Optimal Look-back Horizon for Time Series Forecasting in Federated Learning

Dahao Tang[1], Nan Yang[1,3], Yanli Li[1], Zhiyu Zhu[2], Zhibo Jin[2], Dong Yuan[1,3]
[1]The University of Sydney
[2]University of Technology Sydney
[3]Corresponding authors: n.yang@sydney.edu.au, dong.yuan@sydney.edu.au

Association for the Advancement of Artificial Intelligence

THE UNIVERSITY OF SYDNEY

UTS UNIVERSITY OF TECHNOLOGY SYDNEY

Abstract

Selecting an appropriate look-back horizon remains a fundamental challenge in time series forecasting (TSF), particularly in federated learning scenarios where data is decentralized, heterogeneous, and often non-independent. While recent work has explored horizon selection by preserving forecasting-relevant information in an intrinsic space, these approaches are primarily restricted to centralized and independently distributed settings. This paper presents a principled framework for adaptive horizon selection in federated time series forecasting through an intrinsic space formulation. We introduce a synthetic data generator (SDG) that captures essential temporal structures in client data, including autoregressive dependencies, seasonality, and trend, while incorporating client-specific heterogeneity. Building on this model, we define a transformation that maps time series windows into an intrinsic representation space with well-defined geometric and statistical properties. We then derive a decomposition of the forecasting loss into a Bayesian term (irreducible uncertainty) and an approximation term (finite-sample effects and limited model capacity). Our analysis shows that while increasing the look-back horizon improves identifiability, it also increases approximation error. We prove that the total forecasting loss is minimized at the smallest horizon where the irreducible loss starts to saturate.

Introduction

Selecting the right look-back horizon is critical for time series forecasting, yet existing scaling-law insights assume centralized, IID data and fail under the heterogeneity of federated learning. In decentralized settings, clients differ in dynamics, noise, and sequence structure, making a fixed global horizon suboptimal. We introduce a principled framework that uses a structured Synthetic Data Generator to capture shared temporal patterns and client-specific variability.
Our contributions include:

- A geometry-preserving intrinsic space for heterogeneous multivariate series.
- A tight decomposition linking horizon length to Bayesian and approximation errors.
- A proof that forecasting loss is unimodal in horizon size, yielding a client-adaptive optimal horizon criterion.

Synthetic Data Generator (SDG)

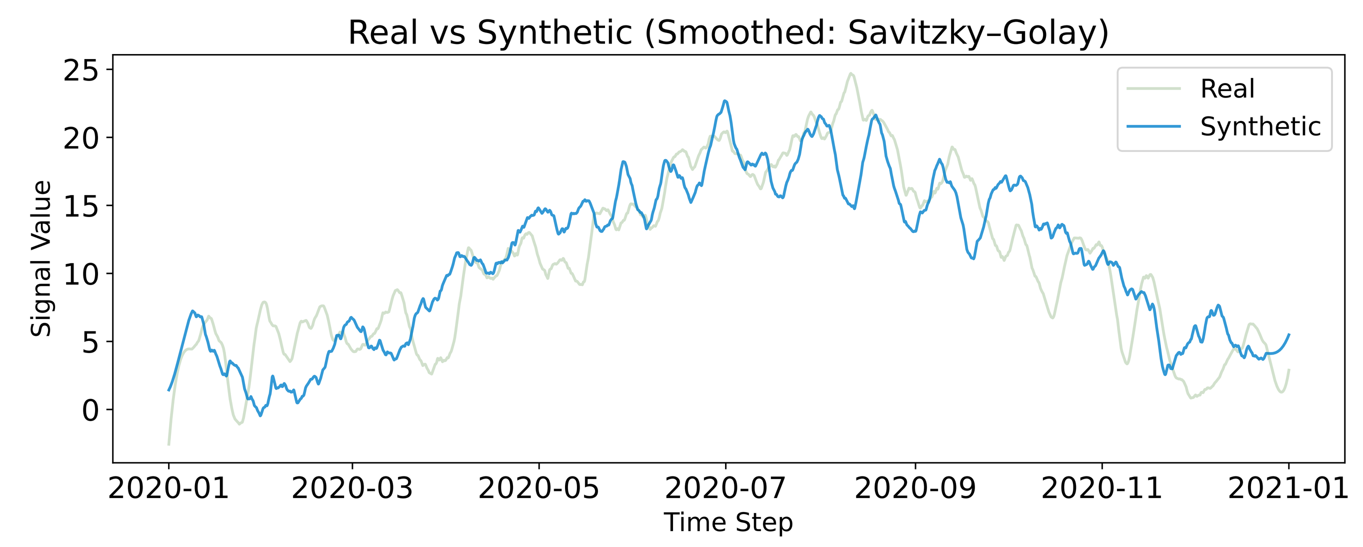

Figure: Comparison between real-world data and data generated by the SDG. The close alignment indicates that the SDG effectively captures the patterns present in real data.

