Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Bounded affine barycentric-coordinate prediction for improving local extreme value forecasting

Naoki Fujiwara

s2620660@u.tsukuba.ac.jp

University of Tsukuba

Yoshito Hirata

University of Tsukuba

Tohru Kawabe

University of Tsukuba

Research Article

Keywords:

Posted Date: November 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11587794/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Bounded affine barycentric-coordinate prediction for improving local extreme value forecasting

Naoki Fujiwara[1*], Yoshito Hirata[1,2] and Tohru Kawabe[1,2]

[1*]Degree Programs in Systems and Information Engineering, Graduate School of Science and Technology, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan.
[2]Institute of Systems and Information Engineering, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan.

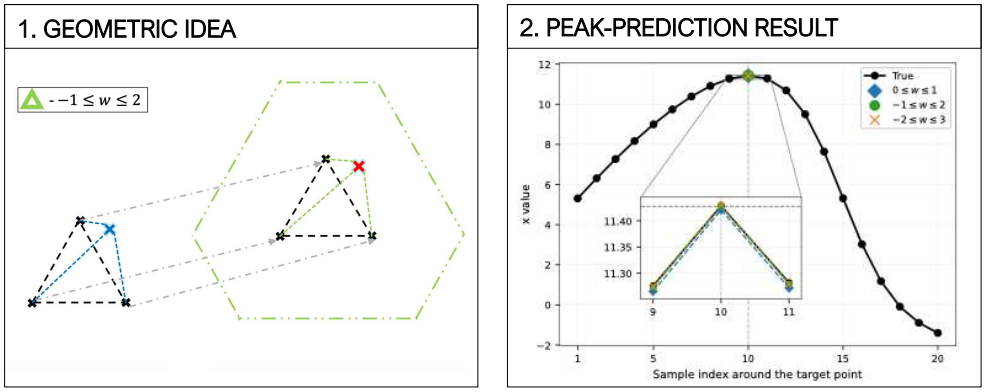
*Corresponding author(s). E-mail(s): s2620660@u.tsukuba.ac.jp;
Contributing authors: hirata@cs.tsukuba.ac.jp;

Abstract

Time series data from complex systems, such as floods and economic systems, are nonlinear. Predicting local maxima and minima is important in real-world applications. However, many local state-space prediction methods represent future states as convex combinations of neighboring observations. As a consequence, predictions are effectively restricted to interpolation within the local neighborhood, which can lead to the underestimation of local extreme values. We propose an extrapolative method of time-series prediction based on barycentric coordinates. We apply the proposed method to two chaotic toy models and to an exploratory benchmark using real-world flood data. We evaluate its performance in both recursive and direct prediction. For the clean Rössler and Lorenz'63 data, the affine method yielded positive median improvements in recursive local-extreme prediction, although the mean improvement was often negative. Under observation noise, most mean and median improvements were negative, and direct prediction was generally unstable. These results indicate that the benefit of bounded affine extrapolation is restricted to selected clean-data and short- to moderate-horizon conditions.

Keywords: Time series prediction, local extreme values, extrapolation, barycentric coordinates

Graphical abstract



1 Introduction

Predicting maxima is important in many fields. As an example from hydrology, we consider small and medium-sized rivers. These rivers respond quickly to rainfall because their drainage areas are small. Their short response times make the rivers prone to sudden flooding. For this reason, we aim to accurately predict the maximum water level in such rivers for disaster prevention.

Extreme events are rare but important targets in time-series prediction. Ding et al. [1] pointed out that standard prediction models can miss such events. They introduced an extreme-value loss and a memory-network module to improve the prediction of extreme events.

However, predicting local maxima remains difficult. Paleologos et al. [2] simulated the daily discharges of two karst springs using neural networks. Their predictions generally tracked the discharge well. However, some configurations showed larger errors at a few local maxima. This suggests that local maxima should be evaluated separately.

Several studies have directly addressed the prediction of local maxima. Hirata and Shiro [3] improved the prediction accuracy for maxima of a flow by using local cross sections. Their method focuses on local maxima by reconstructing the dynamics on sections crossed by the flow. Liu and Hirata [4] treated local maxima as an event series. They used an edit distance for marked point processes as an additional input to a neural network. The edit distance was computed using the efficient exact method developed by Hirata and Sukegawa [5].

These studies show that local maxima require special treatment in time-series prediction. Nevertheless, conventional local prediction methods are often limited to interpolation among neighboring states. As a result, they may not adequately represent future states that exceed the range suggested by nearby observations. This limitation can cause the underestimation of local extreme values.

To overcome this limitation, we introduce a new prediction method that allows extrapolation. We extend a local linear approximation method based on barycentric coordinates and linear programming so that it can perform extrapolation as well as interpolation. In the conventional method, a prediction is represented as a convex

combination of neighboring points by constraining the weights to be nonnegative and to sum to one. In our method, we allow negative weights, thereby extending the representation from a convex combination to an affine combination. This extension enables bounded extrapolation and helps prevent the underestimation of extreme values.

2 Background

Hirata et al. [6] proposed a method for constructing barycentric coordinates using linear programming. The method aims at approximating and predicting high-dimensional dynamical systems. Here, we summarize the work of Hirata et al. [6], based on which we propose the current work.

The method of Hirata et al. [6] extends the barycentric coordinate method proposed by Mees [7]. The method of Mees [7] is designed for low-dimensional systems. The original approach uses geometric partitioning (tessellation). However, geometric partitioning is difficult to apply in high-dimensional spaces. This limitation is overcome by formulating the problem as a linear programming problem [6].

The method is formulated as follows. Let M be a compact manifold, and let

$$f : M \to M$$

be a smooth dynamical system. For an initial state $\xi_0 \in M$, the orbit is given by

$$\xi_n = f^n(\xi_0), \qquad n \in \mathbb{Z}_{\geq 0},$$

where f^n denotes the n-fold composition of f. Let

$$g : M \to \mathbb{R}$$

be a smooth observation function. Then, from the orbit $\{\xi_n\}$, we observe a univariate time series

$$x_n = g(\xi_n), \qquad n \in \mathbb{Z}_{\geq 0}.$$

Using Takens' embedding theorem [8], we reconstruct the state space from the observed scalar time series. Let $m \in \mathbb{N}$ be the embedding dimension and $\tau \in \mathbb{N}$ be the delay. We define the delay-coordinate map

$$G : M \to \mathbb{R}^m$$

by

$$G(\xi) = \big(g(\xi), g(f^\tau(\xi)), \ldots, g(f^{(m-1)\tau}(\xi))\big)^{\mathsf{T}}. \tag{1}$$

Accordingly, for each time index n, the reconstructed vector is written as

$$\mathbf{V}_n = G(\xi_n) = [x_n, x_{n+\tau}, \ldots, x_{n+(m-1)\tau}]^{\mathsf{T}} \in \mathbb{R}^m. \tag{2}$$

Under the assumptions of Takens' embedding theorem, if $m \geq 2d + 1$, where $d = \dim M$, and if g is generic, then G is an embedding. Hence, G is a diffeomorphism

between M and its image $G(M) \subset \mathbb{R}^m$. Writing $F = G \circ f \circ G^{-1}$ for the reconstructed one-step map, the conjugacy gives

$$F^\tau = G \circ f^\tau \circ G^{-1} : G(M) \to G(M). \tag{3}$$

This equation implies that

$$\mathbf{V}_{n+\tau} = F^\tau(\mathbf{V}_n). \tag{4}$$

Based on this formulation, prediction is performed by approximating the dynamics locally in the reconstructed space. Given a target state $\mathbf{V}_n \in \mathbb{R}^m$, let $K_n = \{n_1, n_2, \ldots, n_k\}$ denote the index set of the k nearest neighbors of $\mathbf{V}_n$ in the reconstructed space. The neighboring states $\mathbf{V}_i$ are sorted in ascending order of the Euclidean distances

$$d_i = \|\mathbf{V}_n - \mathbf{V}_i\|, \qquad i \in K_n. \tag{5}$$

Hirata et al. [6] introduce barycentric coordinates for $\mathbf{V}_n$ and its neighboring states $\{\mathbf{V}_i\}_{i \in K_n}$. The weights $\{w_i\}_{i \in K_n}$ are determined using linear programming. In the original barycentric coordinate method, the weights satisfy

$$\mathbf{V}_n \approx \sum_{i \in K_n} w_i \mathbf{V}_i, \qquad \sum_{i \in K_n} w_i = 1, \qquad w_{\min} \le w_i \le w_{\max} \quad (i \in K_n). \tag{6}$$

In the conventional setting,

$$w_{\min} = 0, \text{ and } w_{\max} = 1, \tag{7}$$

so that the approximation is given by a convex combination.

Then, the p-step-ahead state $\hat{\mathbf{V}}_{n+p}$ is predicted as

$$\hat{\mathbf{V}}_{n+p} = \sum_{i \in K_n} w_i \mathbf{V}_{i+p}. \tag{8}$$

Figure 1 illustrates the method. In Figure 1, the blue crosses represent past points. The black crosses forming a triangle represent the nearest neighbors. The red cross represents the predicted point. The black crosses forming another triangle represent the successor points of the nearest neighbors. The constraint $0 \le w_i \le 1$ restricts the prediction region. The predicted point remains inside the convex hull of the successor points.

3 Method

3.1 Conventional barycentric-coordinate prediction and its limitation

The conventional method in Section 2 represents the target state as a convex combination of its neighbors. Denoting the convex weights by $\hat{w}_i$, the conventional constraint

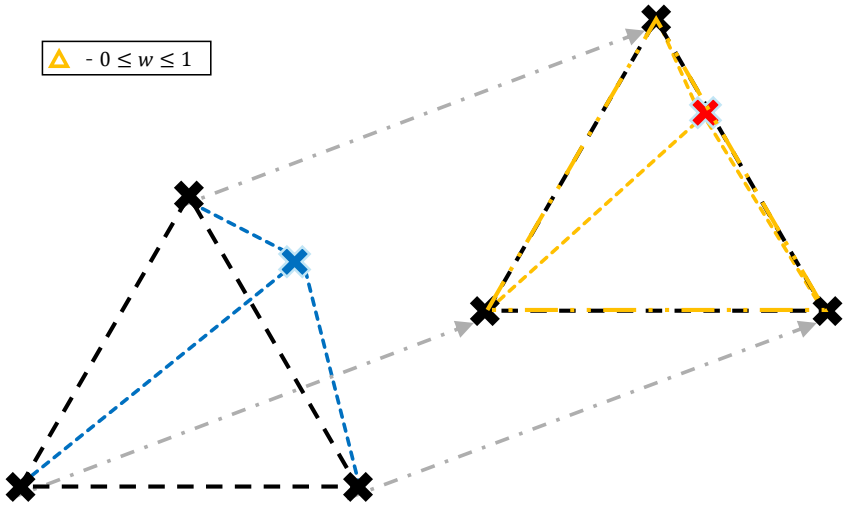


Fig. 1 An example of local linear prediction based on barycentric coordinates with convex-combination constraints

in Eq. (7) reads

$$0 \leq \hat{w}_i \leq 1, \qquad \sum_{i \in K_n} \hat{w}_i = 1, \tag{9}$$

so that the p-step-ahead prediction $\hat{\mathbf{V}}_{n+p} = \sum_{i \in K_n} \hat{w}_i \mathbf{V}_{i+p}$ in Eq. (8) is confined to the convex hull of the successor points $\{\mathbf{V}_{i+p}\}_{i \in K_n}$.

This convex-hull restriction is the limitation that the proposed method addresses. It prevents local predictions from representing successor states outside the hull of neighboring observations, which is problematic when local extreme values require extrapolation. The geometric reason for this limitation is discussed in Subsection 3.4; here, we relax the convex-hull restriction while keeping the prediction bounded.

