformers (Vaswani et al., 2017), and later State-Space Models (SSMs) (Gu & Dao, 2024) were extensively developed and refined in the context of natural language processing. These architectural advances were subsequently transferred to other sequential domains, including time series modeling (Laptev et al., 2017; Hewamalage et al., 2021; Zhang et al., 2023; Hu et al., 2024b; Jin et al., 2024).

RNNs were explicitly conceived as dynamical systems, in which an internal state evolves over time as sequential inputs are processed (Elman, 1990). This formulation is closely related to nonlinear state-space representations, which have long been central to control theory and time series analysis (Kalman, 1960). Consider a linear time-invariant discrete-time state-space model,

$$\boldsymbol{x}_{t+1} = \boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{B}\boldsymbol{u}_t,\ \boldsymbol{y}_{t+1} = \boldsymbol{C}\boldsymbol{x}_{t+1}, \tag{1}$$

where $t \in \mathbb{R}^+$ is a timestamp, $\boldsymbol{x}_t \in \mathbb{R}^{d_x}$ denotes the **internal state** of the system at time t, $\boldsymbol{u}_t \in \mathbb{R}^{d_u}$ a **control input** acting on its dynamics, and $\boldsymbol{y}_t \in \mathbb{R}^{d_y}$ its **measured output**. $\boldsymbol{A} \in \mathbb{R}^{d_x \times d_x}$, $\boldsymbol{B} \in \mathbb{R}^{d_x \times d_u}$, and $\boldsymbol{C} \in \mathbb{R}^{d_y \times d_x}$ are, respectively, the state, input, and output matrices. We call the system defined by Equation 1 a **controlled system**. RNNs extend this formulation by replacing

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.

Table 1: Overview of how different recurrent formulations satisfy the desired consistency properties in the case of time series forecasting. ✓ (or ×) indicates whether the property is satisfied (or not). We propose OSSM, which satisfies all these properties.

Recurrent mode	**Modeling consistency**	**Time propagation consistency**	**Regime consistency**
Shallow	×	✓	×
Deep	✓	✓	×
Encoder only	✓	×	×
SpaceTime (Zhang et al., 2023)	✓	✓	×
OSSM (ours)	✓	✓	✓

the linear state transition with nonlinear and learnable dynamics. More recently, structured SSMs have revisited this formalism to build sequence models that remain recurrent while allowing efficient computation over long sequences (Gu et al., 2022; Gu & Dao, 2024).

This input-state-output interpretation appeared naturally in early recurrent approaches to machine translation, where source words sequentially drove the recurrent state and target words were produced at the output (Castaño et al., 1997; Castano & Casacuberta, 1997; Sundermeyer et al., 2014). From a dynamical-system perspective, the source sequence thus acts as an exogenous input $\boldsymbol{u}$, while the translated sequence constitutes the output $\boldsymbol{y}$. Later encoder-decoder architectures made this separation explicit by using distinct recurrent networks for source encoding and target generation (Cho et al., 2014; Sutskever et al., 2014), a natural choice since source and target follow different sequential distributions and need not share the same dynamics.

While this separation is natural in machine translation, Time Series Forecasting (TSF) is fundamentally different: the observed context and the predicted future are two portions of the same process. When recurrent models (RNNs, LSTMs, and SSMs) are directly transferred to TSF, however, the input-output interpretation becomes less natural. Unlike machine translation, the input and output are the same processes, shifted in time by one timestamp. The underlying system is no longer controlled, but an **autonomous system**. However, $\boldsymbol{u}$ still plays the role of a control input. Assigning this role to $\boldsymbol{u}$ induces a **modeling mismatch**. Deep autoregressive models (Salinas et al., 2020) partially solved this issue by mapping inputs in a different representation space than the predicted outputs. However, when observations are no longer available, such models inject the predictions back into the model input: it becomes an autonomous system, whereas it is a controlled one when observations are available, leading to a **discrepancy at the context-forecasting boundary** that is induced solely by how the dataset is split during learning. A recent SSM-based work, SpaceTime (Zhang et al., 2023), explicitly assumes that $\boldsymbol{u}$ and $\boldsymbol{y}$ are two different processes and predicts future observations to make predictions. However, based on how SpaceTime predicts $\boldsymbol{u}$, the discontinuity remains, and the fact that $\boldsymbol{u}$ and $\boldsymbol{y}$ are different cannot hold. Recent works thus only use such recurrent networks as encoder (Hu et al., 2024a; Xu et al., 2025; Weng et al., 2025), but lose at prediction time the ability of such a unit to **learn how to move forward in time**.

These different designs highlight three desirable consistency properties for recurrent TSF models:

- **Modeling consistency**: the roles assigned to $\boldsymbol{u}$ and $\boldsymbol{y}$ should be consistent with their interpretation in the forecasting problem;
- **Time propagation consistency**: the learned dynamics should be exploited to propagate the latent state throughout the whole time window;
- **Regime consistency**: the dynamics attributed to the underlying process should remain consistent across the context and forecasting intervals.

In this context, we propose a new class of recurrent models called **Observer State-Space Models** (OSSMs). We take a different perspective: we consider the inputs $\boldsymbol{u}$ to be measurements of a dynamical system's outputs $\boldsymbol{y}$. This perspective naturally leads us back to state observers from control theory (Luenberger, 1964; 1966). During the context, available measurements **correct** the estimated latent trajectory through an observer; once forecasting begins, this correction disappears, while the underlying state transition remains unchanged. Thus, the computation necessarily changes

when measurements become unavailable, but the modeled process dynamics are the same in both the context and the horizon interval. This formulation further allows us to leverage tools from control theory to characterize and control the convergence of the estimated latent state. We summarize our contribution in Table 1. We show that simple SSMs and SpaceTime (Zhang et al., 2023) can be recovered as a particular instance of the OSSM framework, but present model mismatches. This work addresses two research questions:

RQ1. Is it possible to define a recurrent model satisfying the three consistency properties?

RQ2. If so, does such a model allow for better performance on the time series forecasting task?

To address RQ1, we revisit Luenberger observers and standard SSMs, introduce the OSSM formulation, and characterize the relationship between OSSM, conventional SSMs, and SpaceTime. To address RQ2, we conduct controlled comparisons between OSSM and corresponding SSM variants on established short- and long-term forecasting benchmarks, using matched parameter counts and training protocols. Our results show that explicitly separating state propagation from measurement correction can yield substantial forecasting gains without increasing model capacity.

2 Background and related work

Observation-driven and free-running dynamics. The discrepancy between recurrent dynamics driven by observed values and those obtained in free-running mode due to Teacher Forcing (Williams & Zipser, 1989) has been studied across several sequential modeling domains (Bengio et al., 2015; Lamb et al., 2016). Closely related difficulties arise when reconstructing dynamical systems, particularly in chaotic regimes, where models trained using observed trajectories may rapidly diverge when recursively rolled out without observations (Pathak et al., 2018; Mikhaeil et al., 2022; Herz et al., 2026). In TSF, methods avoid or mitigate this transition by adapting learning strategies (Teutsch & Mäder, 2022), isolating recurrent dynamics from the forecasting stage (Hu et al., 2024a), refining predictions (Wu et al., 2025), or adapting the dynamics once observations are no longer available (Zhang et al., 2023). Our approach differs in that it revisits the role assigned to observations in the state-space formulation itself: we use observations only as measurements that correct the estimated state when available.

Time series as observations of dynamical systems. Our work belongs to a growing family of forecasting approaches that views time series as observations of an underlying dynamical process. A first line of work builds on Koopman theory (Koopman, 1931), which represents the evolution of observables of a nonlinear dynamical system through a linear operator in a generally infinite-dimensional function space (Liu et al., 2023; Gupta et al., 2024; Wu et al., 2025). Another line of work relies on reconstruction results from nonlinear and chaotic dynamical systems (Hu et al., 2024a; Majeedi et al., 2025). Other methods introduce dynamical inductive biases through state-space representations (Zhang et al., 2023) or explicitly couple temporal evolution with generative mechanisms such as diffusion models (Cachay et al., 2023; Guo et al., 2025).

Learning-based observers. Our work is also related to learning-based state observers. KalmanNet and related neural filtering methods learn or adapt measurement-update mechanisms in state-space models with known or partially known dynamics (Revach et al., 2022), typically in state-estimation settings where a physical model and, during training, latent-state supervision may be available. Neural Luenberger and Kazantzis–Kravaris–Luenberger (KKL) observers (Kazantzis & Kravaris, 1998; Niazi et al., 2022) instead focus on reconstructing the state of nonlinear dynamical systems from measurements. These approaches share the principle of correcting a state estimate with observations, but target state estimation rather than output-only time series forecasting.

Together, these works support a view of time series forecasting in which predicting future observations amounts, at least approximately, to identifying and propagating underlying dynamics from (partial) observations. Our work adopts this perspective and focuses on a complementary question: how should the available observations interact with the estimated dynamical state for forecasting purposes?

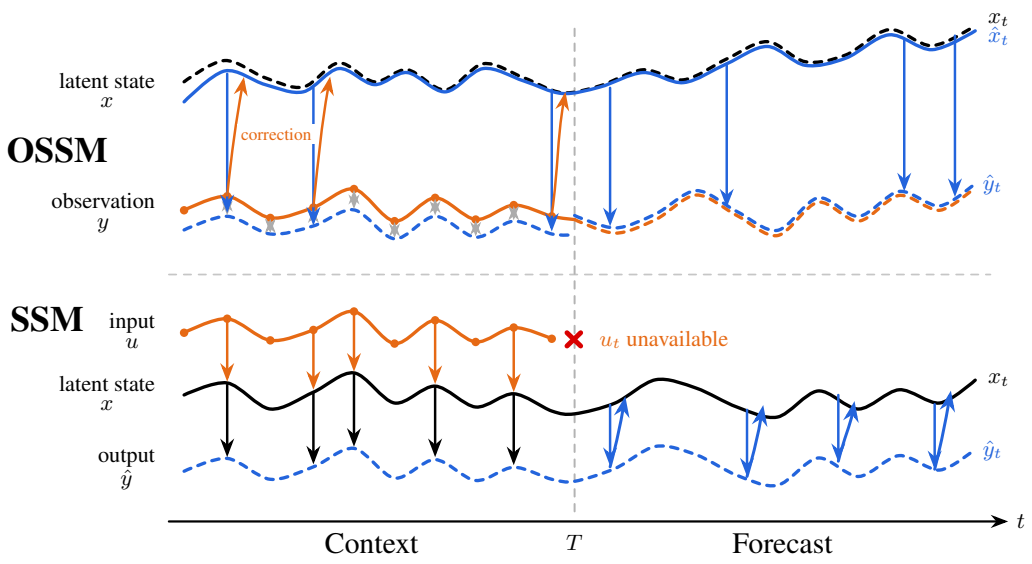


Figure 1: Schematic information flow in OSSM and SSM forecasting. In OSSM, observations correct an autonomous latent-state estimate during context; after T, the correction is removed while the latent dynamics remain unchanged. In SSM, observations act as inputs during context; after T, the input is unavailable and predicted outputs are fed back into the dynamics.

3 Observer State-Space Models

In this section, we introduce **Observer State-Space Models (OSSMs)**, a class of models that aims to provide a recurrent deep learning model that validates the three consistency properties for the time series forecasting task.

3.1 A consistent recurrent model

Let $\boldsymbol{y}$ denote a process from which we have $T \in \mathbb{N}$ historical records $\boldsymbol{y}_{1:T}$. In the time series forecasting task, based on these historical records, we aim to predict the $H \in \mathbb{N}$ future ones. To perform the forecasts, we formulate hypotheses to define our deep learning model.

Hypothesis 3.1. The process $\boldsymbol{y}$ is considered to be derived from an underlying dynamical system.

Following works based on Koopman theory (Koopman, 1931), we formulate the following hypothesis:

Hypothesis 3.2. There exists a latent space where the input process is derived from an underlying linear dynamical system.

Let $\mathcal{E} : \mathbb{R}^{d_y} \to \mathbb{R}^{d_h}$ be a learnable non linear encoder, with $d_y \in \mathbb{N}$ and $d_h \in \mathbb{N}$ the input and hidden dimensions. We thus suppose that $\mathcal{E}_\theta(\boldsymbol{y})$ is derived from a linear latent dynamical system $\boldsymbol{x}$ as it follows:

$$\boldsymbol{x}_{t+1} = \boldsymbol{A}_O \boldsymbol{x}_t, \quad \mathcal{E}(\boldsymbol{y}_t) = \boldsymbol{C}_O \boldsymbol{x}_t \tag{2}$$

with $\boldsymbol{A}_O \in \mathbb{R}^{d_x \times d_x}$ the latent dynamical matrix, $d_x \in \mathbb{N}$ the latent dynamical state dimension, and $\boldsymbol{C}_O \in \mathbb{R}^{d_h \times d_x}$ the latent measurement matrix.

During the learning procedure, the encoder and the latent matrices weights $(\theta_\mathcal{E}, \theta_{A_O}, \theta_{C_O})$ are learned. However, learning the latent dynamics is not sufficient. Indeed, the initial conditions of the latent system's state are unknown. Following the Luenberger observer theory (Luenberger, 1964; 1966), we use the available observations to correct the latent state estimation in order to recover the true latent trajectory:

$$\hat{\boldsymbol{x}}_{t+1} = \boldsymbol{A}_O \hat{\boldsymbol{x}}_t + m_t \boldsymbol{L}_O (\mathcal{E}(\boldsymbol{y}_t) - \mathcal{E}(\hat{\boldsymbol{y}}_t)), \quad \mathcal{E}(\hat{\boldsymbol{y}}_t) = \boldsymbol{C}_O \hat{\boldsymbol{x}}_t \tag{3}$$

with $\boldsymbol{L}_O \in \mathbb{R}^{d_x \times d_h}$ the latent observer matrix, and $m_t \in \{0, 1\}$, a measurement coefficient equal to 1 when a measurement is available, here $\forall t \leq T$, and 0 otherwise. Equation 3 defines what we call the Observer State-Space Model. As $\mathcal{E}(\hat{\boldsymbol{y}})$ lives in the latent space, one should define a nonlinear decoder $\mathcal{D} : \mathbb{R}^{d_h} \to \mathbb{R}^{d_y}$ so the predictions made by the OSSM are projected back in the original space observation.

However, in classical recurrent formulations, the prediction head both propagates the latent state and performs the predictions in the original space. Equation 3 only accounts for the propagation. An OSSM for both is defined as follows:

$$\hat{\boldsymbol{x}}_{t+1} = \boldsymbol{A}_O \hat{\boldsymbol{x}}_t + m_t \boldsymbol{L}_O(\mathcal{E}(\boldsymbol{y}_t) - \hat{\boldsymbol{y}}_t), \quad \hat{\boldsymbol{y}}_t = \boldsymbol{C}_O \hat{\boldsymbol{x}}_t \tag{4}$$

Equation 4 adopts the prediction-head parameterization used in our forecasting architecture. The readout $\hat{\boldsymbol{y}}_t = \boldsymbol{C}_O \hat{\boldsymbol{x}}_t$ is used both as the prediction produced by the head and as the reference term in the innovation. Consequently, we set $d_y = d_h$, so that $\mathcal{E}(\boldsymbol{y}_t) - \hat{\boldsymbol{y}}_t$ is well-defined. In this parameterization, $\mathcal{E}$ should be understood as a learned preprocessing of the observed signal within the shared prediction-head coordinate space, rather than as a separate decoder–encoder pair. The learnable parameters are $(\theta_{\mathcal{E}}, \theta_{\boldsymbol{A}_O}, \theta_{\boldsymbol{C}_O}, \theta_{\boldsymbol{L}_O})$.

Consistency properties

Proposition 3.1 (Modeling consistency)**.** *OSSM satisfies the modeling consistency property.*

In the OSSM formalism, the correction term $(\mathcal{E}(\boldsymbol{y}_t) - \hat{\boldsymbol{y}}_t)$ compares an observed measurement representation with its estimate in the same space. The observed and predicted processes are modeled as deriving from the same latent dynamical system. Their roles are therefore consistent with their theoretical interpretation. OSSM is thus consistent in terms of modeling.

Proposition 3.2 (Time propagation consistency)**.** *OSSM satisfies the time propagation consistency property.*

OSSM moves forward in time according to the latent dynamical matrix $\boldsymbol{A}_O$ on both the context and forecast intervals. The prediction head, therefore, applies the same time-propagation mechanism to historical and predicted data. OSSM is thus consistent in terms of time propagation.

Proposition 3.3 (Regime consistency)**.** *OSSM satisfies the regime consistency property.*

OSSM uses the same latent dynamical matrix $\boldsymbol{A}_O$ on both the context and forecast intervals, whether observations are available or not. The transition to the forecast regime only removes the measurement-correction term; it does not introduce different latent dynamics. OSSM is thus consistent across regimes.

Satisfying the three consistency properties, OSSM provides an affirmative answer to RQ1.

Luenberger observer properties OSSM, in its formulation in Equation 4, is inspired by the Luenberger observer formalism.[1] Under Hypothesis 3.2 and while measurements are available, the observer formulation characterizes the evolution of the state-estimation error. Let $\boldsymbol{e}_t = \boldsymbol{x}_t - \hat{\boldsymbol{x}}_t$ denote the error term between the true latent dynamics and the estimated ones. We have the following recurrence:

$$\begin{aligned}
\boldsymbol{e}_{t+1} &= \boldsymbol{x}_{t+1} - \hat{\boldsymbol{x}}_{t+1} \\
&= \boldsymbol{A}_O \boldsymbol{x}_t - \left[\boldsymbol{A}_O \hat{\boldsymbol{x}}_t + m_t \boldsymbol{L}_O\big(\mathcal{E}(\boldsymbol{y}_t) - \hat{\boldsymbol{y}}_t\big)\right] \\
&= \boldsymbol{A}_O(\boldsymbol{x}_t - \hat{\boldsymbol{x}}_t) - m_t \boldsymbol{L}_O \boldsymbol{C}_O(\boldsymbol{x}_t - \hat{\boldsymbol{x}}_t) \\
&= (\boldsymbol{A}_O - m_t \boldsymbol{L}_O \boldsymbol{C}_O)\, \boldsymbol{e}_t.
\end{aligned} \tag{5}$$

[1] Formulation in Equation 3 is closer to the Luenberger observer formalism as the residual term is the comparison of two latent observations, while in Equation 4, a latent observation and one in the original space are compared.

During the observed context, $m_t = 1$, so

$$\boldsymbol{e}_{t+1} = (\boldsymbol{A}_O - \boldsymbol{L}_O\boldsymbol{C}_O)\boldsymbol{e}_t.$$

Consequently, if the spectral radius ρ of $(\boldsymbol{A}_O - \boldsymbol{L}_O\boldsymbol{C}_O)$ is lower than one, the error term can converge to zero, so the estimated system state can converge to the true one. In the Luenberger observer theory, the matrices $\boldsymbol{A}_O$ and $\boldsymbol{C}_O$ are already known. If $(\boldsymbol{A}_O, \boldsymbol{C}_O)$ is observable (see Appendix A for further details), then one can freely set the weights of $\boldsymbol{L}_O$ so the speed of convergence of the estimated state can be controlled.

Learning both to propagate and to correct In contrast to classical observer design, the latent matrices that define the system are not known beforehand. OSSMs jointly learn the latent dynamics $\boldsymbol{A}_O$, the measurement map $\boldsymbol{C}_O$, and the observer gain $\boldsymbol{L}_O$. Observability does not need to be explicitly enforced: observable pairs $(\boldsymbol{A}_O, \boldsymbol{C}_O)$ form an open dense subset of the parameter space and are therefore obtained almost surely under continuous random initialization. In the experiments, $(\boldsymbol{A}_O, \boldsymbol{C}_O)$ are initialized and $\boldsymbol{L}_O$ is set so the model starts with a chosen $\rho(\boldsymbol{A}_O - \boldsymbol{L}_O\boldsymbol{C}_O)$.

3.2 Coupled Dynamics in SSMs

SSM formulation A linear State-Space Model (SSM) defines the recurrent transformation

$$\boldsymbol{x}_{t+1} = \boldsymbol{A}_S\boldsymbol{x}_t + \boldsymbol{B}_S\boldsymbol{u}_t, \tag{6}$$

$$\boldsymbol{y}_{t+1} = \boldsymbol{C}_S\boldsymbol{x}_{t+1}, \tag{7}$$

with $\boldsymbol{u} \in \mathbb{R}^{d_u}$ the input control process and $\boldsymbol{B} \in \mathbb{R}^{d_x \times d_u}$ is the control matrix. When observations are not available anymore, $\boldsymbol{y}$ is fed to the system state equation through $\boldsymbol{B}$. During the learning process, $\boldsymbol{y}$ is learned to reproduce $\boldsymbol{u}$.

SSM as an entangled OSSM A closed-loop SSM can be expressed exactly as an OSSM under a change of coordinates. This equivalence is central to our formulation: OSSM does not enlarge the class of linear recurrent prediction heads, but directly parameterizes the autonomous dynamics used during forecasting separately from measurement correction.

Indeed, by setting

$$\boldsymbol{A}_O = \boldsymbol{A}_S + \boldsymbol{B}_S\boldsymbol{C}_S \tag{8}$$

$$\boldsymbol{L}_O = \boldsymbol{B}_S \tag{9}$$

$$\boldsymbol{C}_O = \boldsymbol{C}_S \tag{10}$$

We can recover the SSM as an OSSM. Similarly, SpaceTime (Zhang et al., 2023), an SSM designed specifically for time series forecasting purposes, can be rewritten as an OSSM (see Appendix B).[2]

Conventional closed-loop SSMs and OSSMs represent the same class of linear recurrent forecasting heads. Their difference lies in the parameterization: the former represents the context update through an input matrix and readout, whereas OSSM directly parameterizes the autonomous transition used during free-running forecasting and the gain that assimilates available measurements. This separation makes these two roles explicit and independently controllable at initialization.

Rewriting the SSM as an OSSM, while supposing that the observed (encoded) signal $\boldsymbol{u}$ is a measurement which corrects the internal state estimation, some modeling mismatches arise:

- the evolution matrix $\boldsymbol{A}_O$ depends on the observer matrix $\boldsymbol{L}_O$. However, $\boldsymbol{A}_O$ is independent of the observer, which is exogenous to the studied system and is added to tackle an estimation problem;
- the evolution matrix $\boldsymbol{A}_O$ also depends on the readout matrix. However, again, the dynamics matrix is also independent of $\boldsymbol{C}_O$, which represents the channels that are measured by the user.

[2] A recent analysis also connected SpaceTime to Luenberger observers under adversarial attacks (Anand & Pappas, 2026). In contrast, we consider the TSF task and SpaceTime arises as a particular instance of the OSSM class.

In classical control, the system dynamics are specified independently of the observer, which is subsequently designed using the dynamics and the available measurement map. Under an observer-based interpretation of conventional SSMs, however, these roles become entangled: the effective autonomous transition depends on parameters that also govern measurement assimilation and readout.

This observation motivates RQ2: Does explicitly separating autonomous propagation from measurement correction improve time series forecasting? We investigate this question through controlled comparisons between conventional SSMs and their observer-parameterized counterparts. Better results from the OSSM model compared to SSM would validate RQ2.

4 EXPERIMENTS

4.1 EXPERIMENTAL SETUP

The OSSM is designed to perform time-series predictions while preserving the dynamics from the context into the prediction interval. In general, SSM architectures consist of SSM encoder units that learn a representation of the input. The final layer reads the encoded input and performs the predictions.

We compare in the same settings two model families, illustrated in Figure 1:

- an OSSM, which is composed of SSM layers as encoders, and a final OSSM layer as the prediction head.
- an SSM, which is composed of the same encoding units, but the prediction head is an SSM.

Our experiments address a controlled question: for a fixed recurrent forecasting architecture, optimizer, parameter budget, and training protocol, does the observer parameterization improve the standard SSM prediction head? We evaluate the two model families on the TSF task using the TFB benchmark (Qiu et al., 2024), across four forecasting horizons, considering the ETT datasets, Electricity, Traffic, Solar, Exchange, Weather, and Wind for long-term forecasting, using a lookback window of 512, and ILI, NYSE, NASDAQ, and NN5 for short-term forecasting, using a lookback window of 104. Results are aggregated across 6 seeds and hyperparameter configurations using the median improvement from SSM to OSSM with the same configuration, and the fraction of pairs OSSM won. We report five complementary normalized metrics: MSE, MAE, WAPE, sMAPE, and msMAPE, enabling us to assess the consistency of the results across diverse metrics. Results are reported in Table 2. Detailed and raw results can be found in Appendix E.

To promote stable free-running dynamics at initialization, we control the spectral radius of the transition used during forecasting: $\rho(\boldsymbol{A}_O)$ for OSSM and $\rho(\boldsymbol{A}_S + \boldsymbol{B}_S\boldsymbol{C}_S)$ for the SSM baseline. In both cases, this radius is set to 0.9.

For OSSM, we additionally initialize the observer-error dynamics $\boldsymbol{A}_O - \boldsymbol{L}_O\boldsymbol{C}_O$ with spectral radius 0.91. This setting is part of the observer-structured parameterization: it makes both the free-running transition and the measurement-correction dynamics explicit at initialization. All parameters are subsequently optimized without spectral constraints.

4.2 RESULTS

Controlled comparison. Table 2 compares OSSM with its matched SSM counterpart across the full experimental grid. The two models have identical encoders, parameter counts, optimization settings, data splits, and random seeds; they differ only in the parameterization of the forecasting head. Consequently, the comparison isolates whether separating autonomous propagation from measurement correction is beneficial.

