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

Dahao Tang[1], Nan Yang[1], Yanli Li[1], Zhiyu Zhu[2], Zhibo Jin[2], Dong Yuan[1]
[1]The University of Sydney [2]University of Technology Sydney
dahao.tang@sydney.edu.au, n.yang@sydney.edu.au, ...

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)

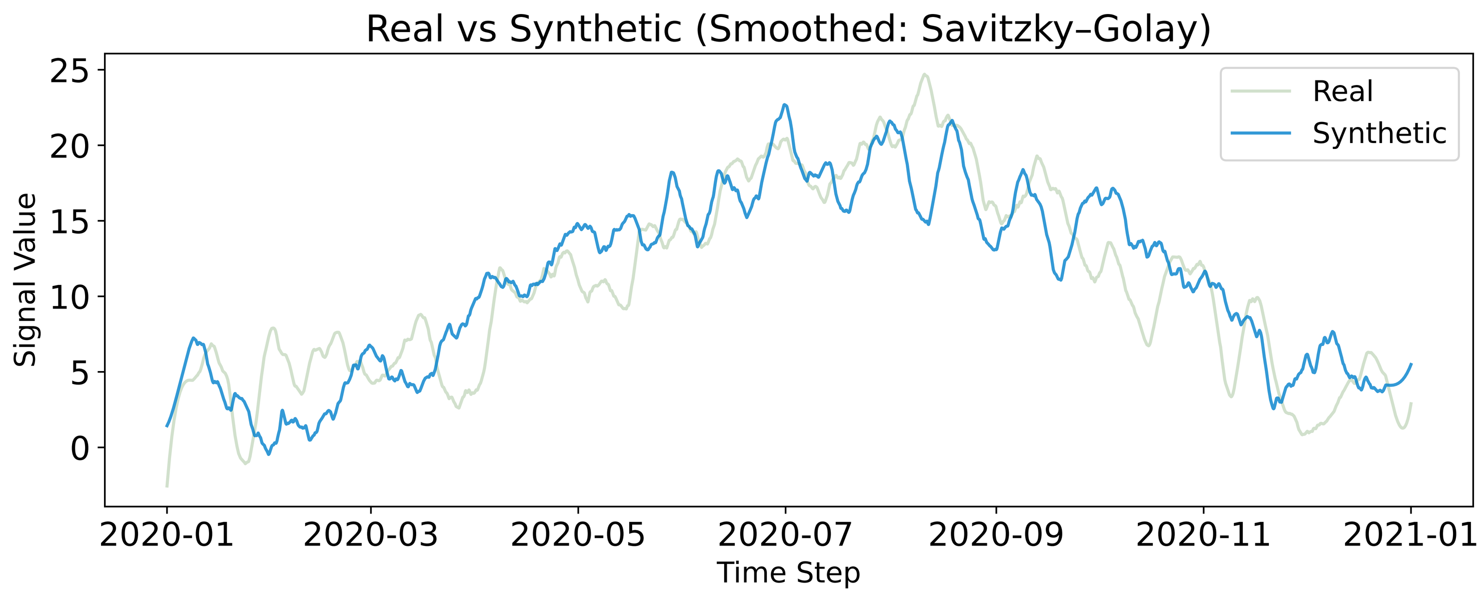

Figure: Comparison between real-world data and data generated by the SDG. The close alignment indicates that the SDG effectively captures the patterns present in real data.

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:

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

where $\rho_k \in (0,1)$ is the spectral radius of the AR companion matrix.

Loss Decomposition

The global prediction loss decomposes as:

$$L(H) = L_{\mathrm{Bayes}}(H) + L_{\mathrm{approx}}(H),$$

separating irreducible client uncertainty from approximation error induced by finite samples and model capacity.

Smallest Sufficient Horizon

For tolerance $\delta > 0$, define:

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

This is the smallest horizon at which additional history yields negligible irreducible error reduction.

Conclusion

The total forecasting loss is unimodal in horizon length. The smallest sufficient horizon minimizes loss by balancing information identifiability against approximation error growth, yielding a principled, client-adaptive horizon selection rule.