3.2 Affine barycentric-coordinate prediction

To overcome the convex-hull restriction, we relax the constraints on the weights in Eq. (7) by choosing $w_{\min} < 0$ and $w_{\max} > 1$. We allow negative weights and weights larger than one while keeping the sum-to-one constraint, which turns the convex combination into an affine combination. Specifically, the affine weights $\tilde{w}_i$ satisfy

$$w_{\min} \leq \tilde{w}_i \leq w_{\max}, \qquad \sum_{i \in K_n} \tilde{w}_i = 1, \tag{10}$$

and the prediction is given by

$$\tilde{\mathbf{V}}_{n+p} = \sum_{i \in K_n} \tilde{w}_i \mathbf{V}_{i+p}. \tag{11}$$

Because the weights are only required to satisfy the sum-to-one constraint together with finite bounds, the prediction is no longer restricted to the convex hull and can extend beyond it within the prescribed affine-weight range. This enables bounded extrapolation beyond the neighboring observations.

Write $\hat{f}_p$ and $\tilde{f}_p$ for the convex and affine predictors at forecast lead p. For the Taylor analysis, assume a differentiable local extension of F^p near $\mathbf{V}_n$ in $\mathbb{R}^m$, denoted by the same symbol. Here, $(F^p)'(\mathbf{V}_n)$ denotes the Jacobian matrix of F^p at $\mathbf{V}_n$. For each neighbor $\mathbf{V}_i$ with $i \in K_n$, we have

$$F^p(\mathbf{V}_i) \approx F^p(\mathbf{V}_n) + (F^p)'(\mathbf{V}_n)(\mathbf{V}_i - \mathbf{V}_n), \qquad i \in K_n. \tag{12}$$

By multiplying both sides of Eq. (12) by $\tilde{w}_i$, we obtain

$$\tilde{w}_i F^p(\mathbf{V}_i) \approx \tilde{w}_i \left[F^p(\mathbf{V}_n) + (F^p)'(\mathbf{V}_n)(\mathbf{V}_i - \mathbf{V}_n)\right]. \tag{13}$$

Summing over $i \in K_n$, we obtain

$$\begin{aligned}\sum_{i\in K_n} \tilde{w}_i F^p(\mathbf{V}_i) &\approx \sum_{i\in K_n} \tilde{w}_i \left[F^p(\mathbf{V}_n) + (F^p)'(\mathbf{V}_n)(\mathbf{V}_i - \mathbf{V}_n)\right] \\ &= \left(\sum_{i\in K_n} \tilde{w}_i\right) F^p(\mathbf{V}_n) + (F^p)'(\mathbf{V}_n) \sum_{i\in K_n} \tilde{w}_i(\mathbf{V}_i - \mathbf{V}_n). \end{aligned} \tag{14}$$

Using the constraint

$$\sum_{i\in K_n} \tilde{w}_i = 1,$$

we obtain

$$\sum_{i\in K_n} \tilde{w}_i F^p(\mathbf{V}_i) \approx F^p(\mathbf{V}_n) + (F^p)'(\mathbf{V}_n) \left(\sum_{i\in K_n} \tilde{w}_i \mathbf{V}_i - \mathbf{V}_n\right). \tag{15}$$

Therefore, the approximation formula for the prediction is given by

$$\tilde{f}_p(\mathbf{V}_n) \approx F^p(\mathbf{V}_n) + (F^p)'(\mathbf{V}_n) \left(\sum_{i\in K_n} \tilde{w}_i \mathbf{V}_i - \mathbf{V}_n\right). \tag{16}$$

This formula means that the approximation $\tilde{f}_p(\mathbf{V}_n)$ can be written by the original value $F^p(\mathbf{V}_n)$ added to the difference $(F^p)'(\mathbf{V}_n)\left(\sum_{i\in K_n} \tilde{w}_i \mathbf{V}_i - \mathbf{V}_n\right)$ of the first order term.

Figure 2 illustrates this affine extrapolation for $w_{\min} = -1$ and $w_{\max} = 2$.

3.3 Boundedness of the affine prediction

Although affine weights allow extrapolation, the prediction stays bounded as long as the neighboring states and the weight range are bounded. Assume that the training

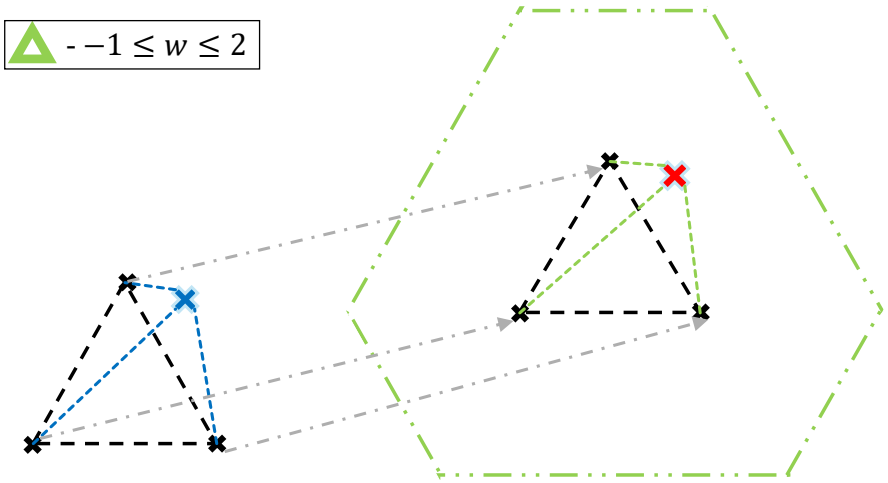


Fig. 2 An example of barycentric coordinates based on an affine combination

data, including all successor values used for prediction, are bounded by a common constant $M > 0$, and let $k = |K_n|$ be the number of neighboring points. Since the affine weights satisfy Eq. (10), there exists

$$W = \max\{|w_{\min}|, |w_{\max}|\} \tag{17}$$

such that $|\tilde{w}_i| \leq W$ for all $i \in K_n$. This yields the following boundedness property.

Remark 1 (Boundedness with bounded affine weights) Let $\tilde{\mathbf{V}} = \sum_{i \in K_n} \tilde{w}_i \mathbf{V}_i$ be an affine barycentric prediction. If $|\tilde{w}_i| \leq W$ for all i and $\|\mathbf{V}_i\| \leq M$ for all neighboring states, then

$$\|\tilde{\mathbf{V}}\| \leq kWM. \tag{18}$$

Therefore, the affine prediction remains bounded.

Proof By the triangle inequality and the homogeneity of the norm,

$$\left\| \sum_{i \in K_n} \tilde{w}_i \mathbf{V}_i \right\| \leq \sum_{i \in K_n} \|\tilde{w}_i \mathbf{V}_i\| = \sum_{i \in K_n} |\tilde{w}_i| \, \|\mathbf{V}_i\|. \tag{19}$$

Because $|\tilde{w}_i| \leq W$ and $\|\mathbf{V}_i\| \leq M$, we obtain

$$\left\| \sum_{i \in K_n} \tilde{w}_i \mathbf{V}_i \right\| \leq \sum_{i \in K_n} WM = kWM. \qquad \square \tag{20}$$

Although the bound in Eq. (18) is conservative, it guarantees that the affine prediction cannot become arbitrarily large when the training data and weights are bounded. This property motivates the use of bounded affine weight ranges.

Applying the same argument to the training successor values in Eq. (11) gives

$$\|\tilde{f}_p(\mathbf{V}_n)\| \leq \sum_{i \in K_n} |\tilde{w}_i| \, \|\mathbf{V}_{i+p}\| \leq kWM. \tag{21}$$

The scalar implementation satisfies the same bound using absolute values and scalar training responses. In recursive prediction, each step again combines one-step successor values from the fixed training data, so this bound applies at every step without requiring the recursive input to lie within the training range. This bounds prediction magnitudes; it does not guarantee forecast accuracy or stability to perturbations.

3.4 Geometric interpretation near attractor boundaries

We next explain geometrically why affine weights can help to predict local extreme values. Consider a target near a local maximum. In such a region, the true successor state may lie near the boundary of the convex hull formed by the successor points $\{\mathbf{V}_{i+p}\}_{i \in K_n}$, or outside the hull.

A convex combination can only reach points inside this convex hull. In particular, each coordinate of a convex prediction cannot exceed the largest corresponding coordinate of the successor points; if the true local maximum exceeds all of them, the convex prediction necessarily underestimates it. An affine combination can reach points in the affine hull outside the convex hull, subject to the weight bounds. The affine combination can thus place the prediction outside the convex hull and move the prediction closer to the true extreme value.

This geometric picture is the mechanism behind the proposed method: the convex-hull restriction can cause the underestimation of local extrema, whereas the affine relaxation allows the prediction to reach the boundary and exterior regions where extrema may occur.

3.5 Expected advantages and limitations

To quantify the advantages and limitations, we compare the prediction errors of the convex and affine predictions. Let $\hat{w}_i$ and $\tilde{w}_i$ denote the convex and affine weights satisfying Eqs. (9) and (10), respectively. Defining the local deviations

$$\mathbf{d}_{\hat{w}} = \sum_{i \in K_n} \hat{w}_i \mathbf{V}_i - \mathbf{V}_n, \qquad \mathbf{d}_{\tilde{w}} = \sum_{i \in K_n} \tilde{w}_i \mathbf{V}_i - \mathbf{V}_n, \tag{22}$$

we write the convex and affine predictions as $\hat{f}_p(\mathbf{V}_n) = \sum_{i \in K_n} \hat{w}_i F^p(\mathbf{V}_i)$ and $\tilde{f}_p(\mathbf{V}_n) = \sum_{i \in K_n} \tilde{w}_i F^p(\mathbf{V}_i)$, which coincide with $\sum_{i \in K_n} \hat{w}_i \mathbf{V}_{i+p}$ and $\sum_{i \in K_n} \tilde{w}_i \mathbf{V}_{i+p}$ for noise-free data. Throughout this comparison, $\|\cdot\|$ denotes the residual norm minimized in the weight computation (the L^1 norm in our implementation; see Subsection 4.1).

As shown in Eq. (16), which holds for any weights satisfying the sum-to-one constraint, the first-order terms of the two predictions are $(F^p)'(\mathbf{V}_n)\mathbf{d}_{\hat{w}}$ and $(F^p)'(\mathbf{V}_n)\mathbf{d}_{\tilde{w}}$. Retaining the Taylor remainders neglected there, the sum-to-one constraint gives, for $w \in \{\hat{w}, \tilde{w}\}$,

$$\sum_{i \in K_n} w_i F^p(\mathbf{V}_i) - F^p(\mathbf{V}_n) = \sum_{i \in K_n} w_i \big(F^p(\mathbf{V}_i) - F^p(\mathbf{V}_n)\big) = (F^p)'(\mathbf{V}_n)\mathbf{d}_w + \mathbf{R}_w, \quad (23)$$

where $\mathbf{R}_w = \sum_{i \in K_n} w_i \mathbf{r}_i$ and $\mathbf{r}_i = F^p(\mathbf{V}_i) - F^p(\mathbf{V}_n) - (F^p)'(\mathbf{V}_n)(\mathbf{V}_i - \mathbf{V}_n)$ is the Taylor remainder for the ith neighbor. For $w = \hat{w}$ and $w = \tilde{w}$, the left-hand side is the prediction error $\hat{f}_p(\mathbf{V}_n) - F^p(\mathbf{V}_n)$ and $\tilde{f}_p(\mathbf{V}_n) - F^p(\mathbf{V}_n)$, respectively. Since F^p is differentiable at $\mathbf{V}_n$, $\|\mathbf{r}_i\|/\|\mathbf{V}_i - \mathbf{V}_n\| \to 0$ as $\mathbf{V}_i \to \mathbf{V}_n$; that is, $\mathbf{r}_i$ is of higher order than $\|\mathbf{V}_i - \mathbf{V}_n\|$. Thus each $\mathbf{r}_i$ is small for close neighbors, but $\mathbf{R}_w$ is not controlled by $\mathbf{d}_w$: its bound $\|\mathbf{R}_w\| \le \sum_{i \in K_n} |w_i| \, \|\mathbf{r}_i\|$ grows with the weight magnitudes, and $\mathbf{R}_w$ may dominate the first-order term when cancellations in the weighted sum make $\mathbf{d}_w$ small.