OSSM improves the five evaluation metrics on most datasets, with the largest and most consistent gains on Electricity, Traffic, NN5, Solar, and ETTh1. These improvements are especially pronounced at longer horizons, where errors accumulated during free-running prediction are more consequential. The effect is not universal: NASDAQ, ETTh2, and Weather exhibit near-parity or mixed outcomes. This heterogeneity is expected, since the benefit of observer correction depends on how informative the available context is about the latent state.

Table 2: Paired OSSM-SSM comparison over 5,376 runs (8 configurations, 4 horizons, and 6 seeds per dataset and head). We report the relative gain $\Delta_M = 100(M_{\text{SSM}} - M_{\text{OSSM}})/M_{\text{SSM}}$ as mean $\pm$ s.d. across seeds. Blue/red bold indicates a strictly positive/negative 1σ interval; black indicates overlap with zero. Win rate is pooled across all five metrics; a diverged head loses all five. Relative gains use only pairs where both heads converge, and thus conservatively underestimate OSSM's advantage.

Dataset	MSE	MAE	WAPE	sMAPE	msMAPE	Win rate
Short-term forecasting						
NN5	$\mathbf{16.1 \pm 2.1}$	$\mathbf{11.7 \pm 1.7}$	$\mathbf{11.9 \pm 1.7}$	$\mathbf{13.5 \pm 2.0}$	$\mathbf{13.6 \pm 2.0}$	94.9%
NYSE	$\mathbf{9.4 \pm 8.4}$	$\mathbf{8.5 \pm 4.3}$	$\mathbf{7.1 \pm 2.8}$	$\mathbf{6.8 \pm 3.2}$	$\mathbf{7.3 \pm 3.3}$	71.0%
ILI	3.1 ± 5.7	2.9 ± 3.8	3.4 ± 3.8	3.9 ± 4.7	4.0 ± 5.4	68.6%
NASDAQ	-1.7 ± 6.0	1.1 ± 3.6	1.2 ± 3.0	0.6 ± 2.4	0.6 ± 2.5	49.0%
Long-term forecasting						
Electricity	$\mathbf{22.9 \pm 4.6}$	$\mathbf{17.1 \pm 3.1}$	$\mathbf{17.0 \pm 3.1}$	$\mathbf{17.6 \pm 3.1}$	$\mathbf{18.4 \pm 3.3}$	89.9%
Solar	$\mathbf{15.0 \pm 2.8}$	$\mathbf{8.9 \pm 1.7}$	$\mathbf{8.9 \pm 1.7}$	$\mathbf{10.0 \pm 2.1}$	$\mathbf{10.2 \pm 2.1}$	69.1%
Traffic	$\mathbf{12.2 \pm 2.5}$	$\mathbf{11.8 \pm 3.0}$	$\mathbf{11.8 \pm 2.9}$	$\mathbf{14.9 \pm 4.2}$	$\mathbf{14.8 \pm 4.1}$	88.0%
ETTh1	$\mathbf{7.2 \pm 1.6}$	$\mathbf{3.9 \pm 0.7}$	$\mathbf{3.9 \pm 0.8}$	$\mathbf{4.2 \pm 1.0}$	$\mathbf{5.1 \pm 1.1}$	66.4%
ETTm1	3.8 ± 4.8	2.0 ± 2.1	$\mathbf{2.1 \pm 2.0}$	$\mathbf{1.8 \pm 1.7}$	$\mathbf{2.3 \pm 2.0}$	59.5%
Wind	$\mathbf{2.6 \pm 1.5}$	$\mathbf{1.7 \pm 0.9}$	$\mathbf{1.8 \pm 1.0}$	$\mathbf{1.3 \pm 0.6}$	$\mathbf{1.4 \pm 0.6}$	59.0%
ETTm2	$\mathbf{1.3 \pm 0.8}$	$\mathbf{0.7 \pm 0.3}$	$\mathbf{0.6 \pm 0.3}$	$\mathbf{0.9 \pm 0.2}$	$\mathbf{1.0 \pm 0.2}$	68.8%
Exchange	0.5 ± 5.8	2.2 ± 2.3	$\mathbf{3.9 \pm 2.3}$	$\mathbf{2.1 \pm 1.2}$	$\mathbf{3.1 \pm 1.4}$	67.9%
ETTh2	-0.6 ± 1.5	-0.2 ± 0.3	-0.2 ± 0.4	$\mathbf{0.5 \pm 0.4}$	0.3 ± 0.4	57.9%
Weather	-1.6 ± 9.1	$\mathbf{1.3 \pm 0.9}$	$\mathbf{1.3 \pm 0.8}$	$\mathbf{0.9 \pm 0.5}$	$\mathbf{1.2 \pm 0.7}$	54.3%

Robustness. Besides average error, OSSM is more reliable during optimization. Over the 5,376 matched runs, 96 OSSM runs diverged, compared with 225 SSM runs. Since both heads have the same capacity and training protocol, this difference suggests that the observer parameterization yields a more favorable recurrent optimization geometry and improves average forecasting accuracy across many datasets.

Interpretation. The results support our second research question in a controlled setting: explicitly distinguishing the dynamics used to propagate the state from the mechanism that assimilates observed measurements can improve recurrent forecasting. They do not imply that OSSM is uniformly preferable for every dataset, nor do they constitute a comparison with all modern TSF architectures.

5 Conclusion and Future Work

Observer State-Space Models (OSSMs) provide a recurrent formulation for time series forecasting in which observations serve as measurements that correct an estimated latent state, rather than as control inputs that define its dynamics. A single autonomous transition is retained across the observed context and the free-running forecast horizon, thereby satisfying the modeling, time-propagation, and regime-consistency properties introduced in this paper. This construction provides a constructive answer to RQ1. It also offers a unified interpretation of conventional SSMs and SpaceTime, revealing that their forecasting updates entangle latent-state propagation with measurement-related terms.

RQ2 is addressed through a strictly controlled evaluation across 14 datasets, 4 horizons, 8 configurations, and 6 seeds. The OSSM and SSM models share the same encoders, parameter counts, optimization protocol, data splits, random seeds, and initialization of free-running stability; only the forecasting-head parameterization differs. Across this 5,376-run grid, OSSM improves performance on most datasets and across five complementary metrics, with particularly consistent gains on Electricity, Traffic, NN5, Solar, and ETTh1. It also substantially reduces optimization failures, with 96 divergent runs versus 225 for the matched SSM baseline.

The results do not establish universal superiority over all forecasting architectures or datasets. Instead, they provide controlled evidence that explicitly separating state propagation from measurement correction is a useful inductive bias for recurrent forecasting, especially when the observed

context is informative about the latent state. Beyond contiguous forecasting, OSSMs naturally extend to intermittent observations, where autonomous propagation bridges missing intervals and new measurements correct the state estimate. Future work can combine this observer structure with stronger encoders and a more general SpaceTime-compatible formulation.

AI USE STATEMENT

In preparing this manuscript, we occasionally used suggestions from LLMs to improve clarity, grammar, and overall readability in both the main content and the code provided. We also used it to search for papers in the related work, especially on chaotic system identification and early adaptation of recurrent models in time series forecasting. All scientific content, including experimental design, data analysis, results, and interpretations, is independently developed by the authors.

ETHICS STATEMENT

This work is foundational research on already existing open-source models, tested on a public benchmark. This work has no impact on society at large, beyond aiming to get a better understanding of deep learning models. This work could reduce model energy consumption by guiding research on dynamic modeling.

REPRODUCIBILITY STATEMENT

For transparency and reproducibility, as detailed in the supplementary material in Appendix C, we base our code on the public repository developed by the TFB benchmark. Along with the paper, we provide the full code, including the studied models and their hyperparameters, used to evaluate them on the TSF task; it is included in the supplementary material.

REFERENCES

Sribalaji C. Anand and George J. Pappas. Adversarial robustness of deep state space models for forecasting, 2026. URL https://arxiv.org/abs/2604.03427.

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate, 2016. URL https://arxiv.org/abs/1409.0473.

Samy Bengio, Oriol Vinyals, Navdeep Jaitly, and Noam Shazeer. Scheduled sampling for sequence prediction with recurrent neural networks. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015. URL https://proceedings.neurips.cc/paper_files/paper/2015/file/e995f98d56967d946471af29d7bf99f1-Paper.pdf.

Alexis-Raja Brachet, Pierre-Yves Richard, and Céline Hudelot. Time series forecasting through the lens of dynamics. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=OL0Vx2FV1g.

Salva Rühling Cachay, Bo Zhao, Hailey Joren, and Rose Yu. DYffusion: A dynamics-informed diffusion model for spatiotemporal forecasting. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL https://openreview.net/forum?id=WRGldGm5Hz.

Asuncion Castano and Francisco Casacuberta. A connectionist approach to machine translation. *5th International Conference On Speech Communication And Technology (EUROSPEECH-97)*, 1997. doi: 10.21437/eurospeech.1997-50. URL https://doi.org/10.21437/eurospeech.1997-50.

Asunción Castaño, Francisco Casacuberta, and Enrique Vidal. Machine translation using neural networks and finite-state models. In *Proceedings of the 7th Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, St John's College, Santa Fe, July 23-25 1997. URL https://aclanthology.org/1997.tmi-1.19/.

Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In Alessandro Moschitti, Bo Pang, and Walter Daelemans (eds.), *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1724–1734, Doha, Qatar, October 2014. Association for Computational Linguistics. doi: 10.3115/v1/D14-1179. URL https://aclanthology.org/D14-1179/.

Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling, 2014. URL https://arxiv.org/abs/1412.3555.

Jeffrey L. Elman. Finding structure in time. *Cognitive Science*, 14(2):179-211, 1990. doi: 10.1207/s15516709cog1402_1. URL https://doi.org/10.1207/s15516709cog1402_1.

Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. In *First Conference on Language Modeling*, 2024. URL https://openreview.net/forum?id=tEYskw1VY2.

Albert Gu, Karan Goel, and Christopher Re. Efficiently modeling long sequences with structured state spaces. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=uYLFoz1vlAC.

Xingzhuo Guo, Yu Zhang, Baixu Chen, Haoran Xu, Jianmin Wang, and Mingsheng Long. Dynamical diffusion: Learning temporal dynamics with diffusion models. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=c5JZEPyFUE.

Priyam Gupta, Peter J. Schmid, Denis Sipp, Taraneh Sayadi, and Georgios Rigas. Mori-zwanzig latent space koopman closure for nonlinear autoencoder, 2024. URL https://arxiv.org/abs/2310.10745.

Andre Herz, Daniel Durstewitz, and Georgia Koppe. Teacher forcing as generalized bayes: Optimization geometry mismatch in switching surrogates for chaotic dynamics. *arXiv preprint arXiv:2604.25904*, 2026.

Hansika Hewamalage, Christoph Bergmeir, and Kasun Bandara. Recurrent neural networks for time series forecasting: Current status and future directions. *International Journal of Forecasting*, 37(1):388–427, 2021. ISSN 0169-2070. doi: https://doi.org/10.1016/j.ijforecast.2020.06.008. URL https://www.sciencedirect.com/science/article/pii/S0169207020300996.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Comput.*, 9(8):1735–1780, November 1997. ISSN 0899-7667. doi: 10.1162/neco.1997.9.8.1735. URL https://doi.org/10.1162/neco.1997.9.8.1735.

Jiaxi Hu, Yuehong HU, Wei Chen, Ming Jin, Shirui Pan, Qingsong Wen, and Yuxuan Liang. Attractor memory for long-term time series forecasting: A chaos perspective. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024a. URL https://openreview.net/forum?id=fEYHZzN7kX.

Jiaxi Hu, Disen Lan, Ziyu Zhou, Qingsong Wen, and Yuxuan Liang. Time-ssm: Simplifying and unifying state space models for time series forecasting, 2024b. URL https://arxiv.org/abs/2405.16312.

Ming Jin, Shiyu Wang, Lintao Ma, Zhixuan Chu, James Y. Zhang, Xiaoming Shi, Pin-Yu Chen, Yuxuan Liang, Yuan-Fang Li, Shirui Pan, and Qingsong Wen. Time-llm: Time series forecasting by reprogramming large language models, 2024. URL https://arxiv.org/abs/2310.01728.

Michael I. Jordan. *Serial Order : a parallel distributed processing approach*, pp. 471-495. 1997. doi: 10.1016/s0166-4115(97)80111-2. URL https://doi.org/10.1016/s0166-4115(97)80111-2.

R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45, 03 1960. ISSN 0021-9223. doi: 10.1115/1.3662552. URL https://doi.org/10.1115/1.3662552.

Nikolaos Kazantzis and Costas Kravaris. Nonlinear observer design using lyapunov's auxiliary theorem. *Systems & Control Letters*, 34(5):241–247, 1998.

B. O. Koopman. Hamiltonian systems and transformation in hilbert space. *Proceedings of the National Academy of Sciences*, 17(5):315–318, 1931. doi: 10.1073/pnas.17.5.315. URL https://www.pnas.org/doi/abs/10.1073/pnas.17.5.315.

Alex M Lamb, Anirudh Goyal ALIAS PARTH GOYAL, Ying Zhang, Saizheng Zhang, Aaron Courville, and Yoshua Bengio. Professor forcing: A new algorithm for training recurrent networks. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016. URL https://proceedings.neurips.cc/paper_files/paper/2016/file/16026d60ff9b54410b3435b403afd226-Paper.pdf.

Nikolay Laptev, Jason Yosinski, Li Erran Li, and Slawek Smyl. Time-series extreme event forecasting with neural networks at uber. In *International conference on machine learning*, volume 34, pp. 1–5. sn, 2017.

Yong Liu, Chenyu Li, Jianmin Wang, and Mingsheng Long. Koopa: Learning non-stationary time series dynamics with koopman predictors. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL https://openreview.net/forum?id=A4zzxu82a7.

D. Luenberger. Observers for multivariable systems. *IEEE Transactions on Automatic Control*, 11 (2):190–197, 1966. doi: 10.1109/TAC.1966.1098323.

David G. Luenberger. Observing the state of a linear system. *IEEE Transactions on Military Electronics*, 8(2):74–80, 1964. doi: 10.1109/TME.1964.4323124.

Abrar Majeedi, Viswanatha Reddy Gajjala, Satya Sai Srinath Namburi GNVV, Nada Magdi Elkordi, and Yin Li. LETS forecast: Learning embedology for time series forecasting. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=LLk1qYQatJ.

Jonas Mikhaeil, Zahra Monfared, and Daniel Durstewitz. On the difficulty of learning chaotic dynamics with rnns. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 11297–11312. Curran Associates, Inc., 2022. doi: 10.52202/068431-0821. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/495e55f361708bedbab5d81f92048dcd-Paper-Conference.pdf.

Muhammad Umar B Niazi, John Cao, Xudong Sun, Amritam Das, and Karl Henrik Johansson. Learning-based design of luenberger observers for autonomous nonlinear systems. *arXiv preprint arXiv:2210.01476*, 2022.

Jaideep Pathak, Brian Hunt, Michelle Girvan, Zhixin Lu, and Edward Ott. Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach. *Phys. Rev. Lett.*, 120:024102, Jan 2018. doi: 10.1103/PhysRevLett.120.024102. URL https://link.aps.org/doi/10.1103/PhysRevLett.120.024102.

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. *Proc. VLDB Endow.*, 17:2363 – 2377, 2024.

Xiangfei Qiu, Xingjian Wu, Yan Lin, Chenjuan Guo, Jilin Hu, and Bin Yang. Duet: Dual clustering enhanced multivariate time series forecasting. In *SIGKDD*, pp. 1185–1196, 2025.

Guy Revach, Nir Shlezinger, Xiaoyong Ni, Adria Lopez Escoriza, Ruud JG Van Sloun, and Yonina C Eldar. Kalmannet: Neural network aided kalman filtering for partially known dynamics. *IEEE Transactions on Signal Processing*, 70:1532–1547, 2022.

David Salinas, Valentin Flunkert, Jan Gasthaus, and Tim Januschowski. Deepar: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, 36(3):1181–1191, 2020. ISSN 0169-2070. doi: https://doi.org/10.1016/j.ijforecast.2019.07.001. URL https://www.sciencedirect.com/science/article/pii/S0169207019301888.

Martin Sundermeyer, Tamer Alkhouli, Joern Wuebker, and Hermann Ney. Translation modeling with bidirectional recurrent neural networks. In Alessandro Moschitti, Bo Pang, and Walter Daelemans (eds.), *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 14–25, Doha, Qatar, October 2014. Association for Computational Linguistics. doi: 10.3115/v1/D14-1003. URL https://aclanthology.org/D14-1003/.

Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. Sequence to sequence learning with neural networks. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K.Q. Weinberger (eds.), *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014. URL https://proceedings.neurips.cc/paper_files/paper/2014/file/5a18e133cbf9f257297f410bb7eca942-Paper.pdf.

Mingtian Tan, Mike A Merrill, Vinayak Gupta, Tim Althoff, and Thomas Hartvigsen. Are language models actually useful for time series forecasting? In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=DV15UbHCY1.

Philipp Teutsch and Patrick Mäder. Flipped classroom: Effective teaching for time series forecasting. *Transactions on Machine Learning Research*, 2022. ISSN 2835-8856. URL https://openreview.net/forum?id=w3x20YEcQK.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

Zixuan Weng, Jindong Han, Wenzhao Jiang, and Hao Liu. Sde: A simplified and disentangled dependency encoding framework for state space models in time series forecasting. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2*, KDD '25, pp. 3168–3179, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400714542. doi: 10.1145/3711896.3737119. URL https://doi.org/10.1145/3711896.3737119.

Ronald J. Williams and David Zipser. A learning algorithm for continually running fully recurrent neural networks. *Neural Computation*, 1(2):270–280, 1989. doi: 10.1162/neco.1989.1.2.270.

Xingjian Wu, Xiangfei Qiu, Hongfan Gao, Jilin Hu, Bin Yang, and Chenjuan Guo. K2vae: A koopman-kalman enhanced variational autoencoder for probabilistic time series forecasting. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=71Mm8GDGYd.

Xiongxiao Xu, Canyu Chen, Yueqing Liang, Baixiang Huang, Guangji Bai, Liang Zhao, and Kai Shu. Sst: Multi-scale hybrid mamba-transformer experts for time series forecasting. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management*, CIKM '25, pp. 3655–3665, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400720406. doi: 10.1145/3746252.3761394. URL https://doi.org/10.1145/3746252.3761394.

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*, 2023.

Michael Zhang, Khaled Kamal Saab, Michael Poli, Tri Dao, Karan Goel, and Christopher Re. Effectively modeling time series with simple discrete state spaces. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=2EpjkjzdCAa.

A LUENBERGER THEORETICAL DETAILS

The convergence of the state estimate is governed by the eigenvalues of $\boldsymbol{A} - \boldsymbol{LC}$. The pair $(\boldsymbol{A}, \boldsymbol{C})$ is observable if

$$\operatorname{rank}\left[\boldsymbol{C}^\top \quad (\boldsymbol{CA})^\top \quad \cdots \quad (\boldsymbol{CA}^{d_x-1})^\top\right]^\top = d_x. \tag{11}$$

When this condition holds, the eigenvalues of $\boldsymbol{A}-\boldsymbol{LC}$ can be freely set through $\boldsymbol{L}$, so the estimation error convergence speed can be controlled.

B SPACETIME THROUGH THE OSSM FORMULATION

B.1 SPACETIME FORECASTING MODEL

SpaceTime (Zhang et al., 2023) introduces an SSM architecture specifically designed for time series modeling. A central difference with standard deep SSMs is the parameterization of the transition matrix as a companion matrix. This structure provides a canonical representation of discrete-time autoregressive processes and allows the resulting SSM to represent AR dynamics that standard SSM parameterizations may fail to capture.

For forecasting, SpaceTime further introduces a closed-loop decoder. We focus here on this mechanism, as it is the component relevant to our OSSM interpretation. They suppose that the input and output are derived from two distinct processes. Thus, when measurements are no longer available, they don't feed the model with their outputs but with an estimate of the input process. Let $\bar{\boldsymbol{u}}_t$ denote the representation received by the decoder. Its state evolves as

$$\boldsymbol{x}_{t+1} = \boldsymbol{A}_{\mathrm{ST}}\boldsymbol{x}_t + \boldsymbol{B}_{\mathrm{ST}}\bar{\boldsymbol{u}}_t, \quad \hat{\boldsymbol{y}}_{t+1} = \boldsymbol{C}_{\mathrm{ST}}\boldsymbol{x}_{t+1} \tag{12}$$

In addition, SpaceTime learns a projection $\boldsymbol{K}_{\mathrm{ST}}$ to reconstruct the next decoder input,

$$\hat{\bar{\boldsymbol{u}}}_{t+1} = \boldsymbol{K}_{\mathrm{ST}}\boldsymbol{x}_{t+1} \tag{13}$$

The resulting autonomous dynamics are therefore

$$\boldsymbol{x}_{t+1} = \left(\boldsymbol{A}_{\mathrm{ST}} + \boldsymbol{B}_{\mathrm{ST}}\boldsymbol{K}_{\mathrm{ST}}\right)\boldsymbol{x}_t \tag{14}$$

B.2 OSSM REFORMULATION

Following the computations performed in Section 3.2, SpaceTime can be written as a particular representation-space OSSM. Since $\bar{\boldsymbol{u}}_t$ is the decoder input representation, while $\boldsymbol{y}_t$ is the output process, we distinguish the map used in the innovation term from the output readout:

$$\boldsymbol{A}_O = \boldsymbol{A}_{\mathrm{ST}} + \boldsymbol{B}_{\mathrm{ST}}\boldsymbol{K}_{\mathrm{ST}}, \tag{15}$$

$$\boldsymbol{L}_O = \boldsymbol{B}_{\mathrm{ST}}, \tag{16}$$

$$\boldsymbol{K}_O = \boldsymbol{K}_{\mathrm{ST}}, \tag{17}$$

$$\boldsymbol{C}_O = \boldsymbol{C}_{\mathrm{ST}}. \tag{18}$$

Here, $\boldsymbol{K}_O$ reconstructs the decoder input representation $\bar{\boldsymbol{u}}_t$, whereas $\boldsymbol{C}_O$ remains the output readout, consistently with the OSSM formulation.

Indeed, in the context,

$$\hat{\boldsymbol{x}}_{t+1} = \left(\boldsymbol{A}_{\mathrm{ST}} + \boldsymbol{B}_{\mathrm{ST}}\boldsymbol{K}_{\mathrm{ST}}\right)\hat{\boldsymbol{x}}_t + \boldsymbol{B}_{\mathrm{ST}}\left(\bar{\boldsymbol{u}}_t - \boldsymbol{K}_{\mathrm{ST}}\hat{\boldsymbol{x}}_t\right) \tag{19}$$

$$= \boldsymbol{A}_{\mathrm{ST}}\hat{\boldsymbol{x}}_t + \boldsymbol{B}_{\mathrm{ST}}\bar{\boldsymbol{u}}_t, \tag{20}$$

which is the original SpaceTime update. The corresponding output is $\hat{\boldsymbol{y}}_{t+1} = \boldsymbol{C}_{\mathrm{ST}}\hat{\boldsymbol{x}}_{t+1}$. Once observations are removed, $m_t = 0$ and the OSSM evolves autonomously as

$$\hat{\boldsymbol{x}}_{t+1} = \left(\boldsymbol{A}_{\mathrm{ST}} + \boldsymbol{B}_{\mathrm{ST}}\boldsymbol{K}_{\mathrm{ST}}\right)\hat{\boldsymbol{x}}_t, \tag{21}$$

which is precisely the closed-loop transition used by SpaceTime for forecasting.

This reformulation exposes the same limitation as in the SSM formulation: the autonomous forecasting transition is coupled to the innovation parameterization through $\boldsymbol{B}_{\mathrm{ST}}\boldsymbol{K}_{\mathrm{ST}}$. Hence, the free-running dynamics are not learned independently of the map used to predict the decoder input representation.

Moreover, although $\boldsymbol{A}_{\mathrm{ST}}$ is constrained to a companion form, the transition governing free-running forecasting, $\boldsymbol{A}_{\mathrm{ST}} + \boldsymbol{B}_{\mathrm{ST}}\boldsymbol{K}_{\mathrm{ST}}$, is generally not. The companion constraint therefore characterizes the context update, whereas the autonomous dynamics used in practice during forecasting follow a different, unconstrained transition.

This correspondence also highlights a limitation of SpaceTime as an OSSM instance: the innovation is defined in the decoder-input representation space through $\bar{\boldsymbol{u}}_t - \boldsymbol{K}_O\hat{\boldsymbol{x}}_t$, whereas the final prediction is read out through $\boldsymbol{C}_O\hat{\boldsymbol{x}}_t$. A more general OSSM formulation may therefore parameterize the innovation map and the output readout independently. Evaluating this decoupled parameterization within the complete SpaceTime architecture would require a dedicated parameter-matched experimental study, which we leave for future work.

C Code and implementation Details

Code The code is available in the supplementary. The code is based on the repository[3] developed for the TFB benchmark (Qiu et al., 2024) under the MIT license. We only provide the `\ossm` folder. To run the model, one should pull the TFB repository and add the ossm folder into the `\ts_benchmark\baselines\` folder. Models are trained using the standard `deep_forecasting_model_base.py` file. In the `\ossm\functional\` folder, we implement the functions developed in the SpaceTime (Zhang et al., 2023) repository.[4]

Everything was done in accordance with the MIT license, which grants the right to use the code and datasets of the TFB benchmark without restriction.

Everything was done under the Apache-2.0 license for the use of the SpaceTime code.