An SDG is a parametric model designed to simulate univariate time series data characterized by seasonality, temporal dependence (AR memory), and trend. For a given client k, feature f, and time step t, the observation $\hat{x}_{f,t,k}$ is:

$$\begin{aligned}\hat{x}_{f,t,k} &= \mathrm{Seasonal}(A_{f,j,k}, T_{f,j,k}, \Theta_{f,j,k}) + \mathrm{AR}_{p,k}(\phi_k) \\ &\quad + \mathrm{Trend}(\beta_{f,k}) + \epsilon_{f,t,k} \\ &= \sum_{j=1}^{J} A_{f,j,k}\cdot \sin\left(\frac{2\pi t}{T_{f,j,k}} + \theta_{f,j,k}\right) \\ &\quad + \sum_{i=1}^{p} \phi_{k,i}\, x_{f,t-i,k} + \beta_{f,k}\, t + \epsilon_{f,t,k}.\end{aligned} \tag{1}$$

Feature Skewness Formulation

In federated learning, each client observes a different distribution of the same features (feature skew). We model this heterogeneity as:

$$x_{f,t,k} = \Lambda_{f,k}\hat{x}_{f,t,k} + \delta_{f,k} \tag{2}$$

where Λ_{fk} is the linear scale (controlling variance σ_f^2) and δ_{fk} is the mean shift for client k.

Intrinsic Space Construction

We construct a geometry-aware representation space that captures the essential temporal structure of non-IID time series. The transformation pipeline involves: (1) Client-wise normalization; (2) Window flattening; (3) Global covariance estimation; (4) Intrinsic dimension estimation; and (5) Projection.
The intrinsic dimension for client k is approximated as:

$$d_{I,k}(H) \approx F\cdot\left(\min\{H, \ell_{\mathrm{AR},k}\} + g_k(H) + 1\right). \tag{3}$$

Here, $\ell_{\mathrm{AR},k}$ denotes the effective AR memory:

$$\ell_{\mathrm{AR},k} = \left\lceil \frac{\ln(1/(1-\epsilon))}{-\ln \rho_k} \right\rceil, \quad \epsilon \in (0,1) \tag{4}$$

where $\rho_k \in (0,1)$ is the spectral radius of the AR companion matrix. $g_k(H)$ reflects the resolved seasonal complexity:

$$g_k(H) = 2\sum_{j=1}^{J} w_{j,k}\cdot \min\left(1, \frac{H}{T^*_{j,k}}\right), \tag{5}$$

$$w_{j,k} = \frac{\sum_{f=1}^{F} A_{f,j,k}^2}{\sum_{f=1}^{F}\sum_{j=1}^{J} A_{f,j,k}^2}. \tag{6}$$

Loss Decomposition

We now formalize this precise decomposition of the prediction loss in the federated setting, showing how the Bayesian and approximation components arise directly from the client-specific data-generating distributions and the server-side evaluation protocol.

Theorem (Federated Loss Decomposition)

For each client $k \in \{1,\dots,K\}$, let (U_k, V_k) denote its data-generating pair, where U_k takes values in a measurable input space $\mathcal{M}(H)$ and V_k in an output space $\mathcal{M}(S)$, both embedded in a real Hilbert space $(\mathcal{H}, \|\cdot\|)$ with the associated Borel σ-algebras.
Let $m_k^*(u) := \mathbb{E}[\, V_k \mid U_k = u\,]$ be the client-specific Bayesian predictor, defined P_{U_k}–almost everywhere. For any measurable, square-integrable predictor $m : \mathcal{M}(H) \to \mathcal{M}(S)$, the server's global predictive loss is

$$L(H,S;m) := \mathbb{E}_{k\sim\pi}\left[\mathbb{E}\left[\,\|V_k - m(U_k)\|^2\right]\right] \tag{7}$$

where $\pi = (\pi_1,\dots,\pi_K)$ is any distribution over clients and the inner expectation is over (U_k,V_k) under client k's distribution. Then the loss decomposes as:

$$L(H,S;m) = L_{\mathrm{Bayes}}(H,S) + L_{\mathrm{approx}}(H,S;m), \tag{8}$$

where the federated Bayesian loss is

$$L_{\mathrm{Bayes}}(H,S) := \mathbb{E}_{k\sim\pi}\left[\mathbb{E}\left[\,\|V_k - m_k^*(U_k)\|^2\right]\right], \tag{9}$$

and the federated approximation loss is

$$L_{\mathrm{approx}}(H,S;m) := \mathbb{E}_{k\sim\pi}\left[\mathbb{E}\left[\,\|m_k^*(U_k) - m(U_k)\|^2\right]\right]. \tag{10}$$

In particular, the total loss separates into the expected irreducible (client-wise Bayes) component and the expected approximation error of the global predictor relative to each client's Bayes-optimal rule. Please refer to the Extended Version for the proof.