Next, we evaluate how the first-order terms amplify the local deviations through the p-step map. Motivated by the expansion and contraction of nearby orbits described by Lyapunov exponents [9], we assume that there exist finite constants $\lambda_{\min}$ and $\lambda_{\max}$ bounding the finite-time amplification of the two local deviations as follows:

$$e^{\lambda_{\min} p}\|\mathbf{d}_{\hat{w}}\| \le \|(F^p)'(\mathbf{V}_n)\mathbf{d}_{\hat{w}}\|, \qquad \|(F^p)'(\mathbf{V}_n)\mathbf{d}_{\tilde{w}}\| \le e^{\lambda_{\max} p}\|\mathbf{d}_{\tilde{w}}\|. \quad (24)$$

Here, $\lambda_{\min}$ and $\lambda_{\max}$ represent effective lower and upper finite-time expansion rates in this norm. They may depend on the target state and the prediction step p and are not identified with the finite-lag divergence estimates used in Section 4. These finite-time bounds are assumptions for this comparison, rather than consequences of the asymptotic Lyapunov exponents.

To compare the local reconstruction residuals, we define

$$b = \frac{\|\mathbf{d}_{\hat{w}}\|}{\|\mathbf{d}_{\tilde{w}}\|}, \quad (25)$$

provided that $\|\mathbf{d}_{\tilde{w}}\| > 0$. Because $w_{\min} < 0$ and $w_{\max} > 1$, the affine feasible set contains the convex feasible set. When both formulations minimize this same residual norm for the same target and neighbor set, the optimal affine residual cannot exceed the optimal convex residual. Consequently, $b \ge 1$ whenever it is defined. Equality holds when a convex-optimal solution is also optimal for the affine formulation. If $\|\mathbf{d}_{\tilde{w}}\| = 0$, b is undefined and this case is excluded from the ratio comparison; a zero reconstruction residual does not imply zero prediction error.

Applying the triangle inequality to Eq. (23) and using Eq. (24), we obtain $\|\hat{f}_p(\mathbf{V}_n) - F^p(\mathbf{V}_n)\| \ge e^{\lambda_{\min} p}\|\mathbf{d}_{\hat{w}}\|(1-\hat{\eta})$ and $\|\tilde{f}_p(\mathbf{V}_n) - F^p(\mathbf{V}_n)\| \le e^{\lambda_{\max} p}\|\mathbf{d}_{\tilde{w}}\|(1+\tilde{\eta})$, where $\hat{\eta} = \|\mathbf{R}_{\hat{w}}\|/(e^{\lambda_{\min} p}\|\mathbf{d}_{\hat{w}}\|)$ and $\tilde{\eta} = \|\mathbf{R}_{\tilde{w}}\|/(e^{\lambda_{\max} p}\|\mathbf{d}_{\tilde{w}}\|)$ measure the remainders relative to the amplified local deviations. Hence, when b is defined, $\tilde{f}_p(\mathbf{V}_n) \ne F^p(\mathbf{V}_n)$,

and $\hat{\eta} < 1$,

$$\log \frac{\|\hat{f}_p(\mathbf{V}_n) - F^p(\mathbf{V}_n)\|}{\|\tilde{f}_p(\mathbf{V}_n) - F^p(\mathbf{V}_n)\|} \geq \log b - (\lambda_{\max} - \lambda_{\min})p - \kappa_{n,p}, \tag{26}$$

where $\kappa_{n,p} = \log\{(1+\tilde{\eta})/(1-\hat{\eta})\} \geq 0$. The first two terms on the right-hand side compare the reconstruction gain with the difference in amplification, and $\kappa_{n,p}$ quantifies how much the Taylor remainders can weaken this comparison; $\kappa_{n,p}$ approaches zero as $\hat{\eta}$ and $\tilde{\eta}$ approach zero. In particular, the bound ensures that the affine prediction error is smaller than the convex prediction error when $\log b > (\lambda_{\max} - \lambda_{\min})p + \kappa_{n,p}$; otherwise, it does not determine which prediction error is smaller. The condition $\hat{\eta} < 1$ is imposed only on the convex prediction; if it fails, the analysis provides no information on the error ratio.

Eq. (26) holds under the stated assumptions: noise-free successor states, the local linearization in Eq. (16), and the amplification rates in Eq. (24). We do not estimate $\lambda_{\min}$, $\lambda_{\max}$, or the remainders, nor do we verify these assumptions under the experimental conditions. Therefore, the bound clarifies the mechanism but does not guarantee which method yields the smaller error in the experiments; in particular, a decrease in the right-hand side of Eq. (26) does not imply a decrease in the actual ratio or in the prediction advantage. The bound also does not cover noisy data, the single coordinate evaluated in the experiments, or complete recursive prediction trajectories. With these limitations in mind, the expected effects on prediction performance can be summarized as follows.

- **Extrema effect.** The affine combination can be beneficial when the convex-hull constraint limits the representation of the target state or the prediction of its successor (Subsection 3.4). Such situations may occur near local extrema, but proximity to a local extremum alone does not guarantee improved prediction accuracy.
- **Horizon effect.** For fixed b, amplification rates, and $\kappa_{n,p}$, the right-hand side of Eq. (26) decreases with p when $\lambda_{\max} > \lambda_{\min}$. In practice, these quantities may also depend on p; in particular, $\kappa_{n,p}$ may increase with p as the nonlinearity of F^p grows, and the condition $\hat{\eta} < 1$ may eventually fail. Intuitively, local approximation errors can be amplified through the nonlinear dynamics over successive prediction steps, consistent with the sensitivity to initial conditions in chaotic systems.
- **Weight-range effect.** A wider admissible range increases W and, by Remark 1, the bound kWM. This enhances the extrapolation capability but may also make the predictions more sensitive to perturbations. A wider range also permits a larger $\sum_{i \in K_n} |\tilde{w}_i|$, which can increase $\|\mathbf{R}_{\tilde{w}}\|$ and hence $\kappa_{n,p}$. Therefore, a wider range does not necessarily improve prediction accuracy; its effect depends on the local geometry of the reconstructed state space and on the data used to identify the weights.
- **Noise effect.** Observation noise can perturb the local geometry, neighbor selection, and estimated weights, potentially reducing the benefit of extrapolation. Wider weight ranges permit larger perturbation amplification, but the actual effect depends on the selected weights and local data.

These considerations suggest that bounded affine barycentric-coordinate prediction is not uniformly superior to convex-combination-based prediction; the effects of the evaluation horizon, prediction scheme, weight range, and observation noise are evaluated experimentally in Section 4.

4 Experiments

We evaluate the proposed method using two toy models and real flood-related data.

In the experiments, we test two affine ranges, $-1 \leq \tilde{w}_i \leq 2$ and $-2 \leq \tilde{w}_i \leq 3$. Both ranges are constructed by extending the convex interval $[0, 1]$ symmetrically by a fixed offset δ: the lower bound becomes $-\delta$ and the upper bound becomes $1 + \delta$, with $\delta = 1$ and $\delta = 2$, respectively. This symmetric offset parameterization keeps the sum-to-one constraint geometrically natural and allows a controlled comparison of extrapolation strength. The choice of $\delta \in \{1, 2\}$ is exploratory; automatic selection of δ is left as future work.

4.1 Experimental Setup

In the experiments, we use two representative chaotic dynamical systems as toy models: the Rössler system [10] and the Lorenz'63 system [11]. In addition, we apply the proposed method to real flood-related data [12].

Numerical computations were performed using NumPy [13] and SciPy [14]. The schematic illustrations (Figs. 1 and 2) were created using Microsoft PowerPoint, and all other figures were generated using Matplotlib. A large language model (LLM) was used to assist in writing the program code for the numerical experiments; the authors reviewed and tested the code and take full responsibility for it. The continuous-time dynamical systems were numerically integrated with the ordinary differential equation (ODE) solver `solve_ivp`. We used its adaptive Dormand–Prince 5(4) explicit Runge–Kutta method (RK45) [15]. The tolerances were set to rtol $= 10^{-9}$ and atol $= 10^{-12}$. The continuous trajectories were evaluated at equally spaced times to obtain discrete time series. Initial conditions and observation noise were generated using NumPy's pseudorandom number generator with fixed seeds.

The weights used in the local prediction are computed by solving a linear programming (LP) problem. The objective is the L^1 norm of the reconstruction residual of the target point. This norm is minimized under the sum-to-one constraint and the prescribed box constraints on the weights. If the solver did not return a solution satisfying the constraints within a tolerance of 10^{-6}, the problem was re-solved with alternative solver algorithms and relaxed solver tolerances; if this also failed, the uniform weights $w_i = 1/k$, which satisfy all constraints, were used. In the toy-model experiments, this fallback was used in 19 of the 1,202,400 weight computations (0.0016%), all with affine weight ranges on clean data; no LP failure occurred for the flood data or the extrapolation example. Thus, no prediction was excluded from the mean absolute error (MAE) calculation.

The Rössler system is defined as follows:

$$\begin{cases} \dot{x} = -y - z \\ \dot{y} = x + ay \\ \dot{z} = b + z(x - c) \end{cases} \qquad (a = 0.2,\ b = 0.2,\ c = 5.7). \tag{27}$$

The Lorenz'63 system is defined as follows:

$$\begin{cases} \dot{x} = \sigma(y - x) \\ \dot{y} = (\rho - z)x - y \\ \dot{z} = xy - \beta z \end{cases} \qquad (\sigma = 10,\ \rho = 28,\ \beta = \tfrac{8}{3}). \tag{28}$$

We use two schemes for p-step-ahead prediction: recursive prediction and direct prediction. In recursive prediction, one-step-ahead prediction is repeated p times. In direct prediction, the p-step-ahead value is predicted in a single step.

To specify a common evaluation horizon T_p for each toy system, we use a Kantz-style finite-lag divergence calculation [16] on the first 4500 samples of the clean seed-0 series. The embedding parameters for this diagnostic are separate from those used by the predictor in Table 1. Let $\mathbf{u}_i = [x_i, x_{i+\tau_L}, \ldots, x_{i+(m_L-1)\tau_L}]^\mathsf{T}$, and let U_i contain the indices j for which $\|\mathbf{u}_i - \mathbf{u}_j\|_2 \le \epsilon_L$ and $|i-j| > 10$. Both i and j are restricted to indices whose successors through lag 50 remain in the training interval. For $I = \{i : |U_i| > 0\}$, we compute

$$S(r) = \frac{1}{|I|} \sum_{i \in I} \log \left(\frac{1}{|U_i|} \sum_{j \in U_i} \|\mathbf{u}_{i+r} - \mathbf{u}_{j+r}\|_2 \right), \qquad r = 1, \ldots, 50. \tag{29}$$

Thus, distances are averaged within each neighborhood before taking the logarithm, and eligible reference points are then weighted equally. We fit a straight line to $S(r)$ over lags 6–15, inclusive, and denote its slope per sampled time step by λ_{fit}. We use $m_L = 3$ for both systems, with $(\tau_L, \epsilon_L) = (6, 0.14)$ for Rössler and $(3, 0.08)$ for Lorenz'63; the Theiler window is 10 samples. These settings were selected retrospectively in a parameter search to retain the evaluation windows used in the existing prediction comparisons, using no test samples. They were not selected by the prediction-embedding procedure, and the prediction settings were not changed.