Resources and implementation All the considered models were trained on Rocky Linux 8.10, on either:

- Intel Xeon Gold 6230 20C @ 2.1GHz with 768 GB memory with up to four NVIDIA Tesla V100 with 32 GB VRAM,
- Intel Xeon Gold 6346 16C @ 3.1GHz with 1024 GB memory with up to four Nvidia HGX A100 with 40 GB memory.

Training was done in PyTorch[5], using the default learning procedure set in the `deep_forecasting_model_base.py` file. Hyperparameters can be found in the `.\scripts` folder presented in `.sh` files and detailed model configuration in Appendix E. Datasets are split following the rolling forecast method to prevent information leakage from the future to the past. We needed at most 1 GPUs with 15 CPUs per task, running for several minutes.

D Experiment details

D.1 Details on the Datasets

We provide, in Table 3, the detailed statistics of the datasets used in our experiments. It includes the domain, the sampling frequency, the variate dimension, and the data split for training, validation, and testing.

D.2 OSSM vs SSM

Encoding units The encoding units for both models are standard SSM, which are freely learned during the learning procedure:

$$\boldsymbol{x}_{t+1} = \boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{B}\boldsymbol{u}_{t+1}, \quad \boldsymbol{y}_t = \boldsymbol{C}\boldsymbol{x}_t$$

[3] https://github.com/decisionintelligence/TFB (pulled in September 2026)

[4] https://github.com/HazyResearch/spacetime/

[5] Torch version was 2.5.1 with cuda toolkit 12.1 and Python 3.11

Table 3: Statistics of multivariate datasets of the TFB benchmark (Qiu et al., 2024) taken from (Qiu et al., 2025).

Dataset	Domain	Frequency	Lengths	Dim	Split	Description
Traffic	Traffic	1 hour	17,544	862	7:1:2	Road occupancy rates measured by 862 sensors on San Francisco Bay area freeways
ETTh1	Electricity	1 hour	14,400	7	6:2:2	Power transformer 1, comprising seven indicators such as oil temperature and useful load
ETTh2	Electricity	1 hour	14,400	7	6:2:2	Power transformer 2, comprising seven indicators such as oil temperature and useful load
ETTm1	Electricity	15 mins	57,600	7	6:2:2	Power transformer 1, comprising seven indicators such as oil temperature and useful load
ETTm2	Electricity	15 mins	57,600	7	6:2:2	Power transformer 2, comprising seven indicators such as oil temperature and useful load
Electricity	Electricity	1 hour	26,304	321	7:1:2	Electricity records the electricity consumption in kWh every 1 hour from 2012 to 2014
Solar	Energy	10 mins	52,560	137	6:2:2	Solar production records collected from 137 PV plants in Alabama
Wind	Energy	15 mins	48,673	7	7:1:2	Wind power records from 2020-2021 at 15-minute intervals
Weather	Environment	10 mins	52,696	21	7:1:2	Recorded every for the whole year 2020, which contains 21 meteorological indicators
Exchange	Economic	1 day	7,588	8	7:1:2	ExchangeRate collects the daily exchange rates of eight countries
NASDAQ	Stock	1 day	1,244	5	7:1:2	Records opening price, closing price, trading volume, lowest price, and highest price
NYSE	Stock	1 day	1,243	5	7:1:2	Records opening price, closing price, trading volume, lowest price, and highest price
NN5	Banking	1 day	791	111	7:1:2	NN5 is from banking, records the daily cash withdrawals from ATMs in UK
ILI	Health	1 week	966	7	7:1:2	Recorded indicators of patients' data from Centers for Disease Control and Prevention

A first feed-forward acts on the channel dimension to prepare the data for the SSM units in series. Then, each SSM unit maps $\boldsymbol{u}^l \to \boldsymbol{u}^{l+1}$, where l is the l^{th} encoding layer.

Prediction head As detailed in the main part, the prediction is either an OSSM or an SSM as described.

E RAW RESULTS

We report every individual run behind the aggregate results. The grid is fully crossed: 14 datasets $\times$ 4 forecast horizons $\times$ 6 random seeds $\times$ 8 architecture configurations $\times$ 2 prediction heads, for a total of 5376 runs. The observer head (OSSM) and the standard head (SSM) share the architecture, the optimizer, the data splits and the seeds, so each (dataset, horizon, seed) cell is a strictly paired comparison, and only the head differs.

All metrics are the normalized MSE and MAE reported by TFB on the test split; Lower is better. Within each cell, the better of the two heads is in bold. A run is called *divergent* when its normalized MSE exceeds 10 or when it produced no metric at all; such runs are kept in the tables, marked "div.", and counted as a loss for the head that diverged. Across the whole grid 321 runs diverged, 225 of them from the SSM head and 96 from the OSSM head. Where a cell was run more than once, the most recent run is reported.

E.1 CONFIGURATION 1

Table 4: Hyper-parameters of configuration 1. Every other setting follows the TFB defaults, and the two heads share all of them, including the random seeds.

Hyper-parameter	Value
Encoder layers e_{layers}	2
State dimension d_x	64
Model dimension d_{model}	64
Learning rate	10^{-4}
Batch size	32

Table 5: Normalized MSE for configuration 1, lower is better. Rows are (dataset, forecast horizon H) pairs, columns are the six random seeds, each split into the observer head (OSSM) and the standard head (SSM). The better of the two is in bold. "div." marks a run that diverged, i.e. reached a normalized MSE above 10 or produced no metric at all; divergence counts as a loss, so the head that converged is shown in bold.

Dataset	H	2021		2022		2023		2024		2025		2026	
		OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM
Short-term forecasting													
NN5	24	**1.201**	1.372	**1.279**	1.403	**1.257**	1.378	**1.250**	1.364	**1.208**	1.332	**1.205**	1.354
	36	**1.203**	1.330	**1.275**	1.353	**1.230**	1.336	**1.240**	1.325	**1.181**	1.310	**1.175**	1.319
	48	**1.201**	1.313	**1.276**	1.333	**1.218**	1.320	**1.238**	1.309	**1.165**	1.296	**1.157**	1.307
	60	**1.208**	1.305	**1.278**	1.321	**1.212**	1.312	**1.241**	1.301	**1.161**	1.292	**1.150**	1.302
NYSE	24	**0.382**	1.020	**0.274**	0.631	0.424	**0.320**	0.356	**0.343**	0.325	**0.287**	0.392	**0.328**
	36	**1.013**	1.346	**0.424**	0.714	**0.470**	0.536	**1.058**	1.319	**0.563**	0.565	0.553	**0.499**
	48	**0.858**	1.659	1.084	**0.996**	**0.699**	0.881	1.436	**1.188**	**1.549**	1.638	1.541	**0.581**
	60	**1.621**	1.978	**1.496**	1.583	**0.979**	1.345	**1.671**	1.918	**1.815**	2.012	1.850	**1.590**
ILI	24	**2.524**	3.197	2.996	**2.790**	3.633	**3.079**	**2.942**	3.896	**3.073**	3.599	**2.485**	2.935
	36	**2.704**	3.579	3.334	**2.761**	3.608	**3.241**	**3.002**	4.047	**3.077**	3.525	**2.782**	3.207
	48	**2.664**	3.763	3.659	**2.771**	3.787	**3.256**	**3.022**	4.115	**3.455**	3.520	3.906	**3.147**
	60	**2.843**	3.948	3.941	**2.884**	4.172	**4.106**	**3.123**	4.300	**3.803**	3.857	3.891	**3.353**
NASDAQ	24	0.547	**0.486**	0.544	**0.529**	0.548	**0.509**	0.552	**0.513**	0.559	**0.516**	0.563	**0.524**
	36	0.822	**0.754**	**0.847**	0.874	0.941	**0.791**	0.807	**0.799**	0.884	**0.818**	0.854	**0.786**
	48	1.068	**1.018**	1.151	**1.118**	1.187	**1.087**	**1.070**	1.126	1.126	**1.084**	1.143	**1.091**
	60	1.204	**1.173**	1.273	**1.268**	1.252	**1.181**	**1.155**	1.203	1.265	**1.216**	1.292	**1.159**
Long-term forecasting													
Electricity	96	**0.272**	0.346	**0.337**	0.345	**0.274**	0.338	**0.269**	0.345	0.344	**0.279**	**0.273**	0.347
	192	**0.284**	0.355	**0.354**	0.355	**0.352**	0.354	**0.345**	0.355	**0.354**	0.354	**0.350**	0.355
	336	**0.362**	0.367	**0.363**	0.847	**0.363**	0.846	**0.362**	0.368	**0.363**	0.845	**0.362**	0.851
	720	**0.389**	0.889	**0.390**	0.883	**0.866**	0.888	**0.388**	0.400	**0.388**	0.841	**0.388**	0.881
Solar	96	0.302	**0.258**	0.369	**0.277**	0.328	**0.323**	**0.319**	0.322	0.255	**0.244**	**0.239**	0.280
	192	**0.379**	0.384	**0.389**	0.399	**0.380**	0.386	**0.377**	0.383	**0.371**	0.375	0.380	**0.369**
	336	**0.382**	0.406	**0.380**	0.531	**0.381**	0.419	**0.380**	0.416	**0.377**	0.405	**0.385**	0.467
	720	0.444	**0.430**	**0.711**	0.754	**0.374**	0.739	**0.391**	0.747	**0.379**	0.760	**0.375**	0.732
Traffic	96	**0.828**	0.953	**0.780**	0.945	**0.836**	0.946	**0.941**	0.941	**0.943**	0.948	**0.938**	0.951
	192	**0.965**	0.977	**0.971**	0.978	**0.863**	0.978	**0.957**	0.970	0.975	**0.969**	**0.969**	0.976
	336	**0.982**	1.384	**0.986**	1.047	**0.992**	0.992	**0.979**	1.393	**0.978**	1.393	**0.986**	1.374
	720	**1.001**	1.415	**1.015**	1.432	**1.021**	1.430	**1.006**	1.431	**1.025**	1.404	**1.009**	1.412
ETTh1	96	**0.501**	0.517	0.507	**0.504**	0.546	**0.515**	**0.508**	0.510	**0.513**	0.514	0.539	**0.519**
	192	0.668	**0.541**	0.525	**0.518**	**0.530**	0.568	0.570	**0.536**	**0.524**	0.530	0.724	**0.533**
	336	**0.560**	0.707	0.547	**0.522**	0.725	**0.601**	0.709	**0.706**	**0.555**	0.712	0.709	**0.604**
	720	0.722	**0.721**	**0.617**	0.716	0.756	**0.730**	0.726	**0.726**	0.733	**0.721**	**0.719**	0.722
ETTm1	96	0.471	**0.458**	0.465	**0.459**	0.456	**0.452**	0.465	**0.446**	0.479	**0.455**	0.463	**0.458**
	192	**0.483**	0.497	**0.495**	0.496	0.496	**0.484**	0.517	**0.489**	**0.477**	0.494	**0.493**	0.493
	336	0.685	**0.662**	**0.506**	0.538	**0.510**	0.566	**0.537**	0.555	**0.515**	0.525	0.513	**0.512**
	720	**0.529**	0.584	**0.543**	0.554	**0.539**	0.714	**0.703**	0.708	**0.545**	0.714	**0.543**	0.711
Wind	96	0.970	**0.898**	**0.925**	0.927	0.955	**0.914**	0.932	**0.928**	0.945	**0.885**	0.936	**0.898**
	192	**1.105**	1.115	**1.105**	1.114	1.121	**1.101**	**1.098**	1.110	**1.099**	1.127	1.105	**1.041**
	336	1.262	**1.249**	1.237	**1.205**	1.255	**1.254**	1.263	**1.259**	1.240	**1.231**	1.303	**1.259**
	720	**1.343**	1.358	**1.295**	1.349	1.388	**1.364**	**1.350**	1.374	1.373	**1.354**	**1.381**	1.415
ETTm2	96	0.252	**0.251**	**0.239**	0.243	0.256	**0.247**	0.257	**0.247**	**0.250**	0.256	0.258	**0.248**
	192	0.316	**0.297**	**0.288**	0.296	0.304	**0.290**	**0.294**	0.296	0.303	**0.288**	**0.290**	0.304
	336	0.345	**0.332**	0.341	**0.335**	**0.345**	0.356	0.343	**0.337**	**0.342**	0.348	0.348	**0.342**
	720	**0.420**	0.423	**0.420**	0.421	0.428	**0.423**	**0.407**	0.419	**0.417**	0.420	**0.419**	0.423

Table 5, continued.

Dataset	H	2021		2022		2023		2024		2025		2026	
		OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM
Exchange	96	0.228	**0.202**	**0.336**	0.337	0.298	**0.286**	**0.251**	0.269	0.297	**0.276**	0.303	**0.287**
	192	**0.375**	0.406	0.634	**0.400**	0.616	**0.422**	**0.396**	0.459	0.479	**0.342**	0.476	**0.409**
	336	**0.709**	0.710	1.002	**0.749**	**0.678**	0.743	**0.590**	0.716	0.840	**0.665**	0.810	**0.713**
	720	**1.280**	1.281	**1.343**	1.344	1.321	**1.320**	**1.304**	1.314	**1.345**	1.356	1.370	**1.370**
ETTh2	96	0.368	**0.362**	0.363	**0.355**	0.373	**0.372**	0.371	**0.370**	0.378	**0.367**	**0.362**	0.364
	192	0.388	**0.381**	0.390	**0.387**	0.412	**0.389**	**0.374**	0.406	0.418	**0.399**	0.390	**0.385**
	336	0.398	**0.387**	0.395	**0.386**	0.394	**0.385**	0.399	**0.383**	**0.391**	0.432	**0.380**	0.392
	720	**0.431**	0.432	0.441	**0.430**	0.432	**0.428**	0.422	**0.421**	**0.432**	0.433	**0.422**	0.423
Weather	96	0.215	**0.198**	0.199	**0.192**	0.198	**0.195**	0.199	**0.196**	0.197	**0.192**	0.200	**0.194**
	192	0.241	**0.234**	**0.235**	0.275	0.282	**0.234**	**0.239**	0.279	**0.232**	0.278	0.233	**0.230**
	336	**0.286**	0.287	**0.284**	0.309	0.311	**0.293**	0.321	**0.307**	0.308	**0.307**	**0.278**	0.285
	720	0.353	**0.346**	0.353	**0.345**	**0.353**	0.368	**0.351**	0.367	**0.340**	0.349	**0.352**	0.367

Table 6: Normalized MAE for configuration 1, lower is better. Rows are (dataset, forecast horizon H) pairs, columns are the six random seeds, each split into the observer head (OSSM) and the standard head (SSM). The better of the two is in bold. "div." marks a run that diverged, i.e. reached a normalized MSE above 10 or produced no metric at all; divergence counts as a loss, so the head that converged is shown in bold.

Dataset	H	2021		2022		2023		2024		2025		2026	
		OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM
Short-term forecasting													
NN5	24	**0.854**	0.944	**0.904**	0.956	**0.888**	0.951	**0.883**	0.943	**0.862**	0.925	**0.865**	0.942
	36	**0.867**	0.934	**0.913**	0.944	**0.884**	0.940	**0.888**	0.934	**0.857**	0.927	**0.859**	0.934
	48	**0.871**	0.932	**0.917**	0.940	**0.882**	0.937	**0.892**	0.931	**0.853**	0.925	**0.854**	0.932
	60	**0.878**	0.931	**0.921**	0.938	**0.881**	0.937	**0.897**	0.931	**0.854**	0.927	**0.853**	0.933
NYSE	24	**0.455**	0.767	**0.351**	0.587	0.469	**0.419**	0.432	**0.422**	0.404	**0.384**	0.434	**0.422**
	36	**0.751**	0.903	**0.463**	0.611	**0.485**	0.558	**0.760**	0.897	0.565	**0.545**	0.524	**0.516**
	48	**0.720**	1.022	0.783	**0.735**	**0.605**	0.730	0.922	**0.819**	**0.976**	1.017	0.968	**0.559**
	60	**0.994**	1.131	**0.936**	0.993	**0.726**	0.952	**1.003**	1.113	**1.067**	1.148	1.075	**0.981**
ILI	24	**1.099**	1.315	1.250	**1.183**	1.339	**1.257**	**1.193**	1.462	**1.222**	1.358	**1.104**	1.203
	36	**1.115**	1.381	1.327	**1.157**	1.346	**1.294**	**1.197**	1.481	**1.222**	1.318	**1.174**	1.248
	48	**1.077**	1.409	1.390	**1.159**	1.381	**1.286**	**1.193**	1.483	**1.298**	1.324	1.433	**1.223**
	60	**1.124**	1.438	1.438	**1.179**	1.453	**1.448**	**1.212**	1.508	**1.378**	1.391	1.450	**1.263**
NASDAQ	24	0.560	**0.524**	0.536	**0.531**	0.547	**0.538**	0.567	**0.536**	0.557	**0.535**	0.555	**0.535**
	36	0.685	**0.649**	**0.676**	0.677	0.706	**0.667**	0.677	**0.661**	0.684	**0.664**	0.684	**0.659**
	48	0.764	**0.744**	0.771	**0.764**	0.791	**0.767**	**0.762**	0.772	0.769	**0.761**	0.779	**0.756**
	60	0.806	**0.793**	0.815	**0.812**	0.813	**0.800**	**0.792**	0.803	0.813	**0.802**	0.818	**0.785**
Long-term forecasting													
Electricity	96	**0.367**	0.431	**0.423**	0.429	**0.368**	0.423	**0.363**	0.430	0.428	**0.373**	**0.366**	0.432
	192	**0.376**	0.436	**0.435**	0.436	**0.434**	0.436	**0.427**	0.436	**0.434**	0.436	**0.432**	0.436
	336	**0.439**	0.444	**0.441**	0.756	**0.441**	0.756	**0.441**	0.445	**0.441**	0.755	**0.440**	0.759
	720	**0.456**	0.774	**0.457**	0.770	**0.761**	0.774	**0.457**	0.465	**0.456**	0.746	**0.456**	0.769
Solar	96	0.419	**0.342**	0.447	**0.386**	0.442	**0.425**	0.435	**0.420**	0.363	**0.340**	**0.349**	0.392
	192	0.487	**0.486**	0.506	**0.497**	**0.486**	0.489	0.487	**0.483**	**0.480**	0.482	0.485	**0.475**
	336	**0.493**	0.503	**0.485**	0.563	**0.494**	0.517	**0.489**	0.510	**0.483**	0.501	**0.493**	0.522
	720	0.530	**0.514**	**0.669**	0.696	**0.484**	0.686	**0.496**	0.691	**0.487**	0.699	**0.488**	0.683
Traffic	96	**0.490**	0.540	**0.458**	0.539	**0.489**	0.539	0.537	**0.536**	**0.537**	0.538	**0.534**	0.541
	192	**0.543**	0.546	**0.544**	0.547	**0.495**	0.545	**0.541**	0.544	**0.541**	0.542	**0.543**	0.545
	336	**0.547**	0.787	**0.549**	0.582	**0.542**	0.559	**0.543**	0.793	**0.542**	0.792	**0.547**	0.784
	720	**0.552**	0.797	**0.553**	0.805	**0.553**	0.805	**0.553**	0.805	**0.555**	0.790	**0.555**	0.796

Table 6, continued.

Dataset	H	2021		2022		2023		2024		2025		2026	
		OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM
ETTh1	96	**0.489**	0.504	**0.491**	0.494	0.528	**0.501**	**0.494**	0.501	**0.498**	0.502	0.517	**0.504**
	192	0.562	**0.519**	0.510	**0.505**	**0.507**	0.530	0.541	**0.516**	**0.509**	0.514	0.580	**0.512**
	336	**0.534**	0.588	0.521	**0.511**	0.596	**0.545**	**0.588**	0.588	**0.530**	0.594	0.586	**0.552**
	720	**0.610**	0.612	**0.570**	0.613	0.641	**0.618**	0.617	**0.616**	0.624	**0.615**	**0.608**	0.613
ETTm1	96	0.468	**0.460**	0.464	**0.461**	0.464	**0.462**	0.473	**0.460**	0.477	**0.466**	0.466	**0.465**
	192	**0.475**	0.483	**0.479**	0.484	0.486	**0.475**	0.496	**0.481**	**0.474**	0.483	0.485	**0.485**
	336	0.550	**0.542**	**0.486**	0.504	**0.486**	0.508	**0.504**	0.504	**0.490**	0.499	0.496	**0.488**
	720	**0.499**	0.519	**0.506**	0.510	**0.504**	0.568	**0.564**	0.567	**0.503**	0.570	**0.505**	0.566
Wind	96	0.704	**0.664**	0.679	**0.676**	0.703	**0.678**	0.691	**0.675**	0.695	**0.659**	0.680	**0.659**
	192	**0.771**	0.772	**0.768**	0.776	0.777	**0.767**	**0.762**	0.770	**0.768**	0.782	0.770	**0.738**
	336	0.833	**0.831**	0.837	**0.817**	0.831	**0.830**	0.830	**0.830**	0.829	**0.820**	0.858	**0.831**
	720	**0.865**	0.869	**0.853**	0.867	0.882	**0.872**	**0.864**	0.874	0.877	**0.868**	**0.878**	0.890
ETTm2	96	0.331	**0.331**	**0.326**	0.329	0.337	**0.331**	0.338	**0.331**	**0.332**	0.337	0.338	**0.329**
	192	0.368	**0.358**	**0.356**	0.358	0.361	**0.358**	**0.358**	0.359	0.363	**0.357**	**0.355**	0.362
	336	0.386	**0.379**	0.385	**0.381**	**0.384**	0.392	0.385	**0.382**	**0.385**	0.388	0.388	**0.383**
	720	**0.428**	0.429	**0.427**	0.429	0.433	**0.430**	**0.422**	0.429	**0.428**	0.428	**0.429**	0.429
Exchange	96	0.358	**0.344**	0.429	**0.425**	**0.388**	0.392	**0.362**	0.387	0.397	**0.386**	0.400	**0.384**
	192	**0.464**	0.478	0.605	**0.469**	0.562	**0.482**	**0.474**	0.514	0.499	**0.437**	0.517	**0.467**
	336	**0.628**	0.658	0.764	**0.676**	**0.626**	0.674	**0.581**	0.656	0.687	**0.623**	0.684	**0.617**
	720	**0.852**	0.853	**0.872**	0.873	0.868	**0.867**	**0.857**	0.862	**0.872**	0.877	0.881	**0.880**
ETTh2	96	0.423	**0.417**	0.413	**0.412**	**0.421**	0.422	0.420	**0.420**	0.425	**0.420**	**0.412**	0.416
	192	0.427	**0.424**	0.430	**0.424**	0.443	**0.431**	**0.419**	0.439	0.446	**0.434**	0.429	**0.425**
	336	0.438	**0.429**	0.434	**0.428**	0.434	**0.429**	0.439	**0.425**	**0.434**	0.460	**0.424**	0.430
	720	**0.457**	0.458	0.464	**0.457**	0.461	**0.458**	0.452	**0.451**	**0.461**	0.461	**0.453**	0.454
Weather	96	0.268	**0.253**	0.258	**0.251**	0.253	**0.253**	0.258	**0.252**	0.258	**0.249**	0.259	**0.257**
	192	0.292	**0.280**	**0.286**	0.314	0.320	**0.286**	**0.288**	0.316	**0.281**	0.315	0.285	**0.283**
	336	0.327	**0.325**	**0.321**	0.336	0.337	**0.326**	0.345	**0.333**	0.334	**0.333**	**0.314**	0.320
	720	0.365	**0.361**	0.366	**0.361**	**0.364**	0.379	**0.367**	0.377	**0.355**	0.365	**0.367**	0.378

E.2 CONFIGURATION 2

Table 7: Hyper-parameters of configuration 2. Every other setting follows the TFB defaults, and the two heads share all of them, including the random seeds.

Hyper-parameter	Value
Encoder layers e_{layers}	2
State dimension d_x	64
Model dimension d_{model}	128
Learning rate	10^{-4}
Batch size	32

Table 8: normalized MSE for configuration 2, lower is better. Rows are (dataset, forecast horizon H) pairs, columns are the six random seeds, each split into the observer head (OSSM) and the standard head (SSM). The better of the two is in bold. "div." marks a run that diverged, i.e. reached a normalized MSE above 10 or produced no metric at all; divergence counts as a loss, so the head that converged is shown in bold.