Smallest Sufficient Horizon

Now we define a key concept, the smallest sufficient horizon, which serves as the optimal look-back horizon that minimizes the forecasting loss.
Formaly, for any tolerance $\delta > 0$, define the smallest sufficient horizon as

$$H_k^*(\delta) := \min\{H : |\Delta L_{\mathrm{Bayes}}^{(k)}(H)| \le \delta\}, \tag{11}$$

at which the Bayesian loss has effectively saturated: further historical context improves the irreducible loss by at most δ. Together, these monotonicity properties imply a unimodal structure for the total loss.

Unimodality and Optimal Horizon

If for a given $\delta > 0$ the Bayesian loss satisfies $\Delta L_{\mathrm{Bayes}}^{(k)}(H) \le -\delta$ for all $H < H_k^*(\delta)$, and the approximation loss satisfies $\Delta L_{\mathrm{approx}}^{(k)}(H;m) \ge \delta$ for all $H \ge H_k^*(\delta)$, then the combined loss obeys that $L^{(k)}(H)$ decreases on $[1, H_k^*(\delta)]$, and $L^{(k)}(H)$ increases on $[H_k^*(\delta), \infty)$.
Consequently, $H_k^*(\delta) \in \arg\min_{H\in\mathbb{N}} L^{(k)}(H)$ with uniqueness up to integer ties.

Proof.

From the Bayesian loss analysis, increasing H reduces seasonal/phase ambiguity and uncovers AR structure, but only up to a finite coverage horizon. Hence, there exists H_0 such that

$$\Delta L_{\mathrm{Bayes}}(H,S) < 0 \quad (H < H_0), \tag{12}$$

while for any $\delta > 0$ we can choose H_0 large enough so that

$$\Delta L_{\mathrm{Bayes}}(H,S) \ge -\delta \quad (H \ge H_0). \tag{13}$$

For the approximation term, the curvature–variance bound on the intrinsic manifold shows that the error grows with both the intrinsic dimension $d_I(H)$ and the factor H/D coming from the effective sample size per window ($\propto D/(HN)$). Since $d_I(H)$ is non-decreasing and eventually saturated, while H/D grows linearly, there exists $\eta > 0$, independent of H, such that

$$\Delta L_{\mathrm{approx}}(H,S) \ge \eta \quad (H \ge H_0). \tag{14}$$

Fix any $\delta \in (0,\eta)$ and define $H^*(\delta)$ as the smallest $H \ge H_0$ with $\Delta L_{\mathrm{Bayes}}(H,S) \ge -\delta$. Then for $H < H^*(\delta)$, we have $\Delta L_{\mathrm{Bayes}}(H,S) < -\delta$ and $\Delta L_{\mathrm{approx}}(H,S) \ge 0$, so $\Delta L(H,S) = \Delta L_{\mathrm{Bayes}}(H,S) + \Delta L_{\mathrm{approx}}(H,S) < -\delta < 0$, and $L(H,S)$ is strictly decreasing. For $H \ge H^*(\delta)$, we have $\Delta L_{\mathrm{Bayes}}(H,S) \ge -\delta$ and $\Delta L_{\mathrm{approx}}(H,S) \ge \eta$, hence $\Delta L(H,S) \ge -\delta + \eta > 0$, so $L(H,S)$ is strictly increasing.
Thus $L(H,S)$ decreases up to $H^*(\delta)$ and increases thereafter, so it is unimodal in H and attains its unique minimum at $H^*(\delta)$ (up to trivial ties), as claimed. $\square$

Hence, before $H_k^*(\delta)$, the reduction in irreducible error outweighs the increase in approximation error; afterwards, the opposite holds. The total loss thus has a single optimal basin, and the smallest sufficient horizon attains the minimum.

Limitation and Discussion

The SDG captures key components, but assumes Gaussian innovations, local stationarity, and stable AR structure, limiting its ability to represent regime shifts, nonlinear patterns, or strong feature interactions. Estimating global covariance also requires privacy-aware aggregation, and treating overlapping windows as independent may overstate sample size.

These assumptions, while standard in theory, are chosen to clearly isolate the effects of horizon length and heterogeneity, providing the first provable foundation for optimal horizon selection and a basis for future extensions.

dahao.tang@sydney.edu.au