The resulting slopes are finite-lag diagnostics rather than validated estimates of a parameter-invariant maximal Lyapunov exponent. With these settings fixed across all 100 clean training series, the median slopes are 0.02469249 for Rössler and 0.09654364 for Lorenz'63. For seed 0, shifting the regression interval by one sample changes the fitted slopes by up to approximately 7.7% and 8.5%, respectively, and later-lag intervals yield smaller slopes. These sensitivities limit a dynamical interpretation of the fitted values. We therefore use Eq. (30) only to document fixed evaluation windows, not to guarantee accurate prediction throughout those windows. The scale parameters $\delta_0 = 10^{-10}$ and $\Delta = 10^{-5}$ are retained for numerical compatibility and are not calibrated to the measured forecast errors.

The improvement rate is defined on the basis of the MAE. The MAE is the average of the absolute differences between the predicted values and the true values. In the toy-model experiments, for each independent test series r, we first compute two MAE values. One is the MAE of the baseline method, and the other is that of the compared method. Both are computed over the evaluation targets within the evaluation horizon T_p. Here, λ_{fit} is measured per sampled time step. Therefore, T_p is expressed as the number of sampled points rather than physical time. We denote these two values by $\mathrm{MAE}_{\mathrm{base},r}$ and $\mathrm{MAE}_{\mathrm{new},r}$, respectively. We then calculate the improvement rate for each test series by Eq. (31). The mean and median improvement rates reported in the toy-model tables are computed over all independent test series.

$$T_p = \left\lfloor \frac{1}{\lambda_{\mathrm{fit}}} \log \frac{\Delta}{\delta_0} + \frac{1}{2} \right\rfloor \tag{30}$$

$$\frac{\mathrm{MAE}_{\mathrm{base}} - \mathrm{MAE}_{\mathrm{new}}}{\mathrm{MAE}_{\mathrm{base}}} \times 100 \tag{31}$$

4.2 Validation of the Embedding Dimension for Toy Models

The embedding parameters were selected using only the training interval of each time series. The delay τ was selected using the first-minimum mutual-information criterion of Fraser and Swinney [17]. For each lag, mutual information was estimated using the Kraskov–Stögbauer–Grassberger (KSG-1) estimator [18] with $k = 4$ nearest neighbors. The delay τ was then set to the first local minimum of the group-median mutual-information curve. We then applied the False Nearest Strands (FNS) method of Kennel and Abarbanel [19] to candidate dimensions $m = 1, \ldots, 12$. As an additional study-specific robustness rule, we selected the smallest m for which the 90th percentile of the FNS ratio across realizations did not exceed 1%. The number of nearest neighbors used for prediction was fixed at $k = 5$ after parameter selection. Table 1 summarizes the embedding parameters used in the toy-model experiments.

Table 1 Embedding parameters used in the toy-model prediction experiments. "FNS unselectable" means that no candidate in $m = 1, \ldots, 12$ satisfied the study-specific 90th-percentile criterion.

System	Data condition	τ	FNS result	Operational m
Rössler	Clean	13	Selected	4
Rössler	0.1%, 0.5%, 1% noise	14	Selected	4
Lorenz'63	Clean and all noise levels	16	Unselectable	3[a]

[a]Selected only by a supplementary conventional false nearest neighbors (FNN) analysis and used as an exploratory operational setting.

For every Rössler condition, the criterion selected $m = 4$. For Lorenz'63, the median FNS ratio at $m = 3$ was between 0% and 0.187%. However, the 90th percentile of the FNS ratio at $m = 3$ ranged from 1.064% to 3.004% across the four noise conditions. No candidate dimension therefore satisfied the prespecified criterion for Lorenz'63. The FNS ratio nevertheless attained its minimum at $m = 3$. A supplementary conventional

false nearest neighbors (FNN) analysis [20] also selected $m = 3$ for every Lorenz'63 condition. We therefore used $m = 3$ only as an exploratory operational setting. It is not interpreted as an FNS-selected dimension.

4.3 Results on Toy Models

Unless otherwise stated, the following toy-model experiments use $m = 4$, $\tau = 13$ for clean Rössler data, $m = 4$, $\tau = 14$ for noisy Rössler data, and the exploratory setting $m = 3$, $\tau = 16$ for all Lorenz'63 data; $k = 5$ is used throughout.

4.3.1 Evaluation at Local Extreme Values

For validation, we generate 100 time series from randomly selected initial conditions. For each series, the first 2000 steps for the Rössler system and the first 1000 steps for the Lorenz'63 system are discarded as transients, and the following 5000 points are used. The first 4500 points are used as training data, and the remaining 500 points are used as test data. In this subsection, we evaluate the prediction performance only at local extreme values. Because affine weights are expected to be most useful near attractor boundaries, we first evaluate prediction accuracy at local extreme values.

For the Rössler system, the integration step size is 0.1. For the reference training series, the finite-lag slope is $\lambda_{\text{fit}} = 0.02372127$ per sampled step. With $\delta_0 = 10^{-10}$ and $\Delta = 10^{-5}$, the unrounded value is 485.3418 samples; Eq. (30) rounds it to the fixed evaluation horizon $T_p = 485$. Figures 3 and 4 show the recursive and direct prediction results for the Rössler system.

Figure 3 shows lower median errors for the affine methods over early and intermediate local extreme values in recursive prediction. The difference remains visible up to the 17th local extreme value and at the last local extreme value within $T_p = 485$.

Figure 4 shows that the reduction in prediction error is concentrated in the first few local extreme values in direct prediction. Near the evaluation horizon, the affine methods have larger median errors than the convex method.

Within the range determined by the evaluation horizon T_p, we compute the MAE for the first local extreme value and the last local extreme value within the evaluation window. For the Rössler system, the last target is the 16th or 17th local extreme value, depending on the series. Tables 2 and 3 show the improvement rates obtained by changing the weight constraint from the convex combination to the affine combination. Both the mean and median improvement rates are reported to examine the influence of outliers.

Table 2 Rössler system – improvement rates for recursive prediction.

Target type	Weight range	Mean improvement (%)	Median improvement (%)
First local extreme value	$-1 \le w \le 2$	24.452	73.374
	$-2 \le w \le 3$	-116.149	77.507
Last local extreme value	$-1 \le w \le 2$	-104.299	32.405
	$-2 \le w \le 3$	-227.759	13.337

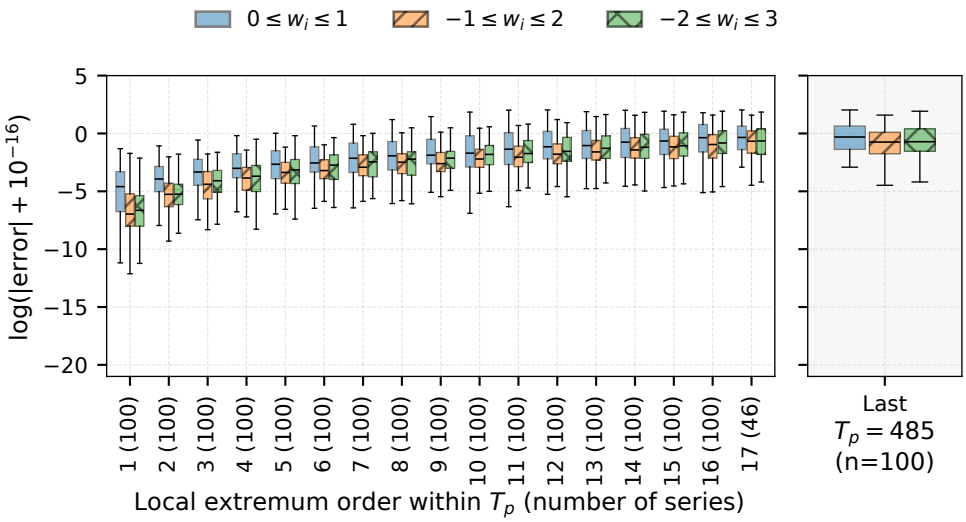


Fig. 3 Results for local extreme values in the Rössler system with recursive prediction. The horizontal axis of the left panel represents the order of the local extreme values within the evaluation horizon $T_p = 485$, and the number in parentheses is the number of series having that local extreme value within T_p. The vertical axis represents the natural logarithm of the absolute prediction error, $\log(|\text{prediction error}| + 10^{-16})$, where 10^{-16} is added to compute the logarithm stably. Each boxplot is computed from those of the 100 time series, generated from different initial conditions, that have the corresponding local extreme value. The right panel shows the last local extreme value within T_p in each series, which is the 16th or 17th local extreme value depending on the series

Table 2 quantifies the recursive-prediction result. The median improvements are 73.374% and 77.507% at the first local extreme value and remain positive at 32.405% and 13.337% at the last local extreme value. The mean improvement is positive only for the narrower affine range at the first local extreme value; the other three mean improvements are negative, indicating substantial degradation in some runs.

Table 3 Rössler system – improvement rates for direct prediction.

Target type	Weight range	Mean improvement (%)	Median improvement (%)
First local extreme value	$-1 \le w \le 2$	37.924	73.545
	$-2 \le w \le 3$	-12.044	76.188
Last local extreme value	$-1 \le w \le 2$	-444.335	-27.027
	$-2 \le w \le 3$	-1398.334	-34.047

Table 3 shows that, at the first local extreme value, the median improvements are 73.545% and 76.188%. The corresponding mean improvement is 37.924% for the narrower affine range but −12.044% for the wider range. At the last local extreme value, both the mean and median improvements become negative. Therefore, for direct

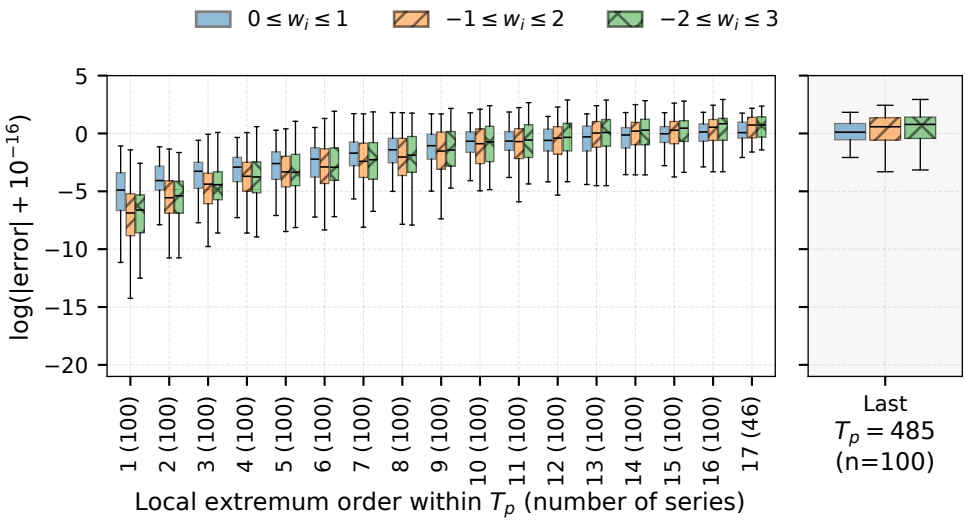


Fig. 4 Results for local extreme values in the Rössler system with direct prediction. The horizontal axis of the left panel represents the order of the local extreme values within the evaluation horizon $T_p = 485$, and the number in parentheses is the number of series having that local extreme value within T_p. The vertical axis represents the natural logarithm of the absolute prediction error, $\log(|\text{prediction error}| + 10^{-16})$, where 10^{-16} is added to compute the logarithm stably. Each boxplot is computed from those of the 100 time series, generated from different initial conditions, that have the corresponding local extreme value. The right panel shows the last local extreme value within T_p in each series, which is the 16th or 17th local extreme value depending on the series

prediction of the Rössler system, affine combinations are useful mainly in typical short-term cases. However, the strongly negative mean improvements near the evaluation horizon show that some runs are substantially degraded.

For the Lorenz'63 system, the integration step size is 0.01. For the reference training series, the finite-lag slope is $\lambda_{\text{fit}} = 0.10925145$ per sampled step. With $\delta_0 = 10^{-10}$ and $\Delta = 10^{-5}$, the unrounded value is 105.3801 samples; Eq. (30) rounds it to the fixed evaluation horizon $T_p = 105$. Figures 5 and 6 show the recursive and direct prediction results for the Lorenz'63 system.