Dataset	H	2021		2022		2023		2024		2025		2026	
		OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM
Short-term forecasting													
NN5	24	**1.156**	1.313	**1.152**	1.322	**1.196**	1.272	**1.096**	1.277	**1.126**	1.365	**1.198**	1.281
	36	**1.188**	1.304	**1.197**	1.299	**1.222**	1.276	**1.082**	1.253	**1.113**	1.322	**1.230**	1.287
	48	**1.205**	1.296	**1.226**	1.290	**1.226**	1.264	**1.067**	1.239	**1.103**	1.308	**1.235**	1.269
	60	**1.221**	1.289	**1.230**	1.286	**1.230**	1.273	**1.061**	1.233	**1.100**	1.301	**1.247**	1.276
NYSE	24	0.270	**0.253**	0.390	**0.284**	0.315	**0.309**	0.341	**0.324**	0.821	**0.440**	0.427	**0.313**
	36	0.678	**0.415**	1.346	**1.014**	**0.468**	0.605	0.607	**0.532**	**0.532**	0.712	0.704	**0.406**
	48	1.083	**0.893**	1.650	**1.008**	0.887	**0.610**	**0.812**	1.042	**0.782**	0.926	1.526	**0.622**
	60	1.369	**1.190**	1.960	**1.462**	1.724	**0.890**	**1.331**	1.720	1.864	**1.327**	1.844	**0.841**
ILI	24	**2.959**	3.218	2.474	**2.469**	**2.640**	2.685	2.571	**2.536**	2.658	**2.313**	**2.535**	2.738
	36	2.952	**2.949**	**2.531**	2.738	**2.682**	2.743	**2.535**	2.565	2.567	**2.474**	**2.544**	3.045
	48	3.169	**2.840**	**2.613**	2.828	**2.620**	2.725	**2.523**	2.673	2.595	**2.574**	**2.486**	2.947
	60	3.582	**2.918**	**2.662**	3.013	**2.682**	2.923	**2.602**	2.767	**2.592**	2.825	**2.534**	3.507
NASDAQ	24	0.582	**0.478**	0.558	**0.484**	0.545	**0.466**	0.488	**0.420**	**0.508**	0.509	0.516	**0.480**
	36	0.850	**0.749**	0.826	**0.751**	0.793	**0.763**	0.813	**0.674**	**0.795**	0.795	0.798	**0.746**
	48	1.322	**1.036**	1.054	**1.034**	1.115	**1.084**	1.032	**1.000**	**1.060**	1.078	1.095	**1.075**
	60	1.286	**1.176**	1.207	**1.138**	1.236	**1.210**	1.310	**1.141**	1.399	**1.200**	**1.194**	1.271
Long-term forecasting													
Electricity	96	**0.271**	0.277	**0.271**	0.344	**0.274**	0.278	**0.272**	0.277	**0.271**	0.276	**0.272**	0.340
	192	**0.350**	0.356	**0.280**	0.355	**0.284**	0.354	**0.350**	0.355	**0.348**	0.356	**0.281**	0.354
	336	**0.361**	0.851	**0.361**	0.370	**0.361**	0.366	**0.362**	0.838	**0.361**	0.855	**0.360**	0.843
	720	**0.390**	0.889	**0.388**	0.889	**0.388**	0.866	**0.389**	0.882	**0.862**	0.883	**0.389**	0.827
Solar	96	**0.201**	0.219	**0.255**	0.261	0.265	**0.229**	**0.194**	0.235	0.247	**0.242**	0.305	**0.229**
	192	0.373	**0.354**	**0.369**	0.384	0.378	**0.367**	**0.363**	0.380	0.363	**0.282**	0.375	**0.365**
	336	**0.380**	0.402	**0.375**	0.405	**0.382**	0.430	**0.376**	0.749	**0.375**	0.737	**0.384**	0.732
	720	**0.381**	0.737	**0.379**	0.727	**0.382**	0.753	**0.715**	0.728	**0.406**	0.728	**0.390**	0.731
Traffic	96	**0.807**	0.931	**0.908**	0.925	**0.815**	0.871	**0.820**	0.936	0.903	**0.870**	0.889	**0.831**
	192	**0.958**	1.383	**0.953**	0.971	**0.956**	0.973	**0.964**	0.971	**0.960**	0.968	**0.965**	0.967
	336	**0.980**	0.993	**0.984**	1.404	**0.974**	0.996	**0.982**	1.394	**0.979**	1.401	**0.985**	1.385
	720	**1.012**	1.428	**1.016**	1.431	**1.028**	1.431	**1.013**	1.431	**1.017**	1.432	**1.014**	1.436
ETTh1	96	**0.512**	0.516	**0.501**	0.502	0.519	**0.509**	0.515	**0.510**	0.527	**0.525**	0.510	**0.501**
	192	**0.531**	0.535	0.531	**0.523**	0.541	**0.522**	0.538	**0.533**	**0.552**	0.712	0.539	**0.513**
	336	**0.709**	0.711	**0.515**	0.697	0.539	**0.526**	**0.536**	0.708	0.704	**0.703**	**0.520**	0.690
	720	0.723	**0.717**	0.731	**0.712**	**0.725**	0.731	**0.556**	0.729	**0.728**	0.730	**0.525**	0.718
ETTm1	96	0.466	**0.441**	0.465	**0.447**	**0.464**	0.471	0.467	**0.465**	0.480	**0.457**	0.459	**0.444**
	192	0.488	**0.485**	0.518	**0.487**	0.495	**0.486**	**0.471**	0.523	**0.489**	0.494	**0.496**	0.499
	336	0.509	**0.493**	0.534	**0.528**	**0.508**	0.515	**0.502**	0.700	**0.511**	0.519	0.500	**0.488**
	720	**0.539**	0.715	**0.545**	0.713	0.707	**0.573**	**0.711**	0.717	**0.538**	0.710	**0.544**	0.713
Wind	96	**0.900**	0.916	0.936	**0.914**	**0.900**	0.908	0.940	**0.892**	0.942	**0.912**	**0.915**	0.924
	192	1.095	**1.083**	1.101	**1.097**	1.163	**1.064**	1.101	**1.063**	1.136	**1.076**	**1.087**	1.143
	336	1.259	**1.240**	**1.232**	1.243	1.277	**1.228**	1.269	**1.213**	1.248	**1.224**	1.274	**1.252**
	720	**1.323**	1.327	**1.326**	1.415	**1.289**	1.386	**1.349**	1.362	**1.335**	1.353	**1.375**	1.418
ETTm2	96	**0.241**	0.245	**0.238**	0.249	0.271	**0.232**	**0.250**	0.254	0.238	**0.237**	**0.254**	0.260
	192	0.300	**0.300**	0.313	**0.295**	**0.301**	0.306	0.306	**0.286**	0.311	**0.291**	**0.285**	0.302
	336	**0.348**	0.352	**0.333**	0.344	0.360	**0.348**	0.342	**0.340**	**0.339**	0.342	**0.328**	0.349
	720	**0.418**	0.422	**0.418**	0.428	**0.416**	0.419	0.422	**0.421**	**0.421**	0.422	0.424	**0.423**

Table 8, continued.

Dataset	H	2021		2022		2023		2024		2025		2026	
		OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM
Exchange	96	0.260	**0.165**	**0.178**	0.242	0.326	**0.197**	**0.259**	0.270	0.211	**0.179**	0.252	**0.249**
	192	**0.420**	0.530	**0.432**	0.537	**0.504**	0.537	**0.418**	0.444	0.524	**0.345**	**0.346**	0.356
	336	0.832	**0.731**	**0.623**	0.675	0.848	**0.732**	0.744	**0.606**	0.827	**0.472**	**0.594**	0.714
	720	**1.312**	1.313	3.437	**1.314**	2.546	**1.314**	**1.316**	1.317	2.111	**1.321**	**1.297**	1.307
ETTh2	96	**0.370**	0.370	0.362	**0.357**	0.369	**0.358**	0.396	**0.385**	0.382	**0.355**	**0.355**	0.357
	192	0.393	**0.384**	**0.386**	0.388	**0.391**	0.393	0.380	**0.375**	0.418	**0.378**	0.394	**0.374**
	336	0.391	**0.381**	0.401	**0.383**	0.401	**0.388**	0.387	**0.380**	0.395	**0.383**	0.390	**0.385**
	720	0.436	**0.425**	0.426	**0.423**	0.432	**0.432**	0.426	**0.426**	0.446	**0.433**	**0.425**	0.426
Weather	96	0.195	**0.192**	0.200	**0.187**	0.194	**0.189**	0.192	**0.186**	0.203	**0.194**	0.197	**0.194**
	192	**0.233**	0.276	0.279	**0.233**	0.245	**0.233**	0.234	**0.234**	0.237	**0.232**	0.235	**0.227**
	336	**0.309**	0.320	0.308	**0.283**	**0.279**	0.286	**0.283**	0.306	**0.274**	0.313	0.307	**0.277**
	720	0.354	**0.346**	0.354	**0.344**	**0.347**	0.367	**0.350**	0.367	0.352	**0.348**	0.353	**0.346**

Table 9: Normalized MAE for configuration 2, lower is better. Rows are (dataset, forecast horizon H) pairs, columns are the six random seeds, each split into the observer head (OSSM) and the standard head (SSM). The better of the two is in bold. "div." marks a run that diverged, i.e. reached a normalized MSE above 10 or produced no metric at all; divergence counts as a loss, so the head that converged is shown in bold.

Dataset	H	2021		2022		2023		2024		2025		2026	
		OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM
Short-term forecasting													
NN5	24	**0.833**	0.915	**0.833**	0.927	**0.860**	0.889	**0.807**	0.889	**0.820**	0.945	**0.863**	0.900
	36	**0.862**	0.924	**0.869**	0.925	**0.886**	0.909	**0.809**	0.887	**0.824**	0.934	**0.891**	0.917
	48	**0.876**	0.925	**0.889**	0.925	**0.892**	0.907	**0.806**	0.886	**0.822**	0.932	**0.897**	0.912
	60	**0.887**	0.924	**0.892**	0.926	**0.896**	0.916	**0.805**	0.884	**0.823**	0.932	**0.906**	0.919
NYSE	24	0.360	**0.359**	0.456	**0.383**	**0.394**	0.400	**0.411**	0.415	0.667	**0.462**	0.465	**0.400**
	36	0.654	**0.473**	0.907	**0.758**	**0.484**	0.561	0.591	**0.536**	**0.528**	0.613	0.595	**0.444**
	48	0.849	**0.695**	1.019	**0.742**	0.691	**0.572**	**0.688**	0.791	**0.638**	0.706	0.964	**0.588**
	60	0.936	**0.852**	1.124	**0.932**	1.021	**0.713**	**0.878**	1.043	1.078	**0.869**	1.073	**0.706**
ILI	24	**1.207**	1.257	1.074	**1.072**	**1.105**	1.150	**1.106**	1.106	1.116	**1.054**	**1.076**	1.163
	36	1.219	**1.187**	**1.073**	1.126	**1.121**	1.154	**1.078**	1.103	1.091	**1.076**	**1.082**	1.227
	48	1.260	**1.174**	**1.096**	1.153	**1.094**	1.142	**1.064**	1.125	1.086	**1.083**	**1.056**	1.192
	60	1.338	**1.185**	**1.105**	1.194	**1.106**	1.184	**1.088**	1.143	**1.088**	1.160	**1.079**	1.314
NASDAQ	24	0.543	**0.514**	0.557	**0.505**	0.541	**0.500**	0.517	**0.481**	**0.524**	0.526	0.532	**0.508**
	36	0.676	**0.639**	0.678	**0.641**	0.664	**0.643**	0.661	**0.614**	**0.653**	0.658	0.661	**0.637**
	48	0.820	**0.745**	0.753	**0.743**	0.766	**0.756**	0.740	**0.728**	**0.746**	0.753	0.756	**0.745**
	60	0.817	**0.793**	0.797	**0.775**	0.807	**0.799**	0.819	**0.780**	0.837	**0.792**	**0.790**	0.810
Long-term forecasting													
Electricity	96	**0.366**	0.371	**0.366**	0.429	**0.368**	0.372	**0.366**	0.371	**0.367**	0.370	**0.366**	0.426
	192	**0.431**	0.437	**0.372**	0.436	**0.376**	0.435	**0.432**	0.436	**0.431**	0.437	**0.372**	0.435
	336	**0.440**	0.759	**0.439**	0.447	**0.438**	0.443	**0.439**	0.751	**0.439**	0.761	**0.439**	0.754
	720	**0.457**	0.775	**0.455**	0.774	**0.455**	0.761	**0.457**	0.770	**0.758**	0.771	**0.456**	0.738
Solar	96	**0.301**	0.316	0.367	**0.339**	0.367	**0.329**	**0.298**	0.325	0.360	**0.333**	0.413	**0.321**
	192	0.490	**0.464**	**0.479**	0.489	0.482	**0.474**	**0.475**	0.484	0.468	**0.360**	**0.469**	0.469
	336	**0.489**	0.501	**0.482**	0.501	**0.490**	0.513	**0.488**	0.687	**0.480**	0.677	**0.493**	0.675
	720	**0.478**	0.684	**0.482**	0.680	**0.481**	0.693	**0.671**	0.679	**0.509**	0.683	**0.493**	0.683
Traffic	96	**0.477**	0.530	**0.529**	0.529	**0.489**	0.507	**0.486**	0.535	0.524	**0.505**	0.529	**0.489**
	192	**0.537**	0.792	**0.538**	0.544	**0.537**	0.546	**0.541**	0.545	**0.538**	0.542	**0.540**	0.542
	336	**0.542**	0.549	**0.544**	0.798	**0.541**	0.551	**0.543**	0.793	**0.542**	0.796	**0.546**	0.787
	720	**0.556**	0.803	**0.555**	0.805	**0.557**	0.805	**0.554**	0.805	**0.556**	0.805	**0.554**	0.807

Table 9, continued.

Dataset	H	2021		2022		2023		2024		2025		2026	
		OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM
ETTh1	96	**0.499**	0.507	**0.492**	0.493	**0.500**	0.501	0.505	**0.497**	0.512	**0.507**	0.500	**0.496**
	192	**0.511**	0.515	0.508	**0.504**	0.522	**0.510**	0.515	**0.514**	**0.528**	0.578	0.518	**0.504**
	336	**0.593**	0.593	**0.500**	0.580	0.518	**0.515**	**0.516**	0.594	**0.587**	0.588	**0.506**	0.579
	720	0.619	**0.614**	0.623	**0.605**	**0.621**	0.623	**0.533**	0.623	**0.619**	0.619	**0.521**	0.613
ETTm1	96	0.464	**0.456**	0.469	**0.459**	**0.465**	0.472	0.468	**0.465**	0.473	**0.465**	0.466	**0.455**
	192	**0.479**	0.480	0.496	**0.475**	0.484	**0.479**	**0.471**	0.499	**0.484**	0.485	**0.476**	0.487
	336	0.484	**0.482**	0.509	**0.492**	**0.484**	0.492	**0.486**	0.565	0.491	**0.491**	**0.478**	0.478
	720	**0.503**	0.569	**0.507**	0.563	0.566	**0.514**	**0.567**	0.571	**0.506**	0.567	**0.508**	0.567
Wind	96	**0.670**	0.678	0.675	**0.667**	0.670	**0.666**	0.694	**0.658**	0.690	**0.670**	**0.679**	0.680
	192	0.762	**0.755**	0.769	**0.763**	0.788	**0.744**	0.769	**0.744**	0.783	**0.755**	**0.755**	0.788
	336	0.833	**0.826**	**0.820**	0.825	0.842	**0.821**	0.834	**0.818**	0.834	**0.824**	0.838	**0.829**
	720	0.864	**0.863**	**0.860**	0.890	**0.851**	0.879	**0.865**	0.870	0.871	**0.867**	**0.877**	0.895
ETTm2	96	**0.327**	0.329	**0.324**	0.331	0.342	**0.323**	**0.330**	0.332	**0.325**	0.326	**0.336**	0.336
	192	**0.359**	0.360	0.366	**0.357**	**0.360**	0.361	0.361	**0.352**	0.364	**0.357**	**0.352**	0.361
	336	0.388	**0.388**	**0.379**	0.382	0.392	**0.386**	0.384	**0.383**	**0.383**	0.384	**0.376**	0.386
	720	0.429	**0.428**	**0.427**	0.434	**0.428**	0.428	0.430	**0.428**	**0.429**	0.429	0.431	**0.429**
Exchange	96	0.371	**0.316**	**0.316**	0.357	0.417	**0.350**	**0.367**	0.382	0.340	**0.321**	**0.357**	0.360
	192	**0.483**	0.550	**0.495**	0.559	**0.529**	0.558	**0.482**	0.493	0.530	**0.437**	**0.415**	0.435
	336	**0.662**	0.666	**0.596**	0.630	**0.661**	0.666	0.648	**0.581**	0.664	**0.515**	**0.568**	0.656
	720	**0.860**	0.861	1.297	**0.863**	1.157	**0.862**	**0.860**	0.861	1.073	**0.864**	**0.854**	0.860
ETTh2	96	**0.418**	0.419	0.415	**0.411**	0.421	**0.415**	**0.416**	0.427	0.427	**0.411**	**0.411**	0.412
	192	0.432	**0.424**	**0.424**	0.427	**0.429**	0.434	**0.419**	0.420	0.448	**0.423**	0.427	**0.418**
	336	0.431	**0.425**	0.434	**0.425**	0.439	**0.430**	0.429	**0.424**	0.434	**0.427**	0.431	**0.427**
	720	0.464	**0.455**	0.454	**0.452**	0.459	**0.459**	**0.456**	0.456	0.468	**0.460**	**0.455**	0.456
Weather	96	0.259	**0.253**	0.254	**0.247**	0.253	**0.248**	0.255	**0.245**	0.265	**0.255**	0.259	**0.254**
	192	**0.285**	0.315	0.317	**0.285**	0.294	**0.283**	0.287	**0.282**	0.290	**0.284**	0.283	**0.279**
	336	**0.338**	0.348	0.333	**0.319**	**0.317**	0.323	**0.318**	0.337	**0.311**	0.339	0.333	**0.313**
	720	0.367	**0.361**	0.365	**0.360**	**0.365**	0.380	**0.365**	0.377	**0.363**	0.365	0.364	**0.361**

E.3 CONFIGURATION 3

Table 10: Hyper-parameters of configuration 3. Every other setting follows the TFB defaults, and the two heads share all of them, including the random seeds.

Hyper-parameter	Value
Encoder layers e_{layers}	2
State dimension d_x	128
Model dimension d_{model}	128
Learning rate	10^{-4}
Batch size	32

Table 11: Normalized MSE for configuration 3, lower is better. Rows are (dataset, forecast horizon H) pairs, columns are the six random seeds, each split into the observer head (OSSM) and the standard head (SSM). The better of the two is in bold. "div." marks a run that diverged, i.e. reached a normalized MSE above 10 or produced no metric at all; divergence counts as a loss, so the head that converged is shown in bold.

Dataset	H	2021		2022		2023		2024		2025		2026	
		OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM
Short-term forecasting													
NN5	24	**1.020**	1.246	**0.993**	1.150	**0.996**	1.324	**1.015**	1.314	**1.022**	1.225	**1.058**	1.272
	36	**1.014**	1.233	**0.980**	1.154	**0.969**	1.310	**0.997**	1.302	**1.004**	1.203	**1.047**	1.237
	48	**0.986**	1.227	**0.973**	1.132	**0.964**	1.297	**0.984**	1.292	**0.997**	1.183	**1.035**	1.226
	60	**0.977**	1.222	**0.967**	1.138	**0.953**	1.294	**0.983**	1.288	**0.984**	1.178	**1.022**	1.235
NYSE	24	**0.265**	0.305	**0.236**	0.261	0.246	**0.244**	**0.261**	0.300	**0.250**	0.262	**0.258**	0.258
	36	**0.352**	0.442	**0.416**	0.445	**0.536**	0.545	**0.443**	0.514	**0.474**	0.477	**0.404**	0.647
	48	**0.566**	0.602	**1.036**	1.220	1.004	**0.779**	**0.760**	1.366	0.732	**0.633**	0.992	**0.653**
	60	**0.721**	1.033	**0.900**	1.361	0.843	**0.741**	**0.738**	0.879	**0.902**	0.944	**0.892**	1.859
ILI	24	2.583	**2.409**	2.587	**2.251**	**2.460**	2.850	2.564	**2.356**	2.530	**2.213**	2.576	**2.399**
	36	2.632	**2.579**	**2.554**	2.564	**2.277**	2.801	2.434	**2.370**	**2.347**	2.459	**2.279**	2.555
	48	**2.461**	2.585	**2.519**	2.631	**2.390**	2.709	**2.379**	2.642	**2.329**	2.632	**2.360**	2.649
	60	**2.381**	2.694	**2.639**	2.774	**2.421**	2.847	**2.458**	2.782	**2.564**	2.749	4.054	**2.729**
NASDAQ	24	0.524	**0.500**	0.539	**0.470**	**0.475**	0.482	0.454	**0.414**	0.524	**0.486**	0.487	**0.476**
	36	0.836	**0.819**	0.834	**0.730**	**0.711**	0.725	0.712	**0.707**	0.817	**0.770**	0.797	**0.777**
	48	1.157	**1.059**	1.151	**1.036**	1.068	**1.028**	1.055	**1.010**	1.079	**1.058**	1.092	**1.072**
	60	1.238	**1.234**	1.302	**1.255**	**1.226**	1.260	**1.170**	1.224	1.193	**1.189**	**1.227**	1.315
Long-term forecasting													
Electricity	96	**0.258**	0.276	**0.268**	0.276	**0.270**	0.276	**0.273**	0.276	**0.271**	0.276	**0.271**	0.275
	192	**0.280**	0.355	**0.280**	0.356	**0.280**	0.357	**0.282**	0.357	**0.280**	0.359	**0.350**	0.360
	336	**0.290**	0.840	**0.290**	0.852	**0.362**	0.862	**0.362**	0.864	**0.361**	0.866	**0.362**	0.858
	720	**0.874**	0.889	**0.388**	0.889	**0.869**	0.888	**0.862**	0.867	**0.888**	0.891	**0.873**	0.879
Solar	96	**0.204**	0.235	**0.234**	0.258	**0.212**	0.260	0.236	**0.222**	**0.209**	0.231	0.239	**0.220**
	192	**0.365**	0.405	**0.269**	0.388	**0.374**	0.395	**0.349**	0.398	**0.360**	0.396	**0.390**	0.398
	336	**0.369**	0.735	**0.382**	0.779	**0.378**	0.754	**0.371**	0.751	**0.374**	0.469	**0.377**	0.721
	720	**0.718**	0.743	0.779	**0.744**	**0.416**	0.746	0.730	**0.728**	**0.400**	0.741	**0.402**	0.743
Traffic	96	**0.786**	0.936	**0.779**	0.830	**0.810**	0.837	**0.767**	0.856	**0.814**	0.943	**0.822**	0.833
	192	**0.951**	0.978	**0.850**	0.970	**0.954**	0.973	**0.851**	0.975	**0.946**	0.966	**0.960**	0.973
	336	**0.975**	1.415	**0.970**	1.412	**0.987**	1.364	**0.982**	1.398	**0.973**	1.415	**0.988**	1.404
	720	**1.416**	1.434	**1.009**	1.423	1.422	**1.421**	**1.411**	1.423	**1.418**	1.420	1.416	**1.046**
ETTh1	96	**0.502**	0.511	**0.430**	0.505	0.512	**0.508**	**0.504**	0.527	**0.520**	0.542	**0.446**	0.501
	192	0.527	**0.524**	**0.538**	0.700	0.525	**0.524**	**0.518**	0.697	0.541	**0.532**	**0.629**	0.711
	336	**0.516**	0.704	**0.674**	0.704	**0.711**	0.714	**0.584**	0.709	**0.538**	0.710	**0.536**	0.598
	720	**0.722**	0.723	0.704	**0.701**	0.729	**0.720**	**0.728**	0.730	**0.712**	0.730	0.722	**0.707**
ETTm1	96	0.460	**0.446**	0.457	**0.387**	0.448	**0.438**	0.469	**0.461**	0.459	**0.389**	0.456	**0.445**
	192	0.481	**0.480**	0.488	**0.488**	0.496	**0.468**	0.492	**0.489**	**0.483**	0.490	0.507	**0.466**
	336	**0.497**	0.692	**0.515**	0.526	**0.518**	0.688	**0.507**	0.524	0.519	**0.510**	0.545	**0.508**
	720	0.739	**0.711**	**0.539**	0.712	**0.706**	0.714	**0.706**	0.714	**0.544**	0.712	**0.539**	0.711
Wind	96	0.900	**0.884**	0.938	**0.882**	**0.901**	0.908	0.916	**0.890**	0.932	**0.925**	0.917	**0.914**
	192	**1.094**	1.105	1.124	**1.098**	1.104	**1.064**	1.113	**1.099**	**1.073**	1.081	1.135	**1.072**
	336	1.296	**1.226**	**1.218**	1.229	1.277	**1.254**	**1.230**	1.249	**1.246**	1.258	**1.238**	1.256
	720	**1.305**	1.350	**1.325**	1.412	1.342	**1.341**	**1.342**	1.416	**1.347**	1.369	1.400	**1.362**
ETTm2	96	0.251	**0.227**	0.248	**0.228**	0.256	**0.240**	**0.233**	0.255	**0.226**	0.229	**0.238**	0.245
	192	**0.298**	0.315	**0.287**	0.296	0.300	**0.297**	0.304	**0.292**	0.302	**0.295**	**0.296**	0.302
	336	**0.338**	0.352	**0.347**	0.351	**0.342**	0.350	**0.341**	0.350	0.348	**0.340**	0.345	**0.327**
	720	**0.419**	0.421	0.423	**0.418**	0.422	**0.420**	0.421	**0.421**	**0.419**	0.422	**0.423**	0.423

Table 11, continued.

Dataset	H	2021		2022		2023		2024		2025		2026	
		OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM
Exchange	96	**0.199**	0.246	0.238	**0.187**	0.280	**0.199**	**0.172**	0.258	**0.227**	0.259	0.253	**0.212**
	192	**0.398**	0.537	**0.391**	0.542	**0.433**	0.542	**0.389**	0.532	**0.433**	0.479	**0.445**	0.538
	336	**0.700**	0.723	**0.689**	0.738	**0.707**	0.727	**0.571**	0.724	**0.635**	0.733	0.711	**0.662**
	720	1.313	**1.308**	1.314	**1.310**	**1.303**	1.308	1.317	**1.314**	**1.320**	1.322	**1.305**	1.307
ETTh2	96	**0.353**	0.357	0.362	**0.354**	0.353	**0.352**	**0.354**	0.370	0.382	**0.373**	**0.357**	0.362
	192	**0.381**	0.390	0.390	**0.383**	**0.381**	0.398	0.401	**0.373**	0.401	**0.381**	0.378	**0.377**
	336	**0.383**	0.418	**0.379**	0.383	0.387	**0.387**	**0.381**	0.386	**0.385**	0.388	0.385	**0.380**
	720	**0.424**	0.425	0.421	**0.421**	0.432	**0.431**	0.426	**0.425**	0.433	**0.431**	0.432	**0.424**
Weather	96	0.193	**0.187**	0.190	**0.181**	0.202	**0.190**	0.192	**0.188**	0.195	**0.193**	0.196	**0.190**
	192	0.243	**0.228**	**0.225**	0.235	**0.237**	0.283	0.237	**0.228**	0.237	**0.233**	**0.236**	0.237
	336	0.274	**0.274**	0.281	**0.279**	**0.279**	0.321	**0.276**	0.280	**0.277**	0.284	**0.278**	0.281
	720	**0.347**	0.366	0.367	**0.361**	0.352	**0.349**	**0.366**	0.368	**0.347**	0.367	**0.367**	0.382

Table 12: Normalized MAE for configuration 3, lower is better. Rows are (dataset, forecast horizon H) pairs, columns are the six random seeds, each split into the observer head (OSSM) and the standard head (SSM). The better of the two is in bold. "div." marks a run that diverged, i.e. reached a normalized MSE above 10 or produced no metric at all; divergence counts as a loss, so the head that converged is shown in bold.