Figure 5 shows lower median errors for the affine methods throughout the evaluation window in recursive prediction. This range includes one to four local extreme values for the Lorenz'63 system, depending on the series.

Figure 6 shows no consistent reduction in median error by the affine methods in direct prediction. The error distributions overlap substantially, and any pointwise difference from the convex method does not persist across the evaluation window.

Within the range determined by the evaluation horizon T_p, we compute the MAE for the first local extreme value and the last local extreme value within the evaluation window. For the Lorenz'63 system, the last target is between the first and fourth local extreme values, depending on the series; if a series has only one local extreme value

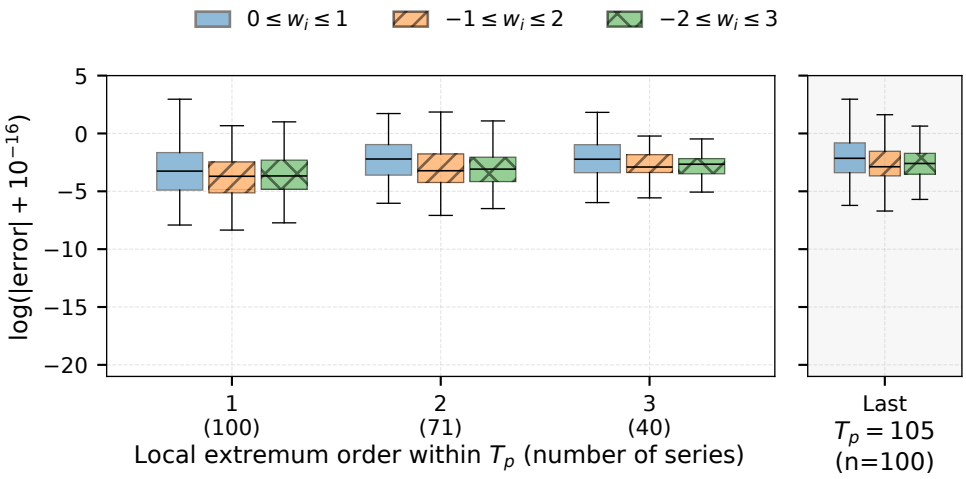


Fig. 5 Results for local extreme values in the Lorenz'63 system with recursive prediction. The horizontal axis of the left panel represents the order of the local extreme values within the evaluation horizon $T_p = 105$, and the number in parentheses is the number of series having that local extreme value within T_p; orders found in fewer than 10 series are not shown. The vertical axis represents the natural logarithm of the absolute prediction error, $\log(|\text{prediction error}| + 10^{-16})$, where 10^{-16} is added to compute the logarithm stably. Each boxplot is computed from those of the 100 time series, generated from different initial conditions, that have the corresponding local extreme value. The right panel shows the last local extreme value within T_p in each series, which is between the first and fourth local extreme values depending on the series

within T_p, the first and last targets coincide. Tables 4 and 5 show the improvement rates.

Table 4 Lorenz'63 system – improvement rates for recursive prediction.

Target type	Weight range	Mean improvement (%)	Median improvement (%)
First local extreme value	$-1 \le w \le 2$	-311.186	31.152
	$-2 \le w \le 3$	-387.798	23.119
Last local extreme value	$-1 \le w \le 2$	-123.721	45.575
	$-2 \le w \le 3$	-257.586	35.435

Table 4 shows positive median improvements of 31.152% and 23.119% at the first local extreme value and 45.575% and 35.435% at the last local extreme value. Thus, the affine combinations improve typical samples in recursive prediction under this exploratory operational setting. The mean improvements are negative in all four cases, showing that large degradations in some runs outweigh the typical improvement.

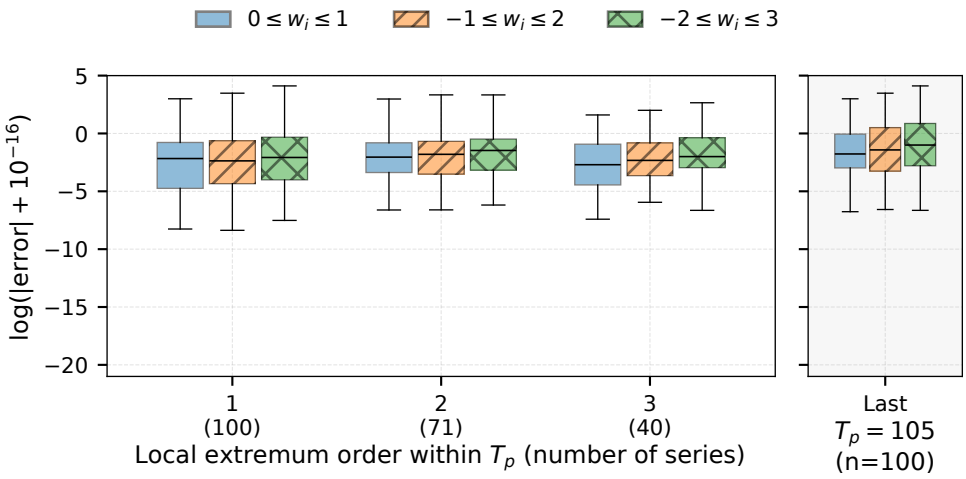


Fig. 6 Results for local extreme values in the Lorenz'63 system with direct prediction. The horizontal axis of the left panel represents the order of the local extreme values within the evaluation horizon $T_p = 105$, and the number in parentheses is the number of series having that local extreme value within T_p; orders found in fewer than 10 series are not shown. The vertical axis represents the natural logarithm of the absolute prediction error, $\log(|\text{prediction error}| + 10^{-16})$, where 10^{-16} is added to compute the logarithm stably. Each boxplot is computed from those of the 100 time series, generated from different initial conditions, that have the corresponding local extreme value. The right panel shows the last local extreme value within T_p in each series, which is between the first and fourth local extreme values depending on the series

Table 5 Lorenz'63 system – improvement rates for direct prediction.

Target type	Weight range	Mean improvement (%)	Median improvement (%)
First local extreme value	$-1 \le w \le 2$	-550.009	-32.994
	$-2 \le w \le 3$	-1004.622	-86.948
Last local extreme value	$-1 \le w \le 2$	-1000.416	-25.635
	$-2 \le w \le 3$	-1653.280	-87.163

Table 5 shows negative mean and median improvements at both the first and last local extreme values. The median improvements range from -25.635% to -87.163%, while the mean improvements are substantially more negative. Thus, the affine combinations are not reliable for direct prediction in the Lorenz'63 system under this exploratory setting.

Overall, affine weights improve typical local-extreme predictions most consistently in recursive forecasting, whereas their performance is unreliable in direct forecasting except for selected short-term Rössler cases.

4.3.2 Evaluation at All Points up to the Evaluation Horizon

We use the same data as in Subsection 4.3.1. In this subsection, we evaluate the prediction performance at all points up to the evaluation horizon.

Figures 7 and 8 show the recursive and direct prediction results for the Rössler system.

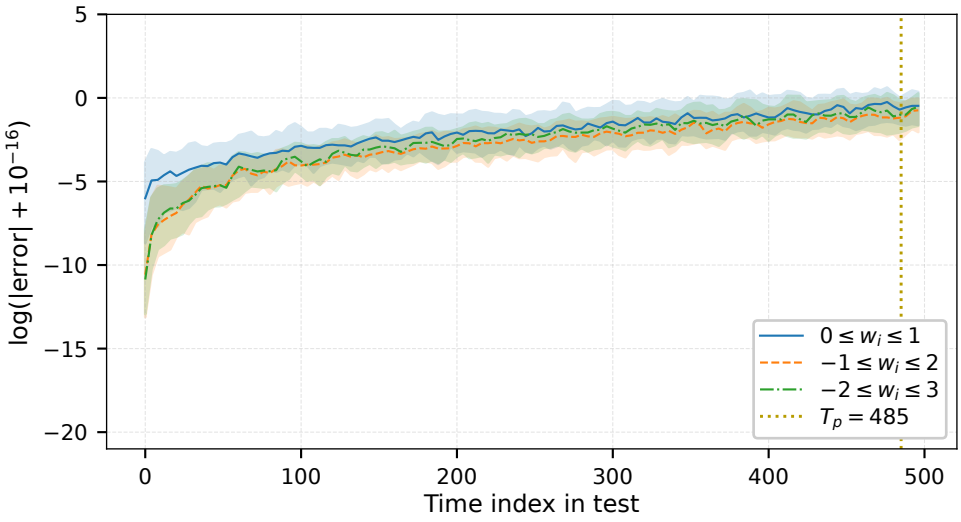


Fig. 7 Evaluation at all points up to the evaluation horizon in the Rössler system with recursive prediction. The horizontal axis represents the time index in the test data. The vertical axis represents the natural logarithm of the absolute prediction error, $\log(|\text{prediction error}| + 10^{-16})$, where 10^{-16} is added to compute the logarithm stably. The thick lines represent the median values over 100 time series generated from different initial conditions. The shaded regions represent the interquartile ranges between the first and third quartiles. The ochre dotted vertical line indicates the evaluation horizon T_p. The prediction points on the left side of this line are within the evaluation window. In this figure, the evaluation horizon is $T_p = 485$

Figures 7 and 8 show that recursive prediction gives more stable results than direct prediction for the Rössler system. In recursive prediction, the affine combinations tend to reduce the prediction error over a wide range of prediction points. In direct prediction, the advantage of the affine combinations becomes weaker, and the prediction error can increase substantially.

Because the validation data are the same as those in Subsection 4.3.1, the evaluation horizon is $T_p = 485$. Table 6 shows the improvement rates over all points up to the evaluation horizon.

Table 6 shows positive median improvements of 43.328% and 36.986% in recursive prediction, but the mean improvement is positive only for the narrower affine range (1.702%) and is negative for the wider range (−56.472%). Both mean and median improvements are negative in direct prediction. These results indicate that,

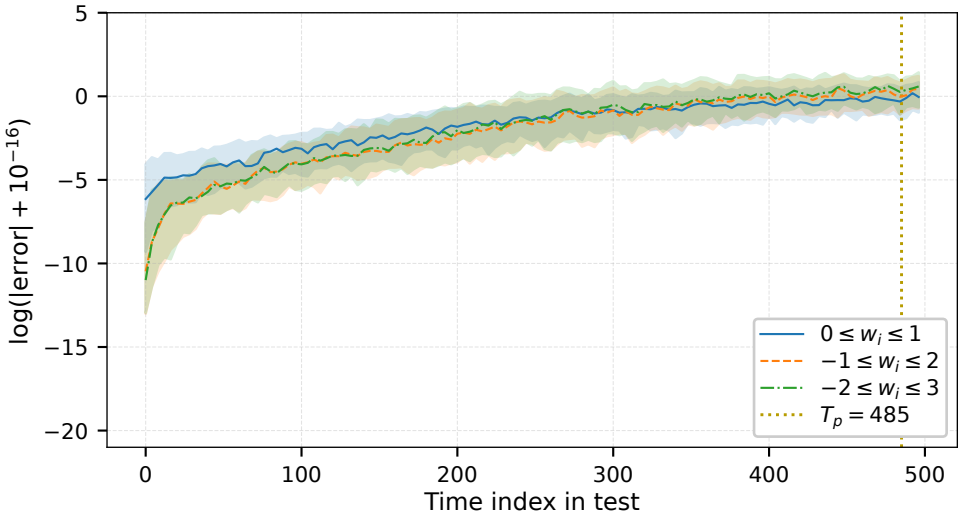


Fig. 8 Evaluation at all points up to the evaluation horizon in the Rössler system with direct prediction. The horizontal axis represents the time index in the test data. The vertical axis represents the natural logarithm of the absolute prediction error, $\log(|\text{prediction error}| + 10^{-16})$, where 10^{-16} is added to compute the logarithm stably. The thick lines represent the median values over 100 time series generated from different initial conditions. The shaded regions represent the interquartile ranges between the first and third quartiles. The ochre dotted vertical line indicates the evaluation horizon T_p. The prediction points on the left side of this line are within the evaluation window. In this figure, the evaluation horizon is $T_p = 485$

Table 6 Improvement rates over all points up to the evaluation horizon for the Rössler system.

Prediction scheme	Weight range	Mean improvement (%)	Median improvement (%)
Recursive prediction	$-1 \leq w \leq 2$	1.702	43.328
	$-2 \leq w \leq 3$	-56.472	36.986
Direct prediction	$-1 \leq w \leq 2$	-58.905	-22.678
	$-2 \leq w \leq 3$	-133.729	-35.351

for the Rössler system, a typical recursive run can benefit from affine combinations, whereas the wider range and direct prediction can substantially degrade performance. Therefore, the advantage of the affine combinations is often more stable in recursive prediction than in direct prediction.

Figures 9 and 10 show the recursive and direct prediction results for the Lorenz'63 system.

Figures 9 and 10 show that the effect of the affine combinations depends on the prediction scheme in the Lorenz'63 system. For recursive prediction, the affine

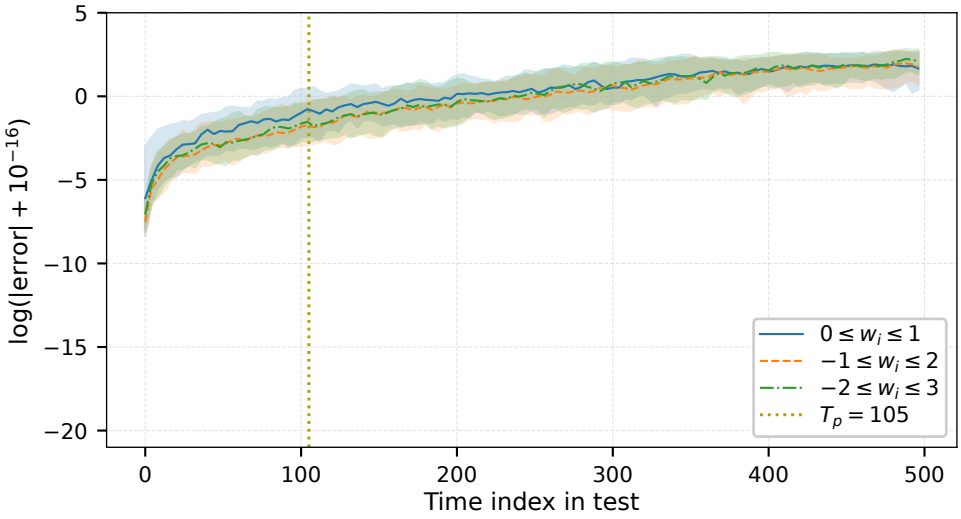


Fig. 9 Evaluation at all points up to the evaluation horizon in the Lorenz'63 system with recursive prediction. The horizontal axis represents the time index in the test data. The vertical axis represents the natural logarithm of the absolute prediction error, $\log(|\text{prediction error}| + 10^{-16})$, where 10^{-16} is added to compute the logarithm stably. The thick lines represent the median values over 100 time series generated from different initial conditions. The shaded regions represent the interquartile ranges between the first and third quartiles. The ochre dotted vertical line indicates the evaluation horizon T_p. The prediction points on the left side of this line are within the evaluation window. In this figure, the evaluation horizon is $T_p = 105$

combinations reduce the median error when all points up to the evaluation horizon are evaluated. In contrast, direct prediction gives negative mean and median improvements.

Because the validation data are the same as those in Subsection 4.3.1, the evaluation horizon is $T_p = 105$. Table 7 shows the improvement rates over all points up to the evaluation horizon.

Table 7 Improvement rates over all points up to the evaluation horizon for the Lorenz'63 system.

Prediction scheme	Weight range	Mean improvement (%)	Median improvement (%)
Recursive prediction	$-1 \leq w \leq 2$	3.605	33.271
	$-2 \leq w \leq 3$	-43.581	18.618
Direct prediction	$-1 \leq w \leq 2$	-192.857	-9.505
	$-2 \leq w \leq 3$	-413.319	-59.804

Table 7 shows positive median improvements of 33.271% and 18.618% in recursive prediction. The corresponding mean improvements are 3.605% and −43.581%, respectively. For direct prediction, both mean and median improvements are negative for

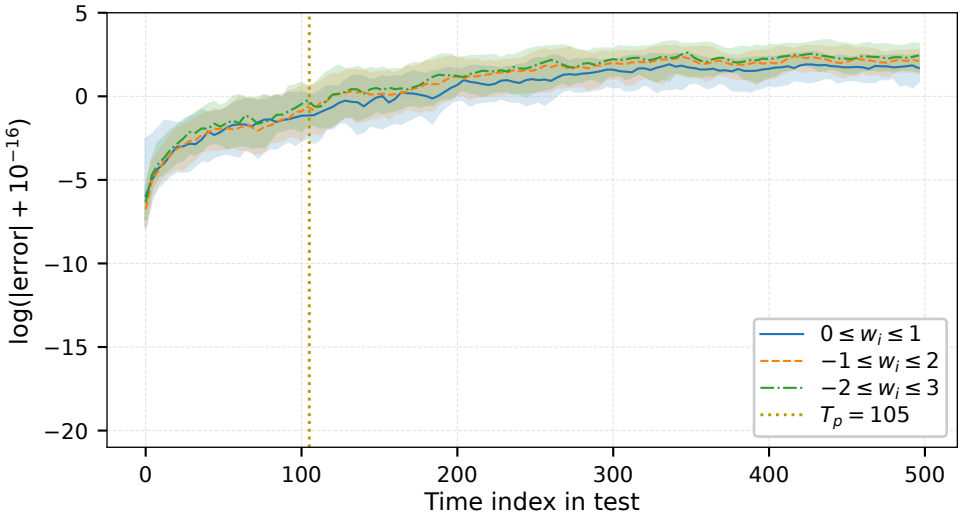


Fig. 10 Evaluation at all points up to the evaluation horizon in the Lorenz'63 system with direct prediction. The horizontal axis represents the time index in the test data. The vertical axis represents the natural logarithm of the absolute prediction error, $\log(|\text{prediction error}| + 10^{-16})$, where 10^{-16} is added to compute the logarithm stably. The thick lines represent the median values over 100 time series generated from different initial conditions. The shaded regions represent the interquartile ranges between the first and third quartiles. The ochre dotted vertical line indicates the evaluation horizon T_p. The prediction points on the left side of this line are within the evaluation window. In this figure, the evaluation horizon is $T_p = 105$

both affine weight ranges. This contrast indicates a typical benefit in recursive prediction under the exploratory Lorenz'63 setting, accompanied by sensitivity to large errors and poor direct-prediction performance.

Overall, all-point evaluation gives positive median improvements in recursive prediction for both toy models, but mean improvements are not uniformly positive. In direct prediction, both mean and median improvements are negative for both systems. Compared with the local-extreme evaluation, the improvements over all points are smaller and less consistent.

4.3.3 Extrapolation Performance

Using $m = 4$, $\tau = 13$, and $k = 5$, we generate a Rössler system time series in which the maximum value appears at the 4459th point. For the prediction of the point at the maximum value, the data up to 10 steps before the target point are used as training data. In the data up to the 4458th point, the maximum value is 11.427455. Figure 11 and Table 8 show the prediction results for the target point at the maximum value.

The true value at the target point is 11.42824.... The prediction with the convex combination, $0 \leq w_i \leq 1$, is 11.42087.... This value is smaller than both the true value and the maximum value contained in the training data. Thus, the convex combination does not predict a value beyond the range of the training data.

In contrast, the predictions with the affine combinations, $-1 \leq w_i \leq 2$ and $-2 \leq w_i \leq 3$, are 11.43014... and 11.43014..., respectively. Both predicted values are larger than the maximum value in the training data. Therefore, the affine combinations predict values outside the observed training range and produce extrapolative predictions. Moreover, both affine predictions are closer to the true value than the convex prediction in this example.

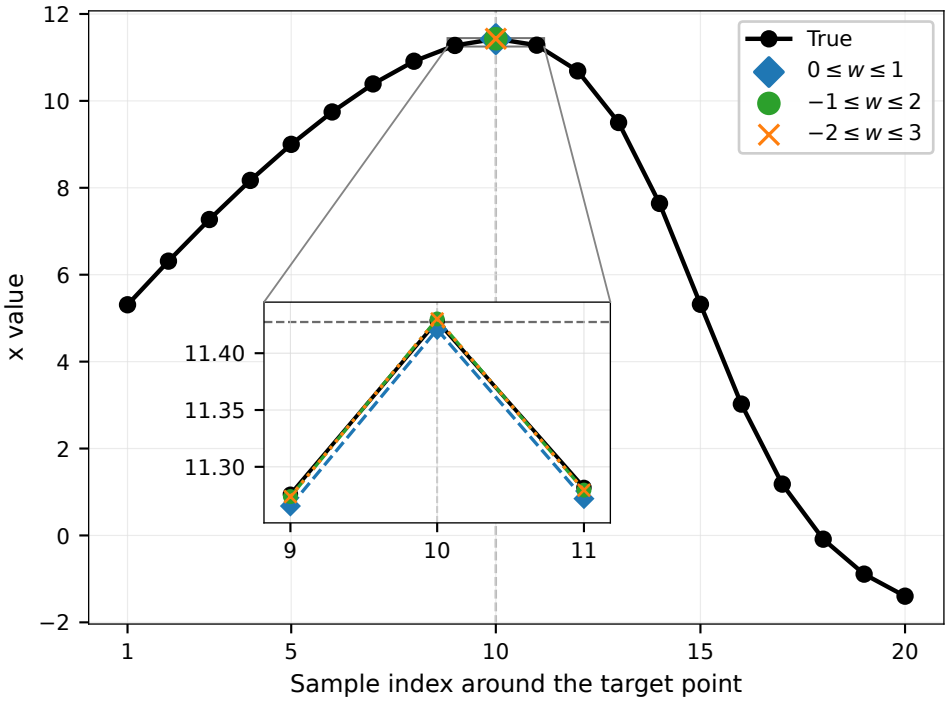


Fig. 11 Prediction values for the target point at the maximum value in the Rössler time series using $m = 4$, $\tau = 13$, and $k = 5$. The horizontal axis represents the sample index around the target point. The vertical axis represents the x-value of the Rössler time series. The black points represent the true values, the blue diamond represents the prediction with the convex combination $0 \leq w_i \leq 1$, the green circle represents the prediction with the affine combination $-1 \leq w_i \leq 2$, and the orange cross represents the prediction with the affine combination $-2 \leq w_i \leq 3$. The maximum value that is not contained in the training data appears at the 10th point on the horizontal axis. The prediction starts 10 points before the maximum value appears, and the predicted values at the 10th point are compared in this figure. The inset enlarges the 9th–11th points around the maximum; its blue dashed, green dash-dotted, and orange dotted curves show the corresponding recursive prediction paths, and the horizontal gray dashed line indicates the maximum value contained in the training data

These results support the interpretation that affine weights are useful when the target value lies near or outside the boundary of the observed training data.

Table 8 Predicted values for the target point at the maximum value in the Rössler time series.

Method or value	Target point (10 steps ahead)
True value	11.42824...
$0 \leq w_i \leq 1$	11.42087...
$-1 \leq w_i \leq 2$	11.43014...
$-2 \leq w_i \leq 3$	11.43014...