Dataset	H	2021		2022		2023		2024		2025		2026	
		OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM
Short-term forecasting													
NN5	24	**0.759**	0.872	**0.749**	0.827	**0.747**	0.924	**0.759**	0.918	**0.760**	0.856	**0.786**	0.881
	36	**0.767**	0.877	**0.753**	0.836	**0.743**	0.928	**0.760**	0.924	**0.763**	0.855	**0.791**	0.875
	48	**0.758**	0.879	**0.754**	0.830	**0.746**	0.926	**0.758**	0.924	**0.765**	0.850	**0.789**	0.874
	60	**0.755**	0.879	**0.753**	0.834	**0.743**	0.928	**0.760**	0.925	**0.762**	0.850	**0.784**	0.882
NYSE	24	**0.355**	0.402	**0.337**	0.364	0.340	**0.337**	**0.373**	0.388	**0.356**	0.358	**0.348**	0.359
	36	**0.394**	0.497	**0.457**	0.480	0.526	**0.522**	**0.477**	0.526	**0.507**	0.511	**0.436**	0.599
	48	**0.529**	0.594	**0.743**	0.834	0.782	**0.633**	**0.637**	0.911	0.624	**0.581**	0.732	**0.562**
	60	**0.629**	0.816	**0.713**	0.889	0.680	**0.634**	**0.669**	0.709	**0.710**	0.737	**0.702**	1.081
ILI	24	1.081	**1.054**	1.115	**1.033**	**1.064**	1.162	1.091	**1.067**	1.064	**1.028**	1.098	**1.073**
	36	**1.089**	1.096	**1.074**	1.103	**1.006**	1.160	**1.048**	1.059	**1.021**	1.074	**1.016**	1.096
	48	**1.049**	1.101	**1.060**	1.112	**1.049**	1.134	**1.038**	1.108	**1.025**	1.104	**1.040**	1.103
	60	**1.053**	1.127	**1.092**	1.139	**1.040**	1.157	**1.055**	1.139	**1.069**	1.129	1.448	**1.118**
NASDAQ	24	0.519	**0.518**	0.531	**0.501**	**0.490**	0.502	0.483	**0.475**	0.524	**0.504**	0.499	**0.498**
	36	**0.660**	0.664	0.671	**0.627**	**0.610**	0.621	**0.616**	0.623	0.669	**0.635**	0.648	**0.638**
	48	0.771	**0.745**	0.767	**0.732**	0.744	**0.728**	0.737	**0.733**	0.755	**0.745**	0.751	**0.742**
	60	**0.800**	0.803	0.813	**0.801**	**0.795**	0.803	**0.786**	0.802	0.791	**0.783**	**0.799**	0.807
Long-term forecasting													
Electricity	96	**0.351**	0.370	**0.364**	0.369	**0.364**	0.369	**0.367**	0.369	**0.365**	0.370	**0.366**	0.369
	192	**0.371**	0.436	**0.372**	0.437	**0.372**	0.438	**0.374**	0.438	**0.372**	0.440	**0.432**	0.440
	336	**0.380**	0.753	**0.380**	0.759	**0.439**	0.765	**0.439**	0.766	**0.439**	0.767	**0.440**	0.763
	720	**0.765**	0.774	**0.455**	0.775	**0.763**	0.774	**0.758**	0.761	**0.774**	0.775	**0.765**	0.769
Solar	96	**0.309**	0.322	**0.333**	0.341	**0.322**	0.339	0.326	**0.317**	**0.320**	0.323	0.348	**0.316**
	192	**0.471**	0.501	**0.371**	0.492	**0.477**	0.494	**0.458**	0.488	**0.464**	0.496	0.500	**0.497**
	336	**0.477**	0.678	**0.493**	0.698	**0.486**	0.690	**0.483**	0.688	**0.477**	0.524	**0.485**	0.674
	720	**0.674**	0.688	0.706	**0.689**	**0.506**	0.690	0.681	**0.681**	**0.503**	0.688	**0.506**	0.688
Traffic	96	**0.464**	0.533	**0.444**	0.489	**0.483**	0.492	**0.446**	0.496	**0.481**	0.537	**0.483**	0.488
	192	**0.537**	0.546	**0.488**	0.545	**0.536**	0.546	**0.484**	0.547	**0.534**	0.542	**0.541**	0.546
	336	**0.542**	0.803	**0.542**	0.802	**0.547**	0.773	**0.543**	0.794	**0.544**	0.802	**0.545**	0.798
	720	**0.796**	0.806	**0.552**	0.801	0.800	**0.799**	**0.795**	0.801	**0.798**	0.799	0.797	**0.571**

Table 12, continued.

Dataset	H	2021		2022		2023		2024		2025		2026	
		OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM
ETTh1	96	**0.488**	0.501	**0.457**	0.496	0.501	**0.498**	**0.494**	0.507	**0.496**	0.512	**0.463**	0.493
	192	**0.506**	0.511	**0.500**	0.571	**0.509**	0.512	**0.507**	0.580	0.516	**0.513**	**0.543**	0.582
	336	**0.506**	0.589	**0.568**	0.583	**0.594**	0.596	**0.552**	0.594	**0.518**	0.591	**0.517**	0.544
	720	**0.618**	0.619	0.601	**0.594**	0.623	**0.614**	**0.622**	0.623	**0.608**	0.619	0.615	**0.603**
ETTm1	96	0.466	**0.458**	0.462	**0.423**	0.455	**0.452**	0.469	**0.459**	0.462	**0.423**	0.461	**0.455**
	192	**0.472**	0.479	**0.478**	0.480	0.483	**0.470**	0.481	**0.478**	**0.474**	0.481	0.492	**0.469**
	336	**0.476**	0.551	**0.491**	0.492	**0.493**	0.551	**0.488**	0.490	0.488	**0.487**	0.511	**0.485**
	720	0.588	**0.568**	**0.505**	0.566	**0.564**	0.570	**0.564**	0.568	**0.508**	0.568	**0.505**	0.566
Wind	96	0.668	**0.657**	0.687	**0.656**	0.667	**0.667**	0.672	**0.654**	0.681	**0.673**	0.681	**0.665**
	192	**0.756**	0.767	0.775	**0.767**	0.772	**0.745**	0.771	**0.765**	**0.751**	0.756	0.780	**0.748**
	336	0.844	**0.816**	**0.813**	0.817	0.845	**0.832**	**0.818**	0.830	**0.825**	0.833	**0.830**	0.831
	720	**0.856**	0.867	**0.856**	0.889	**0.863**	0.864	**0.868**	0.891	**0.866**	0.873	0.884	**0.871**
ETTm2	96	0.331	**0.317**	0.329	**0.320**	0.334	**0.327**	**0.320**	0.333	**0.317**	0.321	**0.325**	0.328
	192	**0.357**	0.368	**0.352**	0.358	**0.358**	0.360	0.360	**0.357**	0.361	**0.357**	**0.359**	0.363
	336	**0.384**	0.387	**0.385**	0.386	**0.384**	0.386	**0.384**	0.386	0.385	**0.383**	0.387	**0.375**
	720	**0.427**	0.428	0.430	**0.428**	0.429	**0.429**	0.428	**0.428**	**0.428**	0.429	**0.429**	0.429
Exchange	96	**0.336**	0.357	0.353	**0.324**	0.383	**0.332**	**0.319**	0.377	**0.352**	0.364	0.362	**0.333**
	192	**0.470**	0.558	**0.457**	0.563	**0.486**	0.563	**0.468**	0.554	**0.485**	0.521	**0.496**	0.561
	336	**0.645**	0.660	**0.637**	0.671	**0.648**	0.663	**0.572**	0.661	**0.596**	0.667	0.652	**0.626**
	720	0.861	**0.859**	0.863	**0.862**	**0.856**	0.860	0.861	**0.859**	**0.863**	0.864	**0.859**	0.860
ETTh2	96	**0.409**	0.412	0.413	**0.409**	0.414	**0.409**	**0.413**	0.415	0.430	**0.420**	**0.413**	0.418
	192	**0.424**	0.429	**0.427**	0.428	**0.425**	0.436	0.430	**0.419**	0.435	**0.422**	0.421	**0.420**
	336	**0.426**	0.450	**0.423**	0.424	0.429	**0.429**	**0.425**	0.428	**0.429**	0.430	0.427	**0.424**
	720	**0.454**	0.455	0.451	**0.450**	0.459	**0.458**	0.455	**0.455**	0.460	**0.458**	0.459	**0.454**
Weather	96	0.253	**0.246**	0.248	**0.239**	0.260	**0.249**	0.252	**0.248**	0.255	**0.251**	0.256	**0.248**
	192	0.297	**0.279**	**0.280**	0.281	**0.288**	0.321	0.289	**0.281**	0.290	**0.280**	0.287	**0.285**
	336	0.312	**0.311**	**0.312**	0.315	**0.314**	0.345	**0.311**	0.317	**0.313**	0.320	**0.312**	0.317
	720	**0.363**	0.377	0.378	**0.371**	**0.363**	0.364	**0.377**	0.382	**0.362**	0.379	**0.378**	0.385

E.4 CONFIGURATION 4

Table 13: Hyper-parameters of configuration 4. Every other setting follows the TFB defaults, and the two heads share all of them, including the random seeds.

Hyper-parameter	Value
Encoder layers e_{layers}	2
State dimension d_x	128
Model dimension d_{model}	128
Learning rate	10^{-3}
Batch size	32

Table 14: Normalized MSE for configuration 4, lower is better. Rows are (dataset, forecast horizon H) pairs, columns are the six random seeds, each split into the observer head (OSSM) and the standard head (SSM). The better of the two is in bold. "div." marks a run that diverged, i.e. reached a normalized MSE above 10 or produced no metric at all; divergence counts as a loss, so the head that converged is shown in bold.

Dataset	H	2021		2022		2023		2024		2025		2026	
		OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM
Short-term forecasting													
NN5	24	**0.800**	1.071	**0.759**	0.854	0.836	**0.823**	**0.810**	1.316	**0.812**	1.198	**0.807**	0.862
	36	**0.843**	0.947	**0.771**	1.165	**1.107**	1.300	**0.819**	0.823	**0.850**	1.300	**0.757**	1.193
	48	**1.090**	1.299	**0.795**	1.306	**0.734**	1.241	**1.190**	1.282	**0.757**	0.842	**1.107**	1.299
	60	1.265	**1.179**	**0.720**	1.290	**1.264**	1.290	**0.764**	1.291	1.274	**1.267**	**0.905**	1.226
NYSE	24	**0.282**	0.591	0.503	**0.227**	**0.228**	0.283	0.275	**0.275**	**0.209**	0.243	0.354	**0.294**
	36	**0.397**	1.471	1.365	**0.906**	**1.444**	1.622	**0.367**	1.338	**0.436**	1.459	**0.438**	1.479
	48	**0.543**	1.797	**0.560**	1.725	**1.738**	1.761	**1.670**	1.726	**1.435**	1.735	**1.519**	1.772
	60	**1.879**	2.078	**2.014**	2.035	2.057	**2.054**	**1.970**	2.011	2.065	**2.041**	2.034	**1.759**
ILI	24	2.696	**1.900**	**2.285**	2.509	**2.288**	2.311	**2.322**	2.533	2.203	**2.096**	2.359	**1.972**
	36	**2.453**	2.638	**2.317**	2.526	**2.673**	3.734	2.478	**2.397**	**2.229**	2.439	4.004	**2.936**
	48	**2.539**	3.647	3.562	**2.627**	**2.522**	4.132	**3.840**	3.908	**2.328**	2.430	4.132	**3.559**
	60	**2.584**	4.239	**2.615**	4.178	**2.476**	4.336	**3.615**	4.248	3.598	**2.686**	4.293	**3.727**
NASDAQ	24	**0.436**	1.253	0.475	**0.447**	**0.462**	0.491	**0.424**	0.465	1.360	**0.502**	**0.450**	0.461
	36	**0.646**	1.302	**0.748**	1.268	**0.785**	0.856	**0.725**	0.917	0.982	**0.812**	**0.778**	1.287
	48	**1.080**	1.198	**1.133**	1.257	**1.024**	1.084	**1.051**	1.164	1.185	**1.126**	**1.077**	1.181
	60	1.401	**1.090**	**1.106**	1.200	**1.171**	1.194	1.198	**1.076**	1.076	**1.059**	1.175	**1.082**
Long-term forecasting													
Electricity	96	**0.718**	0.849	**0.805**	0.854	**0.272**	0.276	**0.261**	0.857	**0.327**	0.842	**0.263**	0.848
	192	**0.845**	0.857	**0.865**	0.872	**0.349**	0.358	**0.739**	0.861	0.860	**0.855**	**0.349**	0.861
	336	**0.839**	0.866	div.	**0.868**	**0.361**	0.868	**0.851**	div.	**0.858**	0.869	**0.840**	div.
	720	div.	div.	**0.894**	div.	1.492	**0.892**	**0.890**	div.	**0.890**	0.892	**0.965**	div.
Solar	96	**0.250**	0.368	**0.197**	0.368	div.	**0.375**	div.	**0.277**	div.	**0.363**	0.597	**0.270**
	192	div.	**0.764**	**0.727**	0.779	**0.664**	0.770	**0.387**	0.397	div.	div.	**0.688**	0.746
	336	**0.760**	0.777	0.724	**0.501**	**0.745**	div.	**0.721**	0.765	div.	div.	div.	**0.763**
	720	div.	**0.758**	**0.746**	div.	**0.748**	0.749	0.747	**0.705**	**0.722**	0.744	div.	div.
Traffic	96	**0.845**	0.942	**0.763**	1.379	**0.750**	0.939	**0.754**	1.419	**0.749**	1.379	1.421	**0.944**
	192	1.428	**1.414**	**1.360**	1.440	**1.325**	1.391	**1.356**	1.618	**1.378**	1.418	1.350	**0.984**
	336	**1.386**	1.397	**1.234**	1.427	**1.383**	1.412	**1.335**	1.432	**1.417**	1.425	**1.395**	1.511
	720	**1.378**	1.460	**1.438**	div.	1.519	**1.440**	**1.435**	1.474	**1.422**	div.	**1.434**	div.
ETTh1	96	**0.448**	0.505	**0.453**	0.697	**0.428**	0.437	**0.467**	0.712	**0.497**	0.708	**0.424**	0.710
	192	0.853	**0.715**	**0.713**	0.713	**0.531**	0.722	0.709	**0.539**	**0.703**	0.719	0.715	**0.714**
	336	**0.522**	0.706	**0.640**	0.700	0.716	**0.712**	**0.711**	0.712	**0.704**	0.710	**0.526**	0.722
	720	**0.714**	div.	**0.733**	div.	**0.749**	div.	**0.749**	div.	**0.758**	div.	**0.708**	div.
ETTm1	96	0.624	**0.453**	**0.442**	0.463	0.630	**0.448**	**0.671**	0.687	**0.487**	0.682	0.462	**0.458**
	192	**0.669**	0.694	**0.489**	0.650	**0.471**	0.672	**0.497**	0.690	**0.504**	0.675	**0.484**	div.
	336	**0.538**	div.	**0.682**	div.	**0.554**	0.698	**0.504**	0.701	**0.698**	0.702	div.	div.
	720	**0.715**	div.	1.352	**0.715**	**0.717**	div.	**0.718**	2.201	**0.728**	div.	**0.712**	div.
Wind	96	**0.887**	0.894	div.	**0.954**	**0.899**	div.	**0.882**	1.243	div.	**1.240**	**0.877**	1.238
	192	div.	**1.314**	**1.083**	1.312	**1.107**	div.	**1.111**	1.249	div.	**1.281**	**1.096**	1.317
	336	**1.194**	div.	div.	div.	**1.378**	div.	**1.353**	div.	**1.238**	1.386	**1.258**	div.
	720	div.	div.	**1.419**	div.	1.487	**1.432**	**1.419**	div.	**1.417**	div.	div.	div.
ETTm2	96	**0.244**	div.	div.	**0.247**	div.	**0.291**	**0.244**	0.248	**0.242**	0.256	**0.243**	0.255
	192	**0.306**	0.319	**0.302**	div.	**0.307**	0.317	**0.294**	0.319	**0.318**	div.	div.	**0.321**
	336	div.	**0.351**	**0.351**	div.	0.352	**0.352**	**0.352**	0.356	0.357	**0.354**	**0.356**	div.
	720	**0.423**	0.423	**0.421**	div.	**0.426**	div.	**0.423**	div.	**0.423**	div.	**0.426**	div.

Table 14, continued.

Dataset	H	2021		2022		2023		2024		2025		2026	
		OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM
Exchange	96	**0.258**	0.430	0.433	**0.425**	**0.279**	0.423	**0.256**	0.423	**0.247**	0.451	**0.276**	0.421
	192	**0.547**	0.590	0.604	**0.547**	**0.591**	div.	0.585	**0.539**	**0.591**	0.603	**0.374**	div.
	336	**0.722**	0.723	0.763	**0.734**	0.735	**0.724**	**0.737**	div.	**0.730**	div.	**0.751**	div.
	720	div.	div.	**1.378**	div.	**1.360**	div.	div.	div.	div.	div.	**1.304**	div.
ETTh2	96	**0.349**	0.367	**0.371**	0.372	**0.355**	0.366	0.360	**0.358**	**0.352**	0.360	**0.357**	0.374
	192	0.385	**0.384**	0.384	**0.382**	**0.385**	0.388	**0.383**	0.384	**0.381**	0.384	0.383	**0.382**
	336	**0.383**	0.384	0.386	**0.384**	**0.395**	div.	**0.395**	div.	0.391	**0.388**	**0.386**	div.
	720	**0.441**	div.	**0.422**	div.	**0.432**	0.433	0.429	**0.425**	**0.432**	div.	**0.442**	div.
Weather	96	0.264	**0.183**	0.193	**0.192**	1.634	**0.194**	**0.174**	0.176	**0.179**	div.	div.	**0.200**
	192	div.	**0.297**	**0.231**	div.	**0.237**	0.298	**0.236**	0.298	0.290	**0.282**	**0.234**	0.298
	336	**0.321**	0.325	div.	**0.326**	**0.279**	0.324	**0.324**	0.327	0.325	**0.325**	**0.298**	0.325
	720	**0.343**	0.362	**0.368**	div.	0.370	**0.367**	div.	div.	**0.369**	div.	div.	**0.362**

Table 15: Normalized MAE for configuration 4, lower is better. Rows are (dataset, forecast horizon H) pairs, columns are the six random seeds, each split into the observer head (OSSM) and the standard head (SSM). The better of the two is in bold. "div." marks a run that diverged, i.e. reached a normalized MSE above 10 or produced no metric at all; divergence counts as a loss, so the head that converged is shown in bold.

Dataset	H	2021		2022		2023		2024		2025		2026	
		OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM
Short-term forecasting													
NN5	24	**0.628**	0.783	**0.606**	0.663	**0.635**	0.644	**0.629**	0.929	**0.624**	0.870	**0.627**	0.658
	36	**0.682**	0.722	**0.623**	0.859	**0.834**	0.926	**0.637**	0.649	**0.678**	0.927	**0.618**	0.879
	48	**0.824**	0.917	**0.657**	0.932	**0.605**	0.902	**0.873**	0.925	**0.616**	0.663	**0.829**	0.930
	60	0.920	**0.868**	**0.606**	0.930	**0.918**	0.928	**0.623**	0.929	0.923	**0.918**	**0.711**	0.899
NYSE	24	**0.379**	0.554	0.507	**0.325**	**0.331**	0.376	**0.361**	0.373	**0.307**	0.342	0.428	**0.388**
	36	**0.445**	0.968	0.923	**0.701**	**0.955**	0.993	**0.419**	0.913	**0.469**	0.962	**0.476**	0.964
	48	**0.517**	1.083	**0.528**	1.057	**1.065**	1.071	**1.043**	1.060	**0.924**	1.062	**0.963**	1.074
	60	**1.098**	1.176	**1.156**	1.159	1.170	**1.166**	**1.145**	1.157	1.172	**1.163**	1.164	**1.043**
ILI	24	1.079	**0.947**	**1.042**	1.101	**1.025**	1.033	**1.023**	1.081	1.058	**1.019**	1.045	**0.986**
	36	**1.047**	1.105	**1.034**	1.098	**1.091**	1.394	1.064	**1.061**	**1.021**	1.056	1.463	**1.231**
	48	**1.093**	1.366	1.359	**1.121**	**1.070**	1.468	**1.415**	1.430	1.048	**1.044**	1.481	**1.345**
	60	**1.085**	1.490	**1.089**	1.487	**1.062**	1.508	**1.356**	1.498	1.338	**1.110**	1.503	**1.386**
NASDAQ	24	**0.465**	0.869	**0.491**	0.492	**0.463**	0.520	**0.465**	0.494	0.898	**0.510**	**0.470**	0.490
	36	**0.588**	0.883	**0.631**	0.867	**0.651**	0.692	**0.610**	0.706	0.735	**0.647**	**0.634**	0.876
	48	**0.742**	0.850	**0.753**	0.784	**0.714**	0.752	**0.738**	0.829	0.846	**0.801**	**0.746**	0.841
	60	0.832	**0.815**	**0.788**	0.793	**0.780**	0.801	**0.787**	0.809	0.811	**0.798**	**0.779**	0.811
Long-term forecasting													
Electricity	96	**0.688**	0.761	**0.736**	0.763	**0.365**	0.370	**0.354**	0.763	**0.415**	0.757	**0.356**	0.761
	192	**0.758**	0.764	**0.766**	0.768	**0.431**	0.438	**0.698**	0.765	0.765	**0.763**	**0.432**	0.765
	336	**0.751**	0.767	div.	**0.768**	**0.439**	0.768	**0.758**	div.	**0.763**	0.768	**0.751**	div.
	720	div.	div.	**0.775**	div.	0.824	**0.776**	**0.775**	div.	**0.775**	0.775	**0.788**	div.
Solar	96	**0.334**	0.474	**0.299**	0.477	div.	**0.479**	div.	**0.365**	div.	**0.469**	0.610	**0.345**
	192	div.	**0.693**	**0.670**	0.706	**0.635**	0.696	**0.488**	0.493	div.	div.	**0.646**	0.684
	336	**0.691**	0.701	0.673	**0.532**	**0.683**	div.	**0.670**	0.697	div.	div.	div.	**0.700**
	720	div.	**0.696**	**0.690**	div.	**0.690**	0.691	0.688	**0.665**	**0.675**	0.687	div.	div.
Traffic	96	**0.512**	0.534	**0.440**	0.794	**0.436**	0.534	**0.442**	0.805	**0.441**	0.792	0.806	**0.537**
	192	0.808	**0.803**	**0.778**	0.809	**0.791**	0.796	**0.801**	0.836	**0.792**	0.805	0.774	**0.548**
	336	**0.790**	0.795	**0.789**	0.805	**0.786**	0.801	**0.762**	0.805	**0.803**	0.805	**0.794**	0.821
	720	**0.804**	0.809	**0.807**	div.	0.822	**0.807**	**0.806**	0.811	**0.800**	div.	**0.806**	div.

Table 15, continued.