4.3.4 Prediction Performance under Noise

We next evaluate the prediction performance when observation noise is added to the toy-model data. The noise is zero-mean Gaussian noise, and its standard deviation is set to 0.1%, 0.5%, or 1% of the standard deviation of the clean time series. For each system, the noise is added to a separately generated set of 100 clean series, in which the first 1000 steps are discarded as transients; this set is common to all noise levels. The same training and test lengths as in the previous subsections are used. The evaluation horizons are kept at $T_p = 485$ for the Rössler system and $T_p = 105$ for the Lorenz'63 system. For the evaluation at local extreme values, the target indices are determined from the clean test data. The improvement rates are computed by using the convex combination $0 \leq w_i \leq 1$ as the baseline.

Figures 12, 13, and 14 show the prediction errors at local extreme values under 0.1%, 0.5%, and 1% observation noise. Compared with the noiseless results in Subsections 4.3.1 and 4.3.2, the advantage of the affine combinations becomes weaker. At 0.1% noise, positive median improvements remain only in isolated local-extreme cases, whereas all mean improvements are negative. At 0.5% and 1% noise, the median improvements are negative for every local-extreme condition reported in Table 9. In particular, the wider affine range $-2 \leq w \leq 3$ often produces larger errors than the convex combination.

Table 9 shows the improvement rates at local extreme values. For the Rössler system under 0.1% noise, the recursive first-extremum medians are -20.694% and -9.600%; the last-extremum medians are 12.473% and -35.007%. For direct prediction at the first extremum, the medians are 9.575% and 0.000%. For the Lorenz'63 system, the narrower recursive range gives a positive median of 8.090% only at the first extremum under 0.1% noise. All mean improvements are negative, and every reported median is negative at 0.5% and 1% noise.

Figures 15, 16, and 17 show the pointwise prediction errors over the test interval under 0.1%, 0.5%, and 1% noise. Table 10 gives the improvement rates computed using all points up to the evaluation horizon. At 0.1% noise, the narrower Rössler recursive range retains a positive median improvement of 22.057%, whereas the wider range gives -18.172%. For Lorenz'63 recursive prediction, the narrower-range median is 0.116%, which is practically zero, and both direct-prediction medians are negative. At 0.5% and 1% noise, all reported median improvements are negative.

Overall, all mean improvements under observation noise are negative, and the improvement rates do not vary monotonically with the noise level. The wider range

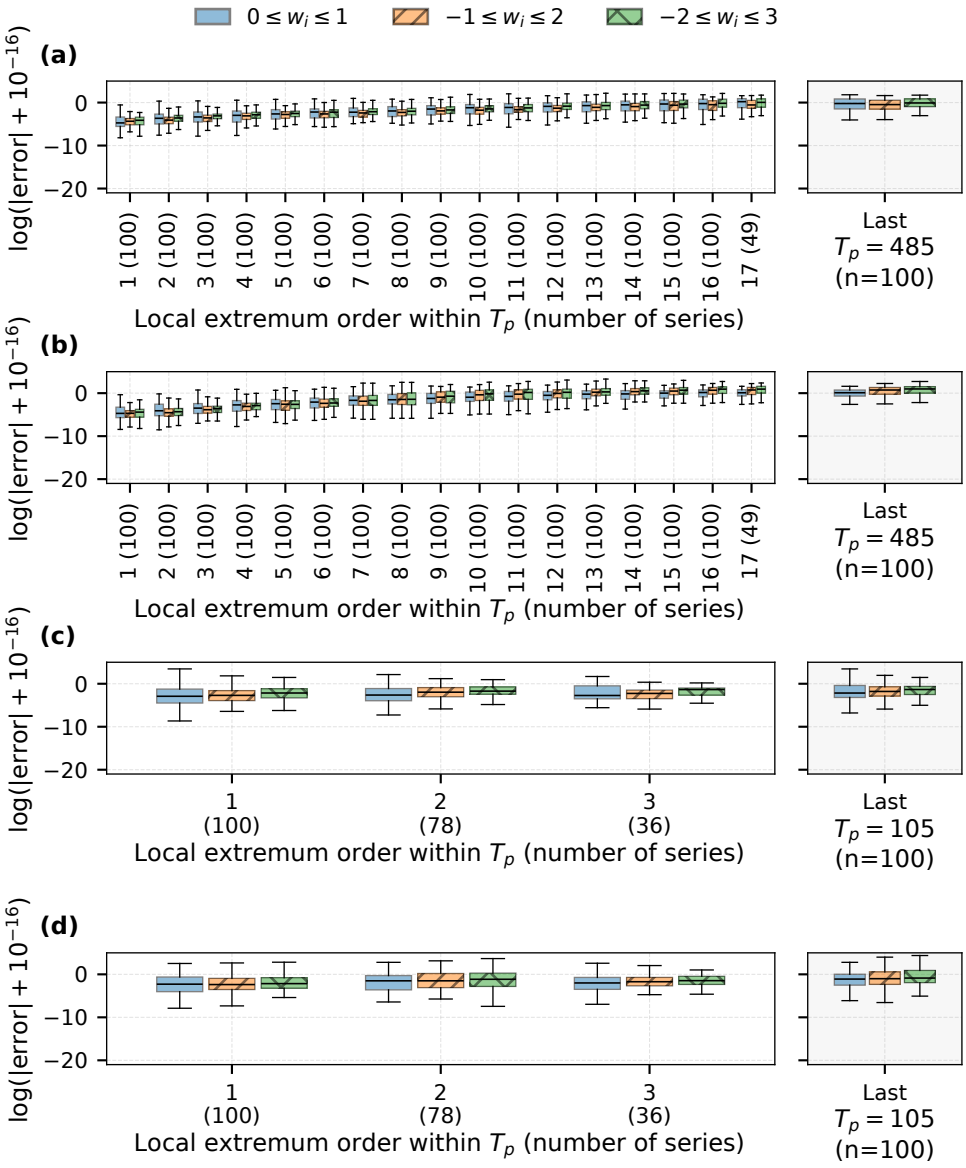


Fig. 12 Prediction errors at local extreme values under 0.1% observation noise. (a) Rössler recursive prediction. (b) Rössler direct prediction. (c) Lorenz'63 recursive prediction. (d) Lorenz'63 direct prediction. Axes and boxplots follow Figs. 3 and 5; numbers in parentheses give the number of series, and Last denotes the last local extreme value within T_p in each series

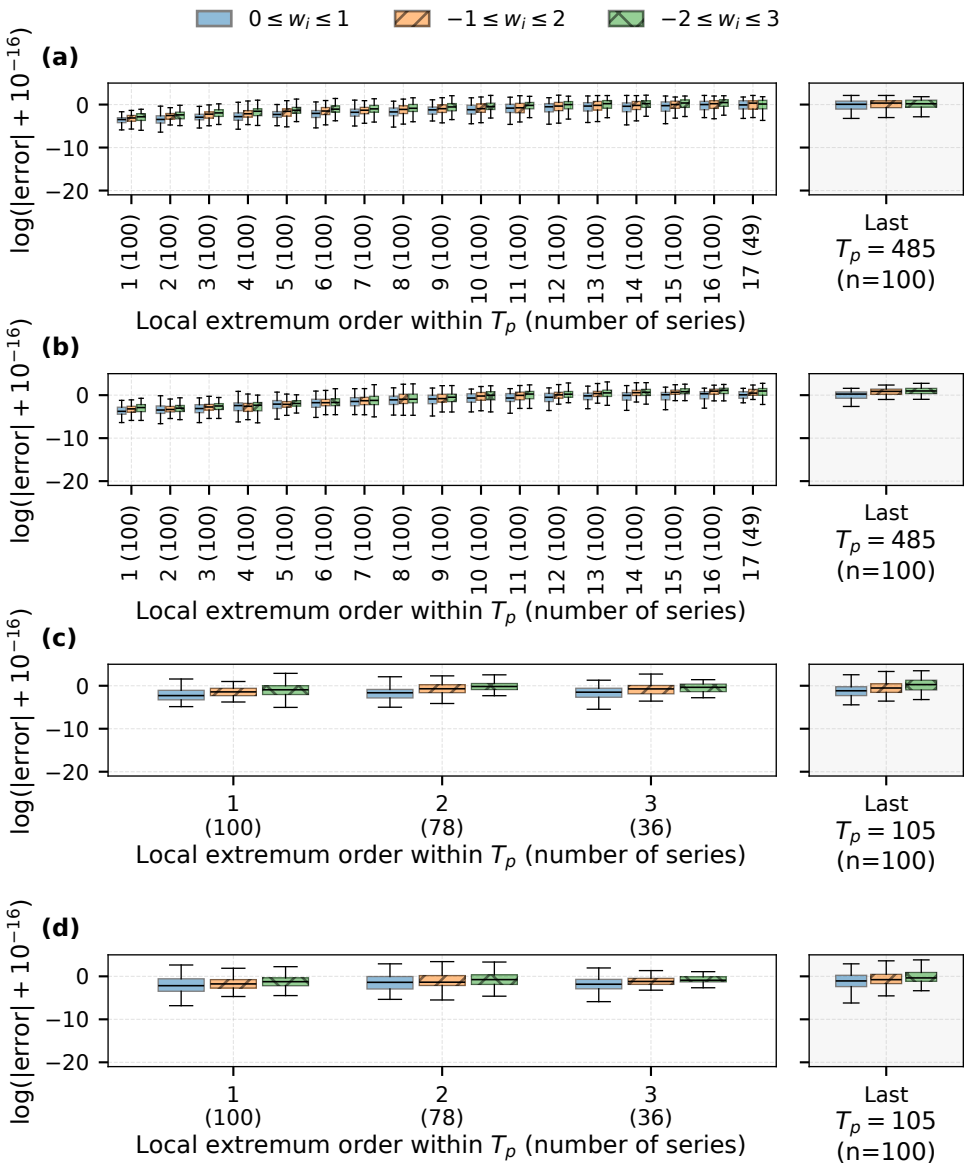


Fig. 13 Prediction errors at local extreme values under 0.5% observation noise. (a) Rössler recursive prediction. (b) Rössler direct prediction. (c) Lorenz'63 recursive prediction. (d) Lorenz'63 direct prediction. Axes and boxplots follow Figs. 3 and 5; numbers in parentheses give the number of series, and Last denotes the last local extreme value within T_p in each series