Dataset	H	2021		2022		2023		2024		2025		2026	
		OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM
ETTh1	96	**0.464**	0.502	**0.464**	0.574	**0.452**	0.458	**0.473**	0.581	**0.490**	0.576	**0.454**	0.579
	192	0.609	**0.587**	0.583	**0.573**	**0.512**	0.588	0.583	**0.516**	**0.579**	0.582	0.582	**0.580**
	336	**0.510**	0.590	**0.555**	0.574	0.597	**0.592**	0.595	**0.594**	0.593	**0.588**	**0.511**	0.599
	720	**0.611**	div.	**0.624**	div.	**0.636**	div.	**0.634**	div.	**0.639**	div.	**0.604**	div.
ETTm1	96	0.520	**0.456**	**0.449**	0.463	0.522	**0.457**	**0.543**	0.554	**0.477**	0.547	**0.464**	0.464
	192	**0.542**	0.558	**0.476**	0.533	**0.474**	0.541	**0.487**	0.553	**0.481**	0.543	**0.480**	div.
	336	**0.497**	div.	**0.548**	div.	**0.502**	0.556	**0.484**	0.565	**0.554**	0.560	div.	div.
	720	**0.568**	div.	0.630	**0.565**	**0.569**	div.	**0.573**	0.921	**0.573**	div.	**0.567**	div.
Wind	96	**0.655**	0.665	div.	**0.702**	**0.664**	div.	**0.657**	0.828	div.	**0.831**	**0.650**	0.828
	192	div.	**0.854**	**0.754**	0.853	**0.771**	div.	**0.774**	0.829	div.	**0.840**	**0.763**	0.860
	336	**0.814**	div.	div.	div.	**0.877**	div.	**0.867**	div.	**0.823**	0.886	**0.830**	div.
	720	div.	div.	**0.894**	div.	0.904	**0.903**	**0.893**	div.	**0.895**	div.	div.	div.
ETTm2	96	**0.328**	div.	div.	**0.330**	div.	**0.356**	**0.324**	0.332	**0.326**	0.336	**0.326**	0.334
	192	**0.364**	0.370	**0.360**	div.	**0.363**	0.369	**0.356**	0.370	**0.369**	div.	div.	**0.372**
	336	div.	**0.387**	**0.386**	div.	**0.388**	0.388	**0.388**	0.390	0.391	**0.389**	**0.390**	div.
	720	0.430	**0.429**	**0.427**	div.	**0.433**	div.	**0.431**	div.	**0.431**	div.	**0.431**	div.
Exchange	96	**0.364**	0.494	0.499	**0.494**	**0.380**	0.493	**0.365**	0.491	**0.355**	0.510	**0.377**	0.491
	192	**0.563**	0.584	0.594	**0.567**	**0.587**	div.	0.581	**0.559**	**0.583**	0.592	**0.448**	div.
	336	**0.661**	0.662	0.682	**0.669**	0.668	**0.663**	**0.666**	div.	**0.666**	div.	**0.668**	div.
	720	div.	div.	**0.887**	div.	**0.880**	div.	div.	div.	div.	div.	**0.858**	div.
ETTh2	96	**0.410**	0.422	0.422	**0.414**	**0.417**	0.422	0.418	**0.413**	**0.412**	0.416	**0.413**	0.417
	192	0.429	**0.429**	0.427	**0.420**	**0.424**	0.433	0.429	**0.423**	0.428	**0.424**	**0.425**	0.425
	336	**0.425**	0.426	0.427	**0.423**	**0.436**	div.	**0.439**	div.	0.433	**0.429**	**0.426**	div.
	720	**0.464**	div.	**0.451**	div.	**0.459**	0.459	0.455	**0.454**	**0.459**	div.	**0.465**	div.
Weather	96	0.313	**0.241**	0.253	**0.252**	0.478	**0.255**	**0.234**	0.237	**0.237**	div.	div.	**0.261**
	192	div.	**0.341**	**0.282**	div.	**0.284**	0.339	**0.286**	0.338	0.328	**0.320**	**0.281**	0.339
	336	**0.347**	0.357	div.	**0.355**	**0.314**	0.354	**0.350**	0.355	**0.350**	0.352	**0.332**	0.352
	720	**0.359**	0.373	**0.380**	div.	0.383	**0.378**	div.	div.	**0.380**	div.	div.	**0.374**

E.5 CONFIGURATION 5

Table 16: Hyper-parameters of configuration 5. Every other setting follows the TFB defaults, and the two heads share all of them, including the random seeds.

Hyper-parameter	Value
Encoder layers e_{layers}	2
State dimension d_x	128
Model dimension d_{model}	128
Learning rate	10^{-3}
Batch size	64

Table 17: normalized MSE for configuration 5, lower is better. Rows are (dataset, forecast horizon H) pairs, columns are the six random seeds, each split into the observer head (OSSM) and the standard head (SSM). The better of the two is in bold. "div." marks a run that diverged, i.e. reached a normalized MSE above 10 or produced no metric at all; divergence counts as a loss, so the head that converged is shown in bold.

Dataset	H	2021		2022		2023		2024		2025		2026	
		OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM
Short-term forecasting													
NN5	24	**1.104**	1.123	**1.215**	1.345	**0.836**	0.925	1.127	**1.001**	**0.949**	1.299	**0.875**	1.263
	36	**1.156**	1.298	**1.157**	1.284	**0.863**	1.295	**0.896**	1.295	**1.161**	1.327	**0.780**	1.254
	48	**0.964**	1.279	**0.751**	1.301	**0.803**	1.281	**1.068**	1.288	**1.247**	1.315	**0.783**	1.289
	60	**1.262**	1.289	**1.211**	1.296	**1.115**	1.286	**1.133**	1.287	**1.281**	1.291	**0.766**	1.280
NYSE	24	**0.369**	1.116	**0.239**	1.210	**0.659**	1.181	**0.221**	0.317	**0.308**	1.064	**0.297**	1.178
	36	**1.474**	1.535	**1.249**	1.486	**1.470**	1.481	**0.466**	1.362	**1.363**	1.473	**1.226**	1.474
	48	**1.626**	1.804	**1.698**	1.709	**0.767**	1.758	**0.797**	1.741	**0.591**	1.744	**1.472**	1.752
	60	**2.002**	2.071	**1.957**	2.022	2.058	**2.053**	2.024	**2.020**	**1.960**	2.034	**1.951**	2.048
ILI	24	2.719	**2.115**	2.281	**2.268**	**2.218**	2.563	**2.326**	2.401	**2.226**	3.887	2.290	**2.219**
	36	**2.629**	3.648	**2.504**	3.560	**2.323**	2.993	**2.514**	2.774	**2.283**	3.701	4.060	**2.441**
	48	**2.416**	3.750	3.719	**3.575**	**2.430**	3.941	2.611	**2.538**	**2.387**	2.472	4.144	**2.588**
	60	**2.539**	2.791	**2.677**	3.980	**3.690**	4.298	**2.801**	4.050	**2.581**	3.028	4.294	**2.643**
NASDAQ	24	**0.472**	0.565	**0.501**	0.515	**0.508**	0.595	**0.471**	0.641	1.356	**0.573**	**0.528**	1.291
	36	**0.780**	0.989	**1.134**	1.149	**0.817**	1.269	**0.724**	0.913	1.286	**1.131**	**0.786**	1.301
	48	**1.064**	1.201	**1.152**	1.181	1.125	**1.095**	**1.151**	1.166	1.184	**1.171**	**1.066**	1.194
	60	1.247	**1.092**	**1.021**	1.077	1.277	**1.051**	1.105	**1.082**	1.077	**1.063**	1.278	**1.080**
Long-term forecasting													
Electricity	96	**0.269**	0.348	**0.332**	0.825	**0.270**	0.819	**0.273**	0.829	**0.789**	0.847	**0.273**	0.388
	192	**0.347**	0.838	**0.353**	0.859	**0.813**	0.858	0.852	**0.851**	**0.747**	0.859	**0.815**	0.861
	336	0.868	**0.865**	**0.861**	0.867	**0.860**	0.866	**0.839**	0.867	**0.821**	0.868	**0.863**	div.
	720	div.	div.	**0.895**	div.	**0.877**	0.891	**0.882**	div.	**0.913**	div.	**0.892**	div.
Solar	96	**0.370**	0.635	**0.238**	0.382	**0.352**	0.706	**0.658**	0.737	**0.339**	0.382	**0.342**	0.647
	192	**0.677**	0.714	**0.397**	0.687	**0.671**	0.770	**0.414**	0.677	div.	**0.406**	**0.689**	0.760
	336	div.	**0.742**	**0.709**	0.738	**0.700**	0.759	**0.725**	0.755	**0.742**	0.764	**0.720**	0.764
	720	**0.729**	0.753	**0.731**	0.746	**0.745**	0.751	div.	**0.463**	div.	**0.751**	div.	**0.735**
Traffic	96	**0.882**	1.345	**0.796**	0.952	**0.901**	1.436	**0.881**	1.435	1.218	**0.960**	**0.920**	0.961
	192	1.412	**1.393**	**1.321**	1.383	1.422	**1.388**	**1.288**	1.435	**1.341**	1.439	**1.360**	1.402
	336	1.404	**1.401**	**1.395**	1.436	1.435	**1.428**	**1.390**	1.429	1.450	**1.423**	**1.391**	1.409
	720	**1.421**	div.	**1.423**	div.	**1.413**	div.	**1.438**	div.	**1.444**	div.	div.	**1.441**
ETTh1	96	**0.447**	0.499	**0.436**	0.707	**0.448**	0.715	**0.599**	0.703	**0.438**	0.709	**0.441**	0.699
	192	**0.518**	0.713	**0.705**	0.731	**0.705**	0.719	**0.686**	0.717	**0.709**	0.718	**0.709**	0.715
	336	**0.718**	div.	**0.698**	div.	**0.709**	0.712	**0.665**	0.705	**0.716**	div.	0.702	**0.696**
	720	**0.756**	div.	**0.721**	div.	div.	div.	**0.719**	div.	0.756	**0.747**	**0.703**	div.
ETTm1	96	0.476	**0.464**	0.459	**0.441**	0.461	**0.444**	div.	**0.682**	0.454	**0.451**	0.481	**0.473**
	192	**0.656**	0.683	**0.486**	0.683	**0.478**	div.	**0.485**	0.680	**0.484**	0.650	**0.479**	0.666
	336	0.700	**0.698**	0.691	**0.677**	**0.682**	0.702	**0.507**	0.704	**0.505**	0.698	**0.683**	0.689
	720	div.	div.	**0.713**	0.716	**0.709**	0.719	**0.718**	0.722	div.	div.	**0.709**	div.
Wind	96	**0.913**	1.247	**0.895**	0.899	**0.911**	1.254	div.	**1.229**	**0.895**	0.978	0.916	**0.897**
	192	**1.112**	1.315	**1.066**	1.113	**1.086**	1.109	**1.091**	1.312	**1.092**	1.306	**1.131**	div.
	336	**1.261**	div.	**1.357**	1.373	**1.233**	1.402	**1.271**	div.	**1.259**	div.	**1.235**	div.
	720	**1.422**	div.	**1.419**	div.	**1.415**	1.422	div.	div.	**1.340**	div.	**1.399**	div.
ETTm2	96	**0.251**	0.292	**0.254**	0.289	**0.287**	0.289	**0.252**	0.292	**0.238**	0.290	**0.247**	0.299
	192	div.	**0.320**	div.	**0.318**	**0.304**	0.318	**0.313**	0.318	**0.317**	0.319	**0.320**	div.
	336	**0.353**	div.	**0.347**	div.	0.354	**0.352**	**0.345**	0.354	0.354	**0.350**	**0.355**	div.
	720	div.	div.	**0.421**	div.	**0.423**	div.	**0.422**	div.	**0.423**	div.	**0.423**	div.

Table 17, continued.

Dataset	H	2021		2022		2023		2024		2025		2026	
		OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM
Exchange	96	**0.268**	0.436	**0.331**	0.427	**0.255**	0.436	**0.424**	0.456	0.457	**0.451**	0.437	**0.421**
	192	0.579	**0.537**	**0.556**	div.	**0.573**	div.	0.549	**0.539**	**0.581**	div.	**0.546**	div.
	336	**0.777**	div.	**0.760**	div.	**0.519**	div.	**0.742**	div.	0.757	**0.735**	**0.766**	div.
	720	div.	div.	**1.312**	div.	**1.315**	div.	div.	div.	div.	div.	**1.302**	div.
ETTh2	96	**0.354**	0.372	**0.369**	0.370	**0.367**	0.369	**0.355**	0.376	**0.357**	0.358	**0.357**	0.377
	192	0.396	**0.383**	**0.381**	0.382	**0.376**	0.386	0.388	**0.384**	0.384	**0.384**	0.382	**0.379**
	336	**0.383**	0.391	**0.389**	div.	**0.388**	div.	**0.384**	div.	**0.409**	div.	**0.389**	div.
	720	**0.439**	div.	div.	div.	**0.443**	div.	**0.432**	div.	div.	**0.443**	**0.429**	div.
Weather	96	0.196	**0.195**	**0.196**	0.200	**0.187**	0.195	0.198	**0.191**	**0.193**	0.196	**0.175**	0.202
	192	**0.227**	0.297	**0.242**	0.297	**0.230**	0.281	**0.232**	0.296	**0.230**	0.296	0.277	**0.245**
	336	div.	**0.323**	**0.310**	div.	0.325	**0.323**	div.	div.	**0.323**	div.	**0.289**	div.
	720	**0.366**	div.	**0.366**	div.	0.375	**0.370**	**0.365**	0.370	div.	div.	**0.342**	div.

Table 18: normalized MAE for configuration 5, lower is better. Rows are (dataset, forecast horizon H) pairs, columns are the six random seeds, each split into the observer head (OSSM) and the standard head (SSM). The better of the two is in bold. "div." marks a run that diverged, i.e. reached a normalized MSE above 10 or produced no metric at all; divergence counts as a loss, so the head that converged is shown in bold.

Dataset	H	2021		2022		2023		2024		2025		2026	
		OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM
Short-term forecasting													
NN5	24	**0.804**	0.814	**0.883**	0.940	**0.642**	0.701	0.838	**0.747**	**0.703**	0.918	**0.670**	0.904
	36	**0.851**	0.924	**0.860**	0.921	**0.660**	0.924	**0.696**	0.925	**0.852**	0.937	**0.631**	0.906
	48	**0.740**	0.919	**0.628**	0.930	**0.641**	0.922	**0.813**	0.924	**0.904**	0.935	**0.638**	0.923
	60	**0.911**	0.927	**0.890**	0.931	**0.835**	0.927	**0.857**	0.926	**0.925**	0.929	**0.634**	0.922
NYSE	24	**0.436**	0.821	**0.334**	0.862	**0.602**	0.849	**0.324**	0.399	**0.393**	0.801	**0.399**	0.852
	36	**0.968**	0.989	**0.879**	0.969	**0.967**	0.970	**0.487**	0.923	**0.925**	0.968	**0.868**	0.969
	48	**1.016**	1.086	**1.049**	1.050	**0.631**	1.069	**0.637**	1.065	**0.558**	1.064	**0.934**	1.068
	60	**1.152**	1.175	**1.135**	1.155	1.170	**1.162**	1.162	**1.158**	**1.129**	1.160	**1.125**	1.166
ILI	24	1.066	**1.009**	**1.029**	1.030	**1.008**	1.104	**1.035**	1.068	**1.012**	1.439	1.048	**1.032**
	36	**1.136**	1.393	**1.071**	1.384	**1.044**	1.137	**1.086**	1.147	**1.004**	1.389	1.476	**1.069**
	48	**1.059**	1.405	1.390	**1.355**	**1.071**	1.430	1.102	**1.096**	**1.037**	1.064	1.481	**1.096**
	60	**1.088**	1.157	**1.126**	1.450	**1.359**	1.495	**1.164**	1.450	**1.082**	1.221	1.501	**1.114**
NASDAQ	24	**0.486**	0.548	**0.510**	0.528	**0.513**	0.580	**0.493**	0.596	0.900	**0.552**	**0.534**	0.873
	36	**0.640**	0.751	**0.796**	0.811	**0.665**	0.866	**0.607**	0.704	0.877	**0.807**	**0.635**	0.878
	48	**0.758**	0.849	**0.825**	0.842	**0.746**	0.750	**0.822**	0.830	0.845	**0.835**	**0.740**	0.848
	60	0.827	**0.816**	**0.768**	0.810	0.807	**0.791**	**0.790**	0.810	0.811	**0.801**	**0.802**	0.811
Long-term forecasting													
Electricity	96	**0.362**	0.432	**0.419**	0.748	**0.364**	0.744	**0.366**	0.750	**0.726**	0.760	**0.366**	0.460
	192	**0.429**	0.753	**0.435**	0.765	**0.737**	0.764	0.761	**0.761**	**0.704**	0.765	**0.738**	0.765
	336	0.768	**0.767**	**0.765**	0.768	**0.764**	0.767	**0.751**	0.767	**0.739**	0.767	**0.765**	div.
	720	div.	div.	**0.776**	div.	**0.766**	0.775	**0.769**	div.	**0.781**	div.	**0.776**	div.
Solar	96	**0.479**	0.627	**0.322**	0.484	**0.443**	0.654	**0.617**	0.670	**0.441**	0.480	**0.447**	0.609
	192	**0.638**	0.657	**0.498**	0.632	**0.636**	0.695	**0.507**	0.651	div.	**0.503**	**0.653**	0.691
	336	div.	**0.682**	**0.664**	0.679	**0.657**	0.692	**0.670**	0.690	**0.681**	0.699	**0.667**	0.699
	720	**0.679**	0.695	**0.683**	0.689	**0.688**	0.694	div.	**0.520**	div.	**0.694**	div.	**0.682**
Traffic	96	**0.526**	0.775	**0.458**	0.538	**0.519**	0.809	**0.524**	0.809	0.734	**0.542**	**0.526**	0.540
	192	0.803	**0.797**	**0.787**	0.792	0.806	**0.795**	**0.767**	0.808	**0.769**	0.809	**0.782**	0.800
	336	0.799	**0.797**	**0.793**	0.806	0.808	**0.805**	**0.791**	0.804	0.810	**0.804**	**0.791**	0.800
	720	**0.799**	div.	**0.802**	div.	**0.796**	div.	**0.808**	div.	**0.809**	div.	div.	**0.807**

Table 18, continued.

Dataset	H	2021		2022		2023		2024		2025		2026	
		OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM
ETTh1	96	**0.463**	0.497	**0.460**	0.568	**0.465**	0.583	**0.532**	0.579	**0.465**	0.576	**0.462**	0.573
	192	**0.503**	0.584	**0.581**	0.596	**0.583**	0.586	**0.575**	0.588	0.585	**0.583**	0.585	**0.580**
	336	**0.597**	div.	**0.590**	div.	**0.591**	0.592	**0.575**	0.593	**0.593**	div.	0.589	**0.580**
	720	**0.638**	div.	**0.613**	div.	div.	div.	**0.616**	div.	0.637	**0.633**	**0.603**	div.
ETTm1	96	0.463	**0.463**	0.466	**0.458**	0.465	**0.456**	div.	**0.548**	0.461	**0.459**	0.474	**0.469**
	192	**0.539**	0.548	**0.480**	0.541	**0.475**	div.	**0.473**	0.545	**0.483**	0.538	**0.473**	0.540
	336	**0.556**	0.559	0.552	**0.548**	**0.548**	0.557	**0.488**	0.566	**0.485**	0.556	**0.548**	0.551
	720	div.	div.	**0.562**	0.568	**0.565**	0.573	**0.569**	0.577	div.	div.	**0.565**	div.
Wind	96	**0.668**	0.828	**0.654**	0.662	**0.665**	0.834	div.	**0.823**	**0.663**	0.717	0.666	**0.659**
	192	**0.769**	0.854	**0.752**	0.773	**0.757**	0.769	**0.756**	0.855	**0.761**	0.855	**0.782**	div.
	336	**0.832**	div.	**0.868**	0.876	**0.820**	0.891	**0.834**	div.	**0.831**	div.	**0.822**	div.
	720	**0.893**	div.	**0.893**	div.	**0.890**	0.895	div.	div.	**0.862**	div.	**0.883**	div.
ETTm2	96	**0.334**	0.356	**0.333**	0.353	**0.353**	0.354	**0.331**	0.356	**0.323**	0.354	**0.327**	0.360
	192	div.	**0.370**	div.	**0.369**	**0.364**	0.370	**0.366**	0.370	**0.369**	0.370	**0.371**	div.
	336	**0.388**	div.	**0.384**	div.	0.389	**0.388**	**0.384**	0.389	0.389	**0.387**	**0.389**	div.
	720	div.	div.	**0.430**	div.	**0.430**	div.	**0.429**	div.	**0.429**	div.	**0.430**	div.
Exchange	96	**0.378**	0.497	**0.419**	0.495	**0.363**	0.498	**0.483**	0.508	0.507	**0.506**	0.501	**0.491**
	192	0.572	**0.560**	**0.571**	div.	**0.578**	div.	0.564	**0.560**	**0.581**	div.	**0.563**	div.
	336	**0.686**	div.	**0.679**	div.	**0.538**	div.	**0.669**	div.	0.678	**0.667**	**0.682**	div.
	720	div.	div.	**0.862**	div.	**0.863**	div.	div.	div.	div.	div.	**0.858**	div.
ETTh2	96	**0.411**	0.419	0.418	**0.413**	**0.418**	0.419	**0.411**	0.420	**0.414**	0.416	**0.413**	0.418
	192	0.435	**0.422**	0.422	**0.420**	0.426	**0.425**	0.432	**0.422**	0.425	**0.424**	0.427	**0.423**
	336	**0.425**	0.431	**0.432**	div.	**0.429**	div.	**0.426**	div.	**0.438**	div.	**0.431**	div.
	720	**0.464**	div.	div.	div.	**0.466**	div.	**0.458**	div.	div.	**0.466**	**0.457**	div.
Weather	96	0.257	**0.255**	**0.253**	0.260	**0.246**	0.253	0.257	**0.250**	**0.254**	0.258	**0.237**	0.257
	192	**0.278**	0.341	**0.296**	0.336	**0.282**	0.318	**0.285**	0.336	**0.283**	0.337	0.314	**0.294**
	336	div.	**0.349**	**0.336**	div.	0.355	**0.349**	div.	div.	**0.349**	div.	**0.326**	div.
	720	**0.376**	div.	**0.377**	div.	**0.382**	0.384	**0.375**	0.383	div.	div.	**0.356**	div.

E.6 CONFIGURATION 6

Table 19: Hyper-parameters of configuration 6. Every other setting follows the TFB defaults, and the two heads share all of them, including the random seeds.

Hyper-parameter	Value
Encoder layers e_{layers}	2
State dimension d_x	256
Model dimension d_{model}	128
Learning rate	10^{-4}
Batch size	32

Table 20: normalized MSE for configuration 6, lower is better. Rows are (dataset, forecast horizon H) pairs, columns are the six random seeds, each split into the observer head (OSSM) and the standard head (SSM). The better of the two is in bold. "div." marks a run that diverged, i.e. reached a normalized MSE above 10 or produced no metric at all; divergence counts as a loss, so the head that converged is shown in bold.

Dataset	H	2021		2022		2023		2024		2025		2026	
		OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM
Short-term forecasting													
NN5	24	**0.939**	1.094	**0.876**	1.118	**0.901**	1.079	**0.927**	1.133	**0.923**	1.128	**0.894**	1.123
	36	**0.893**	1.105	**0.855**	1.097	**0.880**	1.082	**0.887**	1.122	**0.902**	1.112	**0.893**	1.118
	48	**0.886**	1.097	**0.831**	1.077	**0.848**	1.056	**0.870**	1.115	**0.878**	1.110	**0.866**	1.119
	60	**0.877**	1.115	**0.812**	1.085	**0.897**	1.112	**0.859**	1.146	**0.873**	1.144	**0.861**	1.128
NYSE	24	**0.251**	0.268	0.325	**0.321**	0.523	**0.270**	**0.251**	0.374	0.539	**0.270**	**0.251**	0.274
	36	**0.378**	0.743	**0.408**	0.710	**0.412**	0.421	0.583	**0.441**	**0.385**	0.744	0.510	**0.465**
	48	**1.390**	1.511	**0.886**	1.462	**0.925**	1.584	**0.759**	1.592	**0.949**	1.273	**0.589**	0.652
	60	**1.487**	1.957	**1.780**	1.925	**1.407**	1.669	**1.635**	1.890	1.399	**0.753**	**1.147**	1.688
ILI	24	2.644	**2.271**	2.407	**2.333**	**2.481**	2.549	**2.307**	2.448	**2.308**	2.353	2.500	**2.349**
	36	2.470	**2.340**	**2.383**	2.430	**2.305**	2.561	**2.279**	2.562	**2.415**	2.468	2.590	**2.411**
	48	**2.393**	2.520	**2.310**	2.455	**2.350**	2.545	**2.403**	2.554	**2.316**	2.627	**2.463**	2.530
	60	**2.589**	3.148	**2.437**	2.758	**2.597**	2.684	**2.447**	2.864	**2.381**	2.664	**2.637**	2.729
NASDAQ	24	**0.462**	0.475	**0.474**	0.502	0.500	**0.446**	0.459	**0.384**	0.513	**0.484**	0.482	**0.431**
	36	**0.740**	0.787	0.768	**0.754**	0.723	**0.680**	0.756	**0.669**	**0.841**	0.860	0.791	**0.724**
	48	1.114	**1.062**	1.144	**1.099**	**1.078**	1.093	**1.056**	1.058	**1.080**	1.102	1.091	**1.015**
	60	1.199	**1.182**	**1.151**	1.165	**1.173**	1.260	1.245	**1.228**	1.387	**1.190**	1.374	**1.199**
Long-term forecasting													
Electricity	96	**0.270**	0.274	**0.259**	0.275	**0.254**	0.274	**0.257**	0.276	**0.271**	0.277	**0.254**	0.273
	192	**0.280**	0.360	**0.280**	0.359	**0.279**	0.837	**0.279**	0.361	**0.280**	0.362	**0.280**	0.360
	336	**0.300**	0.860	**0.359**	0.864	**0.290**	0.863	**0.292**	0.859	**0.366**	0.864	**0.291**	0.862
	720	**0.882**	0.890	**0.390**	0.887	0.889	**0.871**	0.891	**0.889**	**0.881**	0.891	**0.888**	0.889
Solar	96	**0.230**	0.237	**0.245**	0.250	**0.214**	0.262	**0.245**	0.250	**0.198**	0.272	0.243	**0.220**
	192	**0.248**	0.396	**0.238**	0.398	**0.254**	0.386	**0.249**	0.393	**0.244**	0.400	**0.245**	0.760
	336	**0.382**	0.727	**0.393**	0.763	**0.387**	0.765	**0.388**	0.748	**0.397**	0.462	**0.393**	0.427
	720	**0.387**	div.	**0.736**	div.	0.788	**0.743**	0.442	**0.437**	**0.417**	0.501	0.745	**0.731**
Traffic	96	**0.733**	0.838	**0.712**	0.939	**0.758**	0.835	**0.734**	0.937	**0.711**	0.831	**0.756**	0.840
	192	**0.857**	0.981	**0.816**	0.974	**0.872**	0.973	**0.855**	1.392	**0.818**	1.397	**0.856**	1.385
	336	**1.299**	1.409	**0.981**	1.416	**1.383**	1.397	**0.977**	1.392	**1.379**	1.413	**0.980**	1.412
	720	**1.409**	1.434	**1.430**	1.431	**1.433**	1.433	**1.428**	1.430	**1.431**	1.435	**1.428**	1.429
ETTh1	96	**0.440**	0.511	**0.439**	0.457	**0.435**	0.438	**0.441**	0.527	**0.464**	0.470	**0.468**	0.502
	192	0.644	**0.525**	**0.508**	0.705	**0.469**	0.713	**0.535**	0.713	0.553	**0.529**	0.554	**0.524**
	336	0.694	**0.691**	**0.628**	0.692	**0.599**	0.711	**0.701**	0.702	**0.695**	0.708	**0.527**	0.702
	720	0.716	**0.711**	**0.704**	0.709	0.731	**0.726**	0.713	**0.707**	0.725	**0.716**	0.719	**0.707**
ETTm1	96	**0.361**	0.381	**0.381**	0.381	**0.383**	0.392	**0.379**	0.396	**0.377**	0.381	0.394	**0.377**
	192	**0.465**	0.512	**0.481**	0.495	**0.401**	0.515	**0.422**	0.492	0.493	**0.490**	0.505	**0.489**
	336	**0.505**	0.527	**0.516**	0.537	**0.512**	0.695	**0.509**	0.527	**0.508**	0.532	0.676	**0.488**
	720	**0.711**	0.712	**0.708**	0.711	0.716	**0.566**	**0.711**	0.714	**0.529**	0.775	0.718	**0.716**
Wind	96	0.962	**0.880**	0.902	**0.869**	0.919	**0.908**	**0.893**	0.896	0.924	**0.900**	**0.893**	0.897
	192	1.186	**1.121**	1.087	**1.057**	1.164	**1.071**	1.142	**1.069**	1.154	**1.101**	**1.080**	1.103
	336	1.245	**1.244**	**1.235**	1.256	1.284	**1.251**	**1.224**	1.375	**1.252**	1.269	1.292	**1.242**
	720	**1.346**	1.416	1.373	**1.360**	**1.401**	1.423	**1.318**	1.414	**1.338**	1.420	1.403	**1.346**
ETTm2	96	**0.239**	0.247	**0.228**	0.243	**0.247**	0.249	0.246	**0.244**	**0.229**	0.237	0.244	**0.229**
	192	**0.297**	0.300	0.297	**0.294**	**0.295**	0.314	**0.289**	0.296	**0.282**	0.297	**0.291**	0.302
	336	**0.352**	0.363	**0.345**	0.351	0.355	**0.351**	**0.343**	0.352	**0.339**	0.341	**0.341**	0.342
	720	**0.422**	0.422	0.425	**0.419**	**0.418**	0.422	**0.422**	0.422	**0.422**	0.423	**0.415**	0.424

Table 20, continued.

Dataset	H	2021		2022		2023		2024		2025		2026	
		OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM
Exchange	96	**0.211**	0.227	**0.178**	0.211	**0.215**	0.238	**0.215**	0.278	0.266	**0.229**	0.232	**0.210**
	192	**0.406**	0.541	**0.292**	0.550	**0.535**	0.545	**0.350**	0.542	**0.548**	0.550	**0.378**	0.541
	336	**0.687**	0.727	0.736	**0.733**	0.731	**0.726**	0.734	**0.731**	**0.732**	0.734	**0.715**	0.719
	720	1.314	**1.308**	**1.308**	1.314	**1.052**	1.315	1.314	**1.311**	1.318	**1.316**	**1.300**	1.300
ETTh2	96	0.363	**0.359**	0.367	**0.357**	**0.331**	0.364	**0.332**	0.365	0.370	**0.356**	**0.340**	0.378
	192	0.379	**0.372**	**0.371**	0.391	**0.372**	0.390	**0.376**	0.378	0.385	**0.383**	**0.380**	0.397
	336	**0.383**	0.384	0.384	**0.383**	0.407	**0.387**	**0.383**	0.385	**0.387**	0.388	**0.390**	0.414
	720	0.424	**0.422**	0.424	**0.420**	**0.429**	0.431	0.426	**0.424**	0.433	**0.431**	**0.425**	div.
Weather	96	**0.168**	0.170	0.177	**0.164**	0.176	**0.170**	0.167	**0.164**	0.172	**0.164**	0.169	**0.163**
	192	**0.217**	0.278	**0.225**	0.227	**0.238**	0.238	**0.225**	0.228	**0.231**	0.231	**0.221**	0.232
	336	**0.270**	0.290	**0.275**	0.279	**0.274**	0.282	**0.272**	0.324	**0.271**	0.323	**0.272**	0.323
	720	**0.367**	0.368	**0.367**	0.369	**0.348**	0.364	0.368	**0.350**	**0.367**	0.369	0.348	**0.346**

Table 21: normalized MAE for configuration 6, lower is better. Rows are (dataset, forecast horizon H) pairs, columns are the six random seeds, each split into the observer head (OSSM) and the standard head (SSM). The better of the two is in bold. "div." marks a run that diverged, i.e. reached a normalized MSE above 10 or produced no metric at all; divergence counts as a loss, so the head that converged is shown in bold.

Dataset	H	2021		2022		2023		2024		2025		2026	
		OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM
Short-term forecasting													
NN5	24	**0.719**	0.796	**0.679**	0.804	**0.693**	0.780	**0.715**	0.811	**0.703**	0.810	**0.692**	0.813
	36	**0.707**	0.809	**0.683**	0.802	**0.700**	0.794	**0.707**	0.813	**0.710**	0.810	**0.705**	0.820
	48	**0.710**	0.809	**0.676**	0.796	**0.688**	0.786	**0.703**	0.813	**0.703**	0.812	**0.698**	0.825
	60	**0.707**	0.820	**0.668**	0.801	**0.718**	0.818	**0.698**	0.831	**0.704**	0.831	**0.697**	0.831
NYSE	24	**0.337**	0.363	0.412	**0.406**	0.505	**0.365**	**0.348**	0.449	0.526	**0.367**	**0.337**	0.362
	36	**0.427**	0.619	**0.456**	0.623	**0.454**	0.460	0.576	**0.483**	**0.432**	0.643	0.513	**0.490**
	48	**0.909**	0.953	**0.700**	0.935	**0.707**	0.993	**0.652**	1.001	**0.710**	0.864	**0.556**	0.588
	60	**0.927**	1.126	**1.059**	1.113	**0.912**	1.004	**1.003**	1.105	0.900	**0.634**	**0.828**	1.037
ILI	24	1.079	**0.999**	**1.045**	1.079	**1.099**	1.103	**1.022**	1.084	**1.002**	1.044	1.105	**1.050**
	36	1.050	**1.002**	**1.009**	1.036	**1.066**	1.123	**1.035**	1.103	**1.030**	1.049	1.119	**1.061**
	48	1.035	**1.033**	**0.997**	1.030	**1.034**	1.097	**1.030**	1.096	**1.024**	1.055	**1.072**	1.086
	60	**1.067**	1.269	**1.036**	1.144	**1.119**	1.122	**1.047**	1.158	**1.031**	1.091	**1.078**	1.118
NASDAQ	24	**0.482**	0.493	**0.495**	0.519	0.501	**0.487**	0.481	**0.453**	0.518	**0.505**	0.485	**0.465**
	36	**0.622**	0.644	0.639	**0.637**	0.615	**0.601**	0.625	**0.606**	0.658	**0.657**	0.643	**0.622**
	48	0.769	**0.749**	**0.754**	0.764	**0.749**	0.755	**0.735**	0.741	**0.747**	0.749	0.748	**0.730**
	60	**0.786**	0.793	**0.774**	0.778	**0.784**	0.811	**0.796**	0.799	0.835	**0.783**	0.836	**0.790**
Long-term forecasting													
Electricity	96	**0.365**	0.368	**0.352**	0.368	**0.349**	0.368	**0.351**	0.370	**0.365**	0.370	**0.348**	0.367
	192	**0.372**	0.440	**0.373**	0.440	**0.371**	0.753	**0.371**	0.441	**0.372**	0.442	**0.372**	0.441
	336	**0.390**	0.764	**0.437**	0.766	**0.381**	0.766	**0.382**	0.763	**0.443**	0.766	**0.380**	0.766
	720	**0.771**	0.775	**0.457**	0.774	0.775	**0.764**	0.775	**0.774**	**0.770**	0.775	**0.774**	0.775
Solar	96	0.320	**0.320**	0.337	**0.333**	**0.316**	0.337	0.342	**0.328**	**0.299**	0.345	0.346	**0.316**
	192	**0.353**	0.494	**0.338**	0.495	**0.351**	0.490	**0.351**	0.492	**0.343**	0.496	**0.343**	0.692
	336	**0.491**	0.676	**0.497**	0.693	**0.494**	0.697	**0.491**	0.686	**0.494**	0.532	**0.500**	0.517
	720	**0.480**	div.	**0.684**	div.	0.706	**0.687**	0.520	**0.517**	**0.506**	0.540	0.690	**0.682**
Traffic	96	**0.444**	0.490	**0.428**	0.537	**0.439**	0.494	**0.436**	0.534	**0.428**	0.488	**0.439**	0.490
	192	**0.488**	0.549	**0.460**	0.546	**0.499**	0.546	**0.483**	0.797	**0.453**	0.799	**0.484**	0.794
	336	**0.794**	0.800	**0.544**	0.803	**0.785**	0.792	**0.543**	0.792	**0.781**	0.801	**0.543**	0.801
	720	**0.793**	0.806	**0.805**	0.805	**0.805**	0.806	**0.803**	0.804	**0.805**	0.806	**0.804**	0.804

Table 21, continued.

Dataset	H	2021		2022		2023		2024		2025		2026	
		OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM
ETTh1	96	**0.460**	0.501	**0.458**	0.468	0.459	**0.456**	**0.455**	0.507	0.474	**0.469**	**0.475**	0.499
	192	0.550	**0.509**	**0.493**	0.574	**0.476**	0.582	**0.516**	0.584	0.521	**0.510**	0.526	**0.513**
	336	0.586	**0.584**	**0.547**	0.576	**0.547**	0.594	0.591	**0.591**	**0.583**	0.587	**0.514**	0.588
	720	0.613	**0.609**	**0.600**	0.603	0.623	**0.618**	0.616	**0.609**	0.614	**0.608**	0.613	**0.602**
ETTm1	96	**0.413**	0.415	0.427	**0.418**	0.423	**0.423**	**0.421**	0.425	0.420	**0.414**	0.431	**0.414**
	192	**0.468**	0.485	**0.468**	0.478	**0.432**	0.487	**0.440**	0.479	**0.473**	0.481	0.481	**0.477**
	336	**0.489**	0.491	**0.490**	0.495	**0.484**	0.555	**0.486**	0.492	**0.481**	0.491	0.546	**0.478**
	720	**0.566**	0.567	0.565	**0.562**	0.570	**0.510**	**0.566**	0.572	**0.497**	0.610	0.569	**0.567**
Wind	96	0.700	**0.653**	0.664	**0.645**	0.670	**0.669**	0.663	**0.660**	0.684	**0.668**	**0.657**	0.657
	192	0.802	**0.777**	0.758	**0.740**	0.779	**0.748**	0.779	**0.743**	0.789	**0.764**	**0.746**	0.765
	336	**0.825**	0.826	**0.829**	0.829	0.846	**0.829**	**0.824**	0.877	**0.837**	0.837	0.844	**0.826**
	720	**0.867**	0.892	0.872	**0.869**	**0.885**	0.899	**0.861**	0.892	**0.862**	0.895	0.885	**0.866**
ETTm2	96	**0.326**	0.331	**0.311**	0.327	**0.322**	0.332	**0.326**	0.329	**0.318**	0.325	0.328	**0.318**
	192	**0.358**	0.361	**0.357**	0.358	**0.358**	0.367	**0.354**	0.360	**0.352**	0.359	**0.355**	0.361
	336	**0.387**	0.394	**0.385**	0.386	0.390	**0.388**	**0.384**	0.388	**0.381**	0.383	**0.385**	0.386
	720	0.429	**0.428**	0.431	**0.427**	**0.427**	0.429	**0.428**	0.429	**0.429**	0.429	**0.426**	0.430
Exchange	96	**0.342**	0.345	**0.323**	0.343	**0.347**	0.372	**0.351**	0.385	**0.368**	0.368	**0.343**	0.351
	192	**0.473**	0.561	**0.403**	0.568	**0.558**	0.565	**0.443**	0.560	**0.566**	0.567	**0.453**	0.563
	336	**0.638**	0.662	0.670	**0.669**	0.666	**0.664**	0.665	**0.664**	**0.665**	0.667	**0.655**	0.661
	720	0.861	**0.859**	**0.861**	0.864	**0.781**	0.862	0.860	**0.858**	0.862	**0.862**	**0.857**	0.857
ETTh2	96	0.414	**0.413**	0.419	**0.407**	**0.395**	0.419	**0.397**	0.416	0.422	**0.414**	**0.401**	0.421
	192	0.421	**0.417**	**0.417**	0.431	**0.414**	0.432	**0.419**	0.421	0.427	**0.425**	**0.423**	0.429
	336	**0.426**	0.426	0.426	**0.424**	0.441	**0.430**	**0.426**	0.427	**0.430**	0.430	**0.430**	0.442
	720	0.454	**0.453**	0.453	**0.449**	**0.457**	0.458	0.456	**0.454**	0.460	**0.458**	**0.455**	div.
Weather	96	**0.230**	0.231	0.238	**0.225**	0.237	**0.232**	0.230	**0.226**	0.234	**0.225**	0.232	**0.223**
	192	**0.271**	0.316	0.277	**0.276**	0.288	**0.283**	**0.278**	0.280	0.278	**0.275**	**0.274**	0.280
	336	**0.307**	0.322	**0.310**	0.313	**0.311**	0.316	**0.307**	0.351	**0.308**	0.350	**0.307**	0.351
	720	**0.378**	0.382	**0.378**	0.382	**0.364**	0.378	0.378	**0.364**	**0.378**	0.382	0.362	**0.361**

E.7 CONFIGURATION 7

Table 22: Hyper-parameters of configuration 7. Every other setting follows the TFB defaults, and the two heads share all of them, including the random seeds.

Hyper-parameter	Value
Encoder layers e_{layers}	2
State dimension d_x	256
Model dimension d_{model}	256
Learning rate	10^{-4}
Batch size	32

Table 23: normalized MSE for configuration 7, lower is better. Rows are (dataset, forecast horizon H) pairs, columns are the six random seeds, each split into the observer head (OSSM) and the standard head (SSM). The better of the two is in bold. "div." marks a run that diverged, i.e. reached a normalized MSE above 10 or produced no metric at all; divergence counts as a loss, so the head that converged is shown in bold.

Dataset	H	2021		2022		2023		2024		2025		2026	
		OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM
Short-term forecasting													
NN5	24	**0.813**	0.971	**0.832**	1.027	**0.877**	1.081	**0.802**	1.022	**0.866**	0.979	**0.834**	1.021
	36	**0.770**	0.933	**0.810**	1.031	**0.872**	1.056	**0.770**	0.979	**0.830**	0.995	**0.828**	0.992
	48	**0.752**	1.167	**0.801**	1.055	**0.836**	1.035	**0.749**	0.985	**0.809**	1.029	**0.802**	1.236
	60	**0.723**	0.908	**0.814**	1.011	**0.830**	1.083	**0.793**	1.080	**0.798**	1.020	**0.803**	1.237
NYSE	24	0.263	**0.254**	**0.228**	0.244	**0.240**	0.282	0.238	**0.228**	0.296	**0.257**	**0.235**	0.866
	36	0.485	**0.435**	0.513	**0.504**	**0.398**	0.542	**0.491**	0.870	**0.568**	1.165	**0.771**	0.953
	48	0.862	**0.742**	**1.019**	1.597	**0.788**	1.493	**0.893**	0.955	**0.763**	1.551	**0.469**	1.624
	60	**0.915**	1.718	**1.566**	1.989	**1.034**	1.948	**0.746**	1.995	**0.730**	1.592	**1.032**	1.911
ILI	24	2.503	**2.187**	2.424	**2.316**	2.456	**2.224**	2.715	**2.291**	**2.223**	2.439	2.233	**2.224**
	36	**2.360**	2.389	**2.322**	2.451	2.321	**2.293**	**2.349**	2.461	**2.258**	2.551	2.390	**2.315**
	48	**2.173**	2.555	**2.327**	2.473	**2.246**	2.516	2.510	**2.345**	**2.230**	2.353	**2.404**	2.420
	60	**2.340**	2.650	**2.407**	3.326	2.502	**2.269**	**2.407**	2.635	2.445	**2.381**	**2.487**	2.621
NASDAQ	24	0.627	**0.493**	**0.427**	0.453	0.445	**0.439**	0.429	**0.420**	**0.392**	0.417	0.609	**0.507**
	36	**0.768**	0.839	**0.726**	0.750	0.751	**0.735**	0.778	**0.727**	0.657	**0.600**	**0.770**	0.797
	48	1.166	**1.101**	1.059	**1.045**	**1.069**	1.141	1.126	**1.104**	**1.011**	1.051	**1.022**	1.122
	60	**1.205**	1.331	1.180	**1.176**	1.182	**1.181**	1.343	**1.219**	**1.156**	1.248	1.435	**1.196**
Long-term forecasting													
Electricity	96	**0.249**	0.276	**0.242**	0.275	**0.245**	0.276	**0.254**	0.276	**0.243**	0.275	**0.246**	0.279
	192	**0.277**	0.835	**0.280**	0.358	**0.265**	0.367	**0.278**	0.367	**0.280**	0.852	**0.276**	0.361
	336	**0.860**	0.867	**0.362**	0.864	**0.841**	0.864	**0.291**	0.862	**0.360**	0.864	**0.290**	0.865
	720	0.888	**0.884**	0.890	**0.889**	0.896	**0.890**	**0.877**	0.879	0.895	**0.891**	**0.890**	0.890
Solar	96	0.243	**0.221**	**0.220**	0.236	**0.207**	0.250	**0.209**	0.227	**0.200**	0.251	0.221	**0.214**
	192	**0.240**	0.404	**0.250**	0.393	**0.235**	0.410	**0.243**	0.412	**0.240**	0.739	**0.246**	0.396
	336	**0.381**	0.416	**0.374**	0.477	**0.372**	0.743	**0.401**	0.722	**0.378**	0.762	**0.375**	0.758
	720	0.732	**0.455**	**0.712**	0.721	0.745	**0.730**	**0.393**	div.	0.738	**0.726**	**0.744**	div.
Traffic	96	**0.725**	0.775	**0.742**	0.830	**0.710**	0.828	**0.685**	0.908	**0.676**	0.837	**0.721**	0.935
	192	**0.807**	0.975	**0.854**	1.387	**0.840**	1.369	**0.790**	0.986	**0.810**	1.365	**0.798**	1.383
	336	**1.350**	1.396	**0.876**	1.405	**0.875**	1.413	**0.865**	1.410	**1.356**	1.414	**0.976**	1.415
	720	**1.413**	1.432	**1.401**	1.430	**1.432**	1.432	**1.408**	div.	**1.009**	1.433	1.436	**1.433**
ETTh1	96	**0.431**	0.437	0.484	**0.446**	**0.428**	0.499	0.439	**0.432**	**0.438**	0.512	**0.448**	0.453
	192	**0.531**	0.711	**0.468**	0.525	**0.467**	0.526	**0.519**	0.715	**0.467**	0.545	0.470	**0.461**
	336	**0.537**	0.701	0.711	**0.700**	**0.515**	0.534	0.708	**0.705**	**0.519**	0.701	**0.524**	0.703
	720	**0.699**	0.701	**0.709**	0.716	**0.564**	0.711	**0.720**	0.730	**0.714**	0.719	**0.718**	0.725
ETTm1	96	**0.379**	0.380	0.395	**0.377**	0.393	**0.351**	0.385	**0.360**	**0.367**	0.383	0.388	**0.383**
	192	0.490	**0.480**	0.513	**0.474**	**0.440**	0.478	**0.402**	0.547	**0.481**	0.494	**0.465**	0.478
	336	**0.507**	0.527	**0.507**	0.526	**0.511**	0.688	**0.500**	0.703	**0.516**	0.523	0.513	**0.504**
	720	0.715	**0.589**	**0.605**	div.	**0.713**	0.714	**0.709**	0.713	**0.555**	0.714	**0.538**	0.713
Wind	96	**0.888**	0.917	0.896	**0.883**	**0.877**	0.878	0.902	**0.872**	0.896	**0.878**	0.894	**0.864**
	192	**1.059**	1.105	1.108	**1.069**	1.082	**1.058**	**1.173**	1.307	1.102	**1.081**	**1.076**	1.112
	336	**1.214**	1.381	1.263	**1.242**	**1.192**	1.378	**1.243**	1.390	**1.232**	1.377	1.256	**1.227**
	720	**1.411**	1.421	**1.345**	1.417	**1.325**	1.414	**1.303**	1.417	**1.341**	1.418	**1.354**	1.411
ETTm2	96	**0.235**	0.241	**0.231**	0.243	**0.225**	0.246	**0.227**	0.234	**0.220**	0.247	**0.236**	0.250
	192	0.305	**0.287**	**0.297**	0.316	**0.299**	0.318	**0.283**	0.297	0.298	**0.293**	**0.299**	0.308
	336	0.354	**0.340**	**0.340**	0.364	**0.339**	0.354	**0.335**	0.353	**0.349**	0.352	**0.343**	0.352
	720	**0.421**	0.423	0.427	**0.422**	**0.416**	0.422	**0.428**	div.	**0.415**	0.420	**0.418**	0.422

Table 23, continued.

Dataset	H	2021		2022		2023		2024		2025		2026	
		OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM
Exchange	96	0.274	**0.245**	**0.186**	0.273	0.212	**0.184**	**0.201**	0.236	0.218	**0.179**	**0.198**	0.229
	192	**0.345**	0.556	**0.301**	0.434	**0.381**	0.552	**0.378**	0.553	**0.354**	0.546	**0.441**	0.547
	336	**0.736**	0.750	**0.724**	0.732	0.744	**0.733**	0.752	**0.749**	**0.614**	0.729	**0.605**	0.730
	720	1.336	**1.330**	1.320	**1.316**	**1.319**	div.	1.341	**1.340**	**1.314**	1.317	**1.311**	1.315
ETTh2	96	**0.349**	0.349	**0.355**	0.363	**0.347**	0.371	**0.349**	0.356	**0.329**	0.373	**0.340**	0.361
	192	**0.372**	0.377	0.390	**0.382**	**0.373**	0.385	**0.375**	0.383	**0.359**	0.380	0.397	**0.384**
	336	**0.379**	0.385	**0.380**	0.385	**0.382**	0.383	0.397	**0.388**	**0.382**	0.390	**0.382**	0.384
	720	0.424	**0.423**	0.425	**0.425**	0.426	**0.425**	**0.432**	0.437	**0.426**	0.428	**0.425**	0.426
Weather	96	**0.162**	0.166	0.169	**0.160**	0.166	**0.163**	**0.167**	0.167	0.169	**0.169**	0.169	**0.164**
	192	**0.206**	0.277	**0.209**	0.234	**0.225**	0.231	**0.237**	0.287	**0.221**	0.234	0.230	**0.227**
	336	**0.272**	0.276	**0.276**	0.314	**0.275**	0.323	**0.273**	0.282	0.299	**0.289**	**0.276**	0.308
	720	**0.333**	0.367	0.368	**0.367**	**0.353**	0.368	**0.366**	0.368	**0.368**	0.368	**0.350**	0.371

Table 24: normalized MAE for configuration 7, lower is better. Rows are (dataset, forecast horizon H) pairs, columns are the six random seeds, each split into the observer head (OSSM) and the standard head (SSM). The better of the two is in bold. "div." marks a run that diverged, i.e. reached a normalized MSE above 10 or produced no metric at all; divergence counts as a loss, so the head that converged is shown in bold.

Dataset	H	2021		2022		2023		2024		2025		2026	
		OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM
Short-term forecasting													
NN5	24	**0.635**	0.718	**0.652**	0.764	**0.686**	0.790	**0.633**	0.754	**0.678**	0.729	**0.657**	0.761
	36	**0.627**	0.710	**0.659**	0.775	**0.698**	0.787	**0.631**	0.743	**0.674**	0.752	**0.673**	0.756
	48	**0.624**	0.868	**0.662**	0.793	**0.687**	0.780	**0.628**	0.750	**0.668**	0.773	**0.665**	0.896
	60	**0.612**	0.705	**0.673**	0.772	**0.685**	0.806	**0.657**	0.803	**0.664**	0.768	**0.668**	0.899
NYSE	24	0.363	**0.344**	**0.335**	0.337	**0.341**	0.382	0.334	**0.328**	0.375	**0.354**	**0.330**	0.698
	36	0.504	**0.461**	0.537	**0.513**	**0.449**	0.542	**0.500**	0.686	**0.540**	0.832	**0.651**	0.719
	48	0.693	**0.657**	**0.762**	0.994	**0.671**	0.951	**0.716**	0.727	**0.623**	0.971	**0.462**	1.013
	60	**0.700**	1.031	**1.014**	1.141	**0.806**	1.130	**0.622**	1.145	**0.645**	0.994	**0.791**	1.110
ILI	24	1.042	**1.010**	1.045	**1.042**	1.060	**0.987**	1.132	**1.001**	**0.994**	1.053	1.020	**0.995**
	36	**1.026**	1.060	**1.030**	1.074	1.020	**1.002**	**1.026**	1.072	**1.009**	1.063	1.061	**1.029**
	48	**0.982**	1.083	1.026	**1.018**	**1.013**	1.083	1.055	**1.030**	1.024	**0.992**	1.068	**1.063**
	60	**1.023**	1.101	**1.044**	1.280	1.057	**0.956**	**1.037**	1.103	1.094	**0.987**	**1.066**	1.113
NASDAQ	24	0.552	**0.503**	**0.470**	0.484	**0.474**	0.477	0.471	**0.458**	**0.444**	0.467	0.534	**0.499**
	36	**0.628**	0.662	**0.613**	0.630	**0.622**	0.631	0.632	**0.625**	0.580	**0.567**	**0.623**	0.642
	48	0.765	**0.755**	**0.734**	0.736	**0.740**	0.762	0.757	**0.747**	**0.723**	0.742	**0.722**	0.764
	60	**0.783**	0.820	**0.780**	0.787	**0.782**	0.787	0.823	**0.787**	**0.774**	0.807	0.837	**0.789**
Long-term forecasting													
Electricity	96	**0.344**	0.370	**0.338**	0.369	**0.344**	0.370	**0.349**	0.370	**0.343**	0.369	**0.343**	0.373
	192	**0.369**	0.752	**0.371**	0.439	**0.355**	0.446	**0.370**	0.446	**0.372**	0.762	**0.368**	0.441
	336	**0.764**	0.767	**0.439**	0.766	**0.752**	0.766	**0.380**	0.766	**0.438**	0.766	**0.379**	0.766
	720	0.774	**0.771**	0.775	**0.774**	0.776	**0.775**	**0.767**	0.769	0.777	**0.775**	**0.775**	0.775
Solar	96	0.326	**0.316**	**0.313**	0.317	**0.306**	0.332	**0.294**	0.319	**0.308**	0.333	0.313	**0.308**
	192	**0.333**	0.500	**0.342**	0.492	**0.335**	0.500	**0.339**	0.499	**0.344**	0.681	**0.338**	0.492
	336	**0.491**	0.506	**0.481**	0.530	**0.476**	0.683	**0.496**	0.668	**0.490**	0.698	**0.485**	0.697
	720	0.680	**0.520**	0.668	**0.652**	0.690	**0.681**	**0.498**	div.	0.684	**0.680**	**0.689**	div.
Traffic	96	**0.441**	0.452	**0.443**	0.489	**0.429**	0.484	**0.414**	0.535	**0.406**	0.493	**0.437**	0.532
	192	**0.447**	0.546	**0.485**	0.794	**0.478**	0.785	**0.442**	0.552	**0.446**	0.780	**0.445**	0.792
	336	**0.768**	0.794	**0.494**	0.798	**0.496**	0.802	**0.484**	0.801	**0.769**	0.802	**0.542**	0.803
	720	**0.796**	0.805	**0.788**	0.805	**0.805**	0.805	**0.793**	div.	**0.553**	0.806	0.806	**0.806**

Table 24, continued.

Dataset	H	2021		2022		2023		2024		2025		2026	
		OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM
ETTh1	96	**0.453**	0.455	0.479	**0.459**	**0.457**	0.493	0.460	**0.455**	**0.452**	0.499	0.463	**0.463**
	192	**0.512**	0.577	**0.477**	0.512	**0.478**	0.512	**0.506**	0.585	**0.475**	0.519	0.475	**0.471**
	336	**0.521**	0.579	0.589	**0.579**	**0.507**	0.518	0.592	**0.588**	**0.509**	0.584	**0.513**	0.588
	720	0.603	**0.597**	**0.605**	0.611	**0.539**	0.609	**0.615**	0.622	**0.611**	0.613	**0.613**	0.617
ETTm1	96	0.418	**0.415**	0.427	**0.411**	0.425	**0.401**	0.422	**0.402**	0.412	**0.412**	0.427	**0.418**
	192	0.479	**0.478**	0.497	**0.473**	**0.448**	0.476	**0.430**	0.497	**0.470**	0.478	**0.463**	0.472
	336	**0.486**	0.492	**0.484**	0.493	**0.487**	0.551	**0.482**	0.561	**0.492**	0.495	0.491	**0.487**
	720	0.567	**0.521**	**0.531**	div.	**0.567**	0.568	**0.566**	0.569	**0.513**	0.566	**0.502**	0.566
Wind	96	**0.657**	0.670	0.659	**0.652**	0.648	**0.646**	0.659	**0.647**	**0.654**	0.656	0.656	**0.641**
	192	**0.735**	0.766	0.762	**0.747**	0.756	**0.741**	**0.797**	0.854	0.763	**0.757**	**0.747**	0.775
	336	**0.802**	0.882	0.832	**0.824**	**0.805**	0.879	**0.824**	0.887	**0.813**	0.878	0.829	**0.819**
	720	**0.889**	0.898	**0.865**	0.894	**0.858**	0.890	**0.854**	0.894	**0.860**	0.893	**0.867**	0.891
ETTm2	96	**0.321**	0.325	**0.318**	0.327	**0.315**	0.327	**0.318**	0.323	**0.307**	0.329	**0.324**	0.332
	192	0.360	**0.355**	**0.356**	0.368	**0.358**	0.370	**0.352**	0.361	0.359	**0.358**	**0.355**	0.363
	336	0.392	**0.384**	**0.384**	0.396	**0.383**	0.389	**0.381**	0.388	**0.385**	0.387	**0.385**	0.388
	720	**0.428**	0.429	0.432	**0.428**	**0.426**	0.429	**0.432**	div.	**0.425**	0.427	**0.427**	0.428
Exchange	96	0.375	**0.367**	**0.321**	0.376	0.344	**0.328**	**0.348**	0.356	0.356	**0.320**	**0.341**	0.354
	192	**0.434**	0.571	**0.410**	0.492	**0.455**	0.569	**0.454**	0.564	**0.455**	0.566	**0.491**	0.567
	336	**0.665**	0.675	**0.661**	0.666	0.674	**0.669**	0.674	**0.672**	**0.592**	0.665	**0.582**	0.667
	720	0.869	**0.866**	0.864	**0.862**	**0.865**	div.	0.869	**0.868**	**0.862**	0.863	**0.861**	0.863
ETTh2	96	0.405	**0.405**	**0.403**	0.415	**0.401**	0.419	**0.407**	0.412	**0.395**	0.422	**0.394**	0.413
	192	**0.417**	0.419	0.428	**0.422**	**0.418**	0.428	**0.420**	0.425	**0.408**	0.421	0.432	**0.427**
	336	**0.423**	0.426	**0.424**	0.427	0.426	**0.426**	0.436	**0.429**	**0.425**	0.428	**0.424**	0.425
	720	0.454	**0.453**	0.453	**0.453**	0.455	**0.455**	**0.459**	0.465	**0.453**	0.455	**0.454**	0.454
Weather	96	**0.223**	0.224	0.229	**0.220**	0.228	**0.224**	0.230	**0.228**	0.230	**0.230**	0.232	**0.222**
	192	**0.263**	0.316	**0.264**	0.284	**0.275**	0.278	**0.286**	0.324	**0.271**	0.283	0.285	**0.276**
	336	**0.309**	0.312	**0.309**	0.339	**0.309**	0.352	**0.304**	0.318	0.325	**0.323**	**0.312**	0.334
	720	**0.350**	0.378	0.378	**0.378**	**0.364**	0.378	**0.377**	0.380	**0.380**	0.381	**0.361**	0.386

E.8 CONFIGURATION 8

Table 25: Hyper-parameters of configuration 8. Every other setting follows the TFB defaults, and the two heads share all of them, including the random seeds.

Hyper-parameter	Value
Encoder layers e_{layers}	4
State dimension d_x	128
Model dimension d_{model}	128
Learning rate	10^{-3}
Batch size	32

Table 26: normalized MSE for configuration 8, lower is better. Rows are (dataset, forecast horizon H) pairs, columns are the six random seeds, each split into the observer head (OSSM) and the standard head (SSM). The better of the two is in bold. "div." marks a run that diverged, i.e. reached a normalized MSE above 10 or produced no metric at all; divergence counts as a loss, so the head that converged is shown in bold.

Dataset	H	2021		2022		2023		2024		2025		2026	
		OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM
Short-term forecasting													
NN5	24	0.882	**0.879**	**0.907**	1.198	0.876	**0.843**	0.893	**0.866**	**0.834**	0.898	**0.830**	0.876
	36	**0.929**	1.268	0.900	**0.840**	**1.296**	1.300	**1.303**	1.304	**1.089**	1.186	**1.108**	1.203
	48	**0.823**	1.310	1.230	**0.925**	**1.291**	1.293	**0.989**	1.325	**0.734**	1.287	**0.836**	1.271
	60	**1.252**	1.297	**1.174**	1.219	**1.248**	1.285	**0.865**	1.308	**0.796**	1.220	**0.803**	1.191
NYSE	24	0.521	**0.331**	**0.247**	0.492	**0.250**	0.292	**0.644**	1.189	**0.231**	0.250	**0.235**	0.253
	36	**0.454**	1.569	**0.364**	1.358	1.217	**0.501**	**0.510**	1.526	**0.459**	0.471	**0.439**	1.452
	48	**0.685**	1.681	**1.568**	1.655	**0.708**	1.702	1.727	**1.616**	**1.459**	1.806	**0.748**	1.940
	60	**2.002**	2.091	**1.116**	1.905	**1.937**	2.027	**1.892**	2.097	**1.914**	2.000	**1.942**	2.125
ILI	24	**2.191**	2.367	**2.255**	2.446	3.119	**2.581**	2.594	**1.604**	2.507	**2.359**	**2.109**	2.248
	36	3.488	**2.689**	**2.386**	2.438	**2.576**	2.987	2.474	**2.421**	**2.266**	2.298	2.533	**2.407**
	48	**2.828**	3.547	**2.449**	2.559	**2.439**	2.637	**2.514**	3.766	2.440	**2.288**	**2.552**	3.250
	60	**2.655**	4.099	4.044	**2.640**	4.095	**4.000**	**2.485**	2.612	**2.506**	4.058	**2.470**	2.632
NASDAQ	24	**0.475**	0.530	**0.435**	0.542	**0.497**	0.499	**0.427**	0.461	0.482	**0.470**	**0.431**	0.463
	36	**0.753**	0.815	**0.715**	1.284	**0.768**	0.819	**0.779**	0.999	0.930	**0.782**	**0.734**	1.286
	48	**1.077**	1.184	**1.063**	1.206	**1.129**	1.203	**1.155**	1.190	**1.116**	1.187	**1.053**	1.177
	60	1.158	**1.092**	1.254	**1.097**	1.345	**1.108**	1.204	**1.086**	1.156	**1.098**	**1.166**	1.238
Long-term forecasting													
Electricity	96	0.960	**0.862**	**0.265**	0.834	**0.323**	0.844	**0.245**	0.278	**0.249**	0.867	0.916	**0.857**
	192	**0.775**	0.862	div.	**0.849**	**0.826**	0.863	**0.280**	0.863	div.	**0.356**	**0.839**	0.864
	336	div.	**0.869**	**0.368**	div.	**0.866**	0.867	**0.867**	0.871	**0.866**	0.870	**0.871**	0.879
	720	0.890	**0.889**	div.	div.	**0.893**	div.	0.897	**0.893**	0.891	**0.886**	div.	div.
Solar	96	**0.247**	div.	div.	**0.320**	div.	div.	**0.237**	0.282	0.638	**0.355**	div.	div.
	192	**0.696**	0.773	0.701	**0.676**	div.	div.	div.	div.	**0.360**	0.772	div.	**0.759**
	336	**0.402**	div.	**0.405**	0.763	**0.722**	0.770	div.	**0.721**	**0.726**	0.754	**0.758**	div.
	720	0.789	**0.750**	**0.722**	0.736	div.	**0.751**	0.796	**0.734**	0.741	**0.730**	div.	**0.750**
Traffic	96	1.289	**0.931**	**1.235**	1.373	**1.237**	1.421	2.374	**1.443**	**0.751**	0.931	**0.874**	1.383
	192	**1.297**	1.389	1.347	**1.326**	**0.958**	1.743	**1.390**	1.436	**1.331**	1.364	**0.914**	1.389
	336	**1.409**	div.	**1.382**	1.525	**1.461**	div.	div.	**1.436**	**1.394**	1.429	**1.396**	1.450
	720	**1.413**	1.453	**1.508**	div.	div.	div.	1.669	**1.441**	**1.663**	div.	**1.431**	1.446
ETTh1	96	**0.432**	0.509	**0.497**	0.721	**0.464**	0.707	**0.429**	0.705	**0.516**	0.615	**0.428**	0.711
	192	**0.541**	0.700	**0.521**	0.718	**0.709**	0.719	**0.620**	0.716	**0.629**	0.719	0.709	**0.522**
	336	**0.706**	div.	div.	**0.709**	**0.715**	0.717	**0.631**	0.704	**0.704**	div.	0.713	**0.706**
	720	**0.747**	div.	**0.782**	div.	**0.757**	0.759	div.	div.	0.757	**0.723**	0.758	**0.743**
ETTm1	96	**0.470**	0.636	0.641	**0.455**	0.582	**0.469**	**0.469**	0.660	0.654	**0.641**	**0.386**	0.454
	192	**0.477**	div.	0.493	**0.485**	**0.468**	0.689	**0.491**	0.684	div.	**0.650**	div.	**0.482**
	336	**0.688**	div.	**0.683**	0.707	0.687	**0.549**	**0.497**	div.	0.723	**0.511**	div.	**0.697**
	720	**0.715**	0.718	**1.149**	div.	**0.706**	div.	**0.531**	div.	**0.716**	0.721	div.	**0.732**
Wind	96	**0.889**	1.234	**0.974**	1.174	0.918	**0.894**	**0.940**	1.251	**0.946**	1.232	**0.923**	div.
	192	**1.135**	1.309	**1.095**	1.319	**1.131**	1.138	**1.117**	1.321	**1.117**	1.336	**1.222**	div.
	336	1.381	**1.378**	1.248	**1.238**	**1.337**	1.386	**1.381**	div.	**1.227**	1.383	**1.239**	div.
	720	div.	div.	div.	div.	**1.402**	div.	div.	div.	**1.418**	1.425	**1.421**	div.
ETTm2	96	div.	**0.253**	**0.255**	0.294	**0.254**	0.295	0.256	**0.242**	**0.283**	0.297	**0.252**	0.270
	192	div.	**2.041**	0.342	**0.322**	0.365	**0.319**	**0.320**	0.320	**0.299**	0.323	**0.314**	0.318
	336	**0.361**	div.	**0.353**	0.356	0.360	**0.353**	**0.355**	0.358	**0.360**	div.	div.	div.
	720	div.	**0.425**	**0.425**	div.	**0.422**	div.	**0.423**	0.432	**0.424**	div.	0.426	**0.424**

Table 26, continued.

Dataset	H	2021		2022		2023		2024		2025		2026	
		OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM
Exchange	96	0.466	**0.260**	0.450	**0.331**	**0.276**	0.449	**0.383**	0.470	**0.247**	0.352	**0.342**	0.447
	192	**0.555**	div.	**0.593**	0.597	**0.543**	0.564	0.566	**0.550**	**0.596**	div.	0.571	**0.554**
	336	**0.785**	div.	**0.747**	div.	**0.742**	div.	0.760	**0.735**	div.	div.	**0.782**	div.
	720	1.374	**1.330**	div.	**1.320**	**1.374**	div.	div.	**1.315**	**1.312**	div.	**1.361**	div.
ETTh2	96	**0.362**	0.374	**0.368**	0.372	**0.359**	0.373	**0.369**	0.374	**0.378**	0.395	**0.364**	0.365
	192	0.384	**0.384**	**0.387**	div.	**0.384**	0.387	0.488	**0.388**	**0.388**	div.	0.806	**0.384**
	336	**0.385**	0.404	0.394	**0.387**	**0.397**	div.	**0.394**	div.	**0.395**	div.	0.407	**0.386**
	720	0.440	**0.433**	**0.430**	div.	div.	div.	div.	div.	**0.429**	div.	div.	div.
Weather	96	0.181	**0.179**	**0.180**	0.202	**0.184**	0.269	**0.185**	div.	0.274	**0.271**	**0.193**	0.202
	192	**0.241**	0.294	0.283	**0.247**	**0.255**	0.297	**0.227**	0.235	0.247	**0.246**	**0.234**	0.251
	336	**0.308**	0.396	0.324	**0.324**	**0.285**	0.287	div.	div.	div.	**0.324**	div.	**0.325**
	720	div.	div.	div.	div.	div.	**0.372**	0.378	**0.367**	**0.351**	div.	0.401	**0.368**

Table 27: normalized MAE for configuration 8, lower is better. Rows are (dataset, forecast horizon H) pairs, columns are the six random seeds, each split into the observer head (OSSM) and the standard head (SSM). The better of the two is in bold. "div." marks a run that diverged, i.e. reached a normalized MSE above 10 or produced no metric at all; divergence counts as a loss, so the head that converged is shown in bold.

Dataset	H	2021		2022		2023		2024		2025		2026	
		OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM
Short-term forecasting													
NN5	24	0.683	**0.673**	**0.696**	0.862	0.680	**0.650**	0.666	**0.660**	**0.643**	0.681	**0.636**	0.668
	36	**0.727**	0.913	0.711	**0.660**	**0.928**	0.930	0.930	**0.929**	**0.815**	0.866	**0.833**	0.859
	48	**0.676**	0.932	0.898	**0.713**	0.930	**0.928**	**0.762**	0.936	**0.611**	0.923	**0.687**	0.914
	60	**0.910**	0.930	**0.862**	0.889	**0.908**	0.927	**0.696**	0.933	**0.655**	0.888	**0.633**	0.878
NYSE	24	0.516	**0.410**	**0.363**	0.526	**0.354**	0.377	**0.584**	0.862	**0.329**	0.339	**0.346**	0.350
	36	**0.504**	1.001	**0.429**	0.920	0.850	**0.509**	**0.530**	0.992	0.504	**0.483**	**0.482**	0.952
	48	**0.638**	1.040	**0.991**	1.035	**0.645**	1.053	1.061	**1.014**	**0.952**	1.082	**0.658**	1.123
	60	**1.153**	1.178	**0.801**	1.115	**1.129**	1.162	**1.111**	1.178	**1.117**	1.152	**1.125**	1.187
ILI	24	**1.001**	1.065	**1.030**	1.063	1.140	**1.100**	1.111	**0.869**	1.059	**1.027**	**0.992**	1.024
	36	1.348	**1.106**	**1.039**	1.068	**1.077**	1.162	**1.064**	1.065	**1.023**	1.043	1.069	**1.064**
	48	**1.157**	1.340	**1.069**	1.128	**1.084**	1.108	**1.082**	1.398	1.059	**1.051**	**1.084**	1.276
	60	**1.114**	1.448	1.438	**1.106**	1.466	**1.461**	**1.093**	1.108	**1.093**	1.451	**1.068**	1.124
NASDAQ	24	**0.487**	0.532	**0.494**	0.535	**0.510**	0.521	**0.472**	0.498	**0.504**	0.506	**0.469**	0.489
	36	**0.621**	0.648	**0.614**	0.869	**0.624**	0.660	**0.641**	0.733	0.710	**0.649**	**0.616**	0.875
	48	**0.746**	0.843	**0.740**	0.851	**0.804**	0.850	**0.828**	0.846	**0.763**	0.843	**0.760**	0.844
	60	**0.772**	0.816	**0.810**	0.818	**0.813**	0.820	**0.783**	0.814	**0.776**	0.817	**0.776**	0.794
Long-term forecasting													
Electricity	96	0.792	**0.765**	**0.359**	0.753	**0.410**	0.759	**0.336**	0.371	**0.341**	0.767	0.781	**0.764**
	192	**0.722**	0.765	div.	**0.760**	**0.746**	0.765	**0.371**	0.765	div.	**0.437**	**0.754**	0.767
	336	div.	**0.768**	**0.445**	div.	**0.767**	0.767	**0.767**	0.768	**0.767**	0.768	**0.769**	0.773
	720	0.774	**0.774**	div.	div.	**0.776**	div.	0.776	**0.775**	0.775	**0.772**	div.	div.
Solar	96	**0.321**	div.	div.	**0.422**	div.	div.	**0.310**	0.340	0.634	**0.465**	div.	div.
	192	**0.668**	0.697	0.663	**0.642**	div.	div.	div.	div.	**0.470**	0.693	div.	**0.691**
	336	**0.510**	div.	**0.503**	0.689	**0.668**	0.698	div.	**0.674**	**0.668**	0.691	**0.696**	div.
	720	0.696	**0.693**	**0.675**	0.683	div.	**0.692**	0.699	**0.684**	0.687	**0.680**	div.	**0.692**
Traffic	96	0.768	**0.532**	**0.737**	0.790	**0.716**	0.807	0.928	**0.812**	**0.429**	0.532	**0.520**	0.796
	192	**0.769**	0.795	0.776	**0.761**	**0.538**	0.851	**0.796**	0.809	**0.775**	0.782	**0.535**	0.796
	336	**0.800**	div.	**0.787**	0.820	**0.811**	div.	div.	**0.806**	**0.793**	0.806	**0.794**	0.810
	720	**0.795**	0.809	**0.819**	div.	div.	div.	0.834	**0.807**	**0.850**	div.	**0.805**	0.807

Table 27, continued.

Dataset	H	2021		2022		2023		2024		2025		2026	
		OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM	OSSM	SSM
ETTh1	96	**0.448**	0.503	**0.478**	0.586	**0.467**	0.574	**0.451**	0.576	**0.503**	0.534	**0.453**	0.575
	192	**0.517**	0.581	**0.506**	0.589	**0.584**	0.589	**0.546**	0.578	**0.545**	0.591	0.582	**0.511**
	336	**0.585**	div.	div.	**0.595**	**0.597**	0.598	**0.550**	0.583	**0.592**	div.	0.595	**0.585**
	720	**0.632**	div.	**0.641**	div.	**0.639**	0.639	div.	div.	0.638	**0.619**	0.638	**0.629**
ETTm1	96	**0.465**	0.524	0.532	**0.463**	0.506	**0.470**	**0.470**	0.534	0.534	**0.529**	**0.422**	0.457
	192	**0.466**	div.	**0.478**	0.480	**0.468**	0.556	**0.481**	0.547	div.	**0.530**	div.	**0.475**
	336	**0.550**	div.	**0.548**	0.558	0.550	**0.505**	**0.483**	div.	0.560	**0.488**	div.	**0.556**
	720	**0.568**	0.572	**0.622**	div.	**0.565**	div.	**0.503**	div.	**0.570**	0.575	div.	**0.580**
Wind	96	**0.657**	0.824	**0.711**	0.800	0.680	**0.661**	**0.692**	0.829	**0.692**	0.823	**0.673**	div.
	192	**0.781**	0.852	**0.768**	0.857	**0.780**	0.784	**0.772**	0.858	**0.770**	0.860	**0.816**	div.
	336	0.878	**0.877**	0.826	**0.822**	**0.859**	0.884	**0.879**	div.	**0.814**	0.880	**0.824**	div.
	720	div.	div.	div.	div.	**0.884**	div.	div.	div.	**0.891**	0.895	**0.892**	div.
ETTm2	96	div.	**0.333**	**0.335**	0.358	**0.332**	0.358	0.331	**0.326**	**0.349**	0.360	**0.332**	0.341
	192	div.	**0.533**	0.384	**0.372**	0.381	**0.370**	**0.371**	0.371	**0.359**	0.374	**0.367**	0.369
	336	**0.393**	div.	**0.389**	0.390	0.391	**0.388**	**0.389**	0.391	**0.392**	div.	div.	div.
	720	div.	**0.430**	**0.432**	div.	**0.429**	div.	**0.431**	0.433	**0.430**	div.	0.430	**0.430**
Exchange	96	0.511	**0.373**	0.509	**0.412**	**0.392**	0.503	**0.458**	0.516	**0.357**	0.420	**0.416**	0.507
	192	**0.568**	div.	**0.589**	0.590	**0.559**	0.574	0.576	**0.568**	**0.589**	div.	0.578	**0.571**
	336	**0.689**	div.	**0.675**	div.	**0.670**	div.	0.676	**0.670**	div.	div.	**0.689**	div.
	720	0.884	**0.866**	div.	**0.865**	**0.871**	div.	div.	**0.864**	**0.862**	div.	**0.880**	div.
ETTh2	96	0.421	**0.419**	**0.423**	0.426	0.420	**0.420**	**0.422**	0.427	**0.429**	0.439	0.419	**0.415**
	192	0.428	**0.424**	**0.432**	div.	0.431	**0.424**	0.463	**0.425**	**0.432**	div.	0.504	**0.423**
	336	**0.430**	0.444	0.438	**0.429**	**0.436**	div.	**0.438**	div.	**0.438**	div.	0.447	**0.427**
	720	0.464	**0.459**	**0.458**	div.	div.	div.	div.	div.	**0.457**	div.	div.	div.
Weather	96	0.240	**0.239**	**0.243**	0.262	**0.247**	0.317	**0.246**	div.	0.320	**0.319**	**0.253**	0.261
	192	**0.292**	0.332	0.323	**0.295**	**0.301**	0.334	**0.278**	0.285	**0.296**	0.300	**0.284**	0.303
	336	**0.333**	0.394	**0.351**	0.354	**0.318**	0.322	div.	div.	div.	**0.349**	div.	**0.352**
	720	div.	div.	div.	div.	div.	**0.387**	0.384	**0.378**	**0.365**	div.	0.387	**0.378**