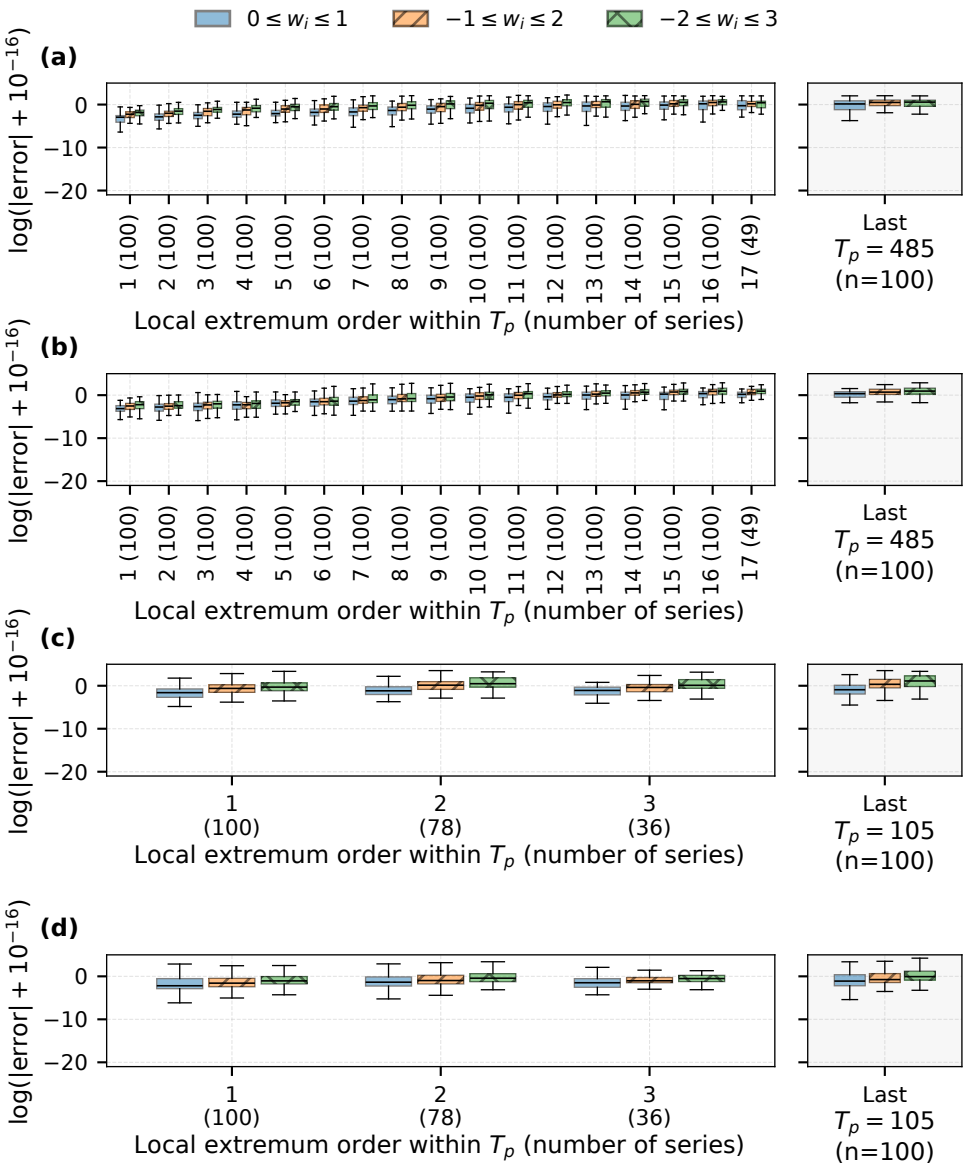


Fig. 14 Prediction errors at local extreme values under 1% observation noise. (a) Rössler recursive prediction. (b) Rössler direct prediction. (c) Lorenz'63 recursive prediction. (d) Lorenz'63 direct prediction. Axes and boxplots follow Figs. 3 and 5; numbers in parentheses give the number of series, and Last denotes the last local extreme value within T_p in each series

Table 9 Mean and median improvement rates at local extreme values under observation noise. Here, "First" denotes the first local extreme value after the prediction starts. "Last" denotes the last local extreme value within T_p in each series ($T_p = 485$ for the Rössler system and $T_p = 105$ for the Lorenz'63 system).

Noise	System	Prediction	Weight range	First mean (%)	First median (%)	Last mean (%)	Last median (%)
0.1%	Rössler	Recursive	$-1 \leq w \leq 2$	-227.566	-20.694	-684.231	12.473
			$-2 \leq w \leq 3$	-445.643	-9.600	-2832.201	-35.007
		Direct	$-1 \leq w \leq 2$	-178.055	9.575	-1236.556	-59.915
			$-2 \leq w \leq 3$	-286.083	0.000	-2329.365	-110.982
0.1%	Lorenz'63	Recursive	$-1 \leq w \leq 2$	-1254.088	8.090	-706.287	-25.790
			$-2 \leq w \leq 3$	-1887.655	-94.173	-712.120	-49.652
		Direct	$-1 \leq w \leq 2$	-140.516	-13.840	-333.506	-41.123
			$-2 \leq w \leq 3$	-295.609	-70.950	-670.623	-130.178
0.5%	Rössler	Recursive	$-1 \leq w \leq 2$	-791.343	-50.093	-209.234	-7.953
			$-2 \leq w \leq 3$	-1035.736	-120.766	-419.367	-13.765
		Direct	$-1 \leq w \leq 2$	-377.310	-15.814	-438.746	-75.814
			$-2 \leq w \leq 3$	-511.313	-36.873	-747.274	-99.757
0.5%	Lorenz'63	Recursive	$-1 \leq w \leq 2$	-389.501	-99.996	-829.176	-84.298
			$-2 \leq w \leq 3$	-744.029	-250.534	-4410.641	-369.303
		Direct	$-1 \leq w \leq 2$	-491.905	-47.767	-698.233	-63.995
			$-2 \leq w \leq 3$	-945.867	-121.349	-1568.465	-126.226
1%	Rössler	Recursive	$-1 \leq w \leq 2$	-752.287	-131.605	-454.665	-29.778
			$-2 \leq w \leq 3$	-848.850	-199.118	-530.960	-52.328
		Direct	$-1 \leq w \leq 2$	-344.970	-32.262	-448.009	-43.348
			$-2 \leq w \leq 3$	-533.550	-88.769	-547.873	-64.719
1%	Lorenz'63	Recursive	$-1 \leq w \leq 2$	-635.603	-207.871	-1480.601	-194.867
			$-2 \leq w \leq 3$	-1521.704	-310.317	-2678.756	-615.362
		Direct	$-1 \leq w \leq 2$	-327.099	-30.284	-586.054	-41.207
			$-2 \leq w \leq 3$	-734.232	-138.412	-1120.890	-131.597

$-2 \leq w \leq 3$ is generally less stable than $-1 \leq w \leq 2$, and improvements under observation noise are limited to selected cases.

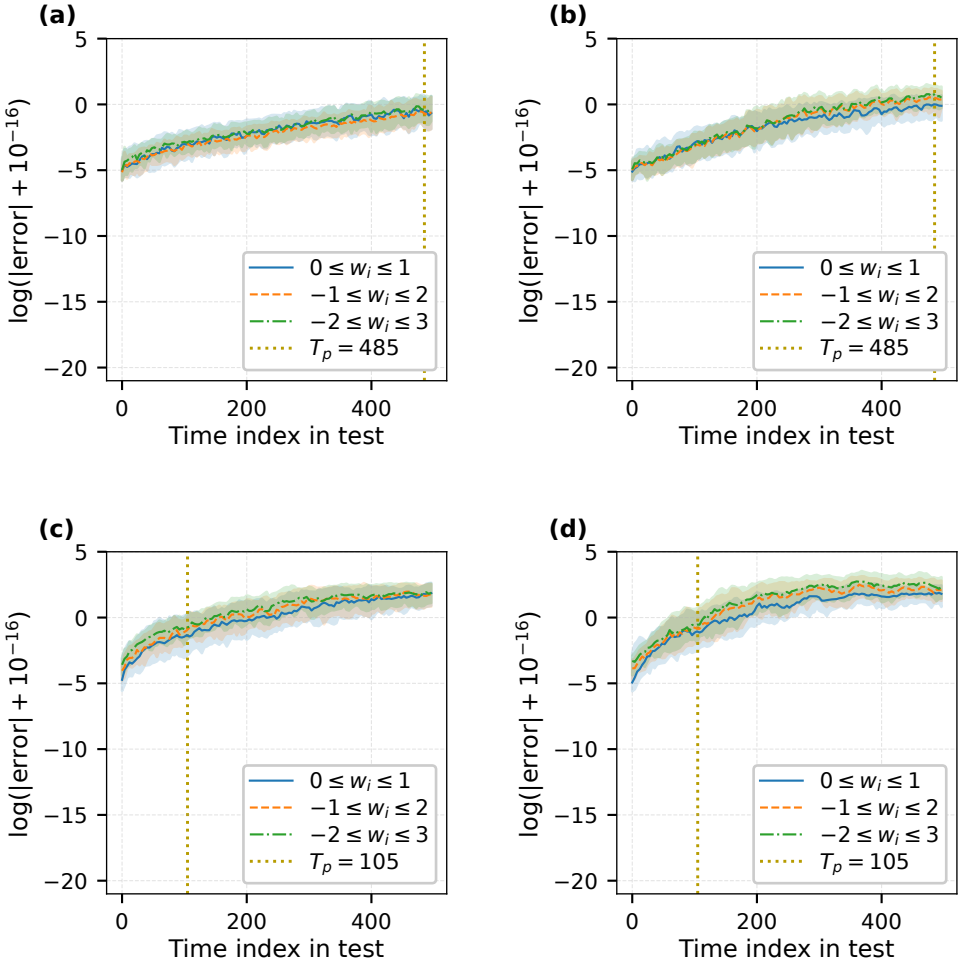


Fig. 15 Prediction errors over the test interval under 0.1% observation noise. (a) Rössler recursive prediction. (b) Rössler direct prediction. (c) Lorenz'63 recursive prediction. (d) Lorenz'63 direct prediction. Axes, medians, interquartile shading, and dotted evaluation-horizon lines follow Figs. 7 and 9

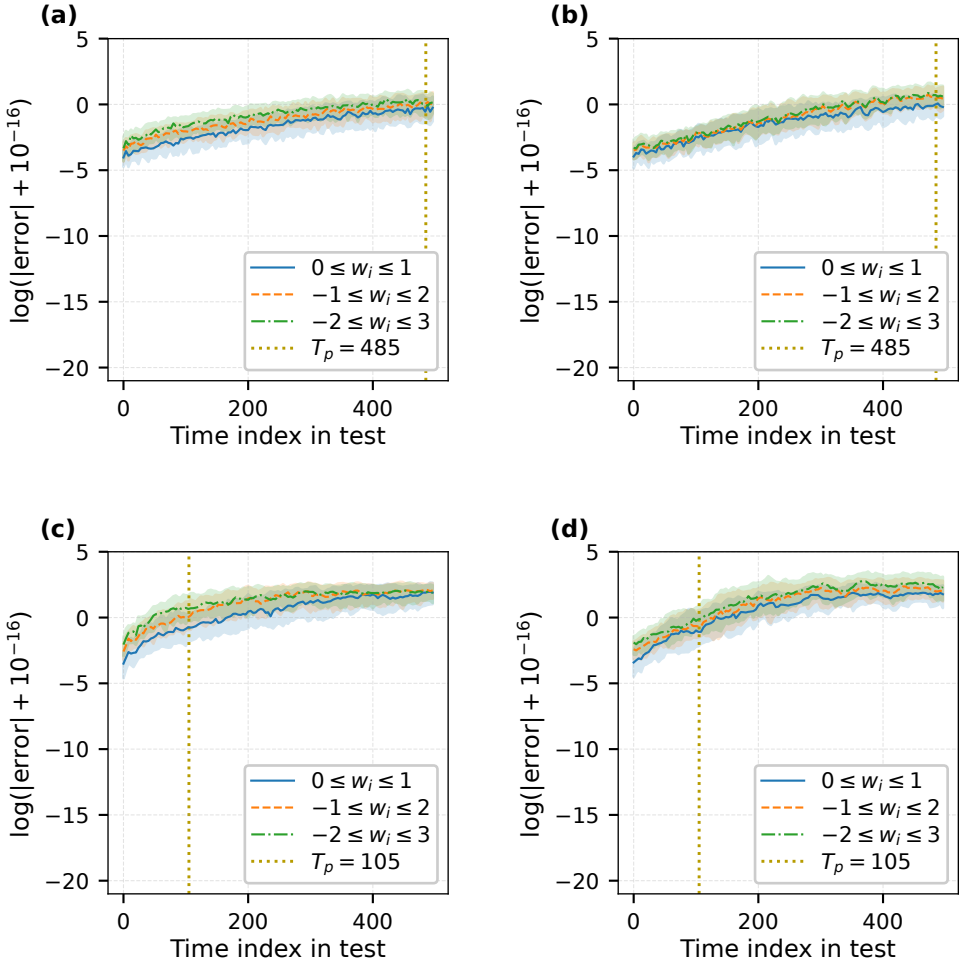


Fig. 16 Prediction errors over the test interval under 0.5% observation noise. (a) Rössler recursive prediction. (b) Rössler direct prediction. (c) Lorenz'63 recursive prediction. (d) Lorenz'63 direct prediction. Axes, medians, interquartile shading, and dotted evaluation-horizon lines follow Figs. 7 and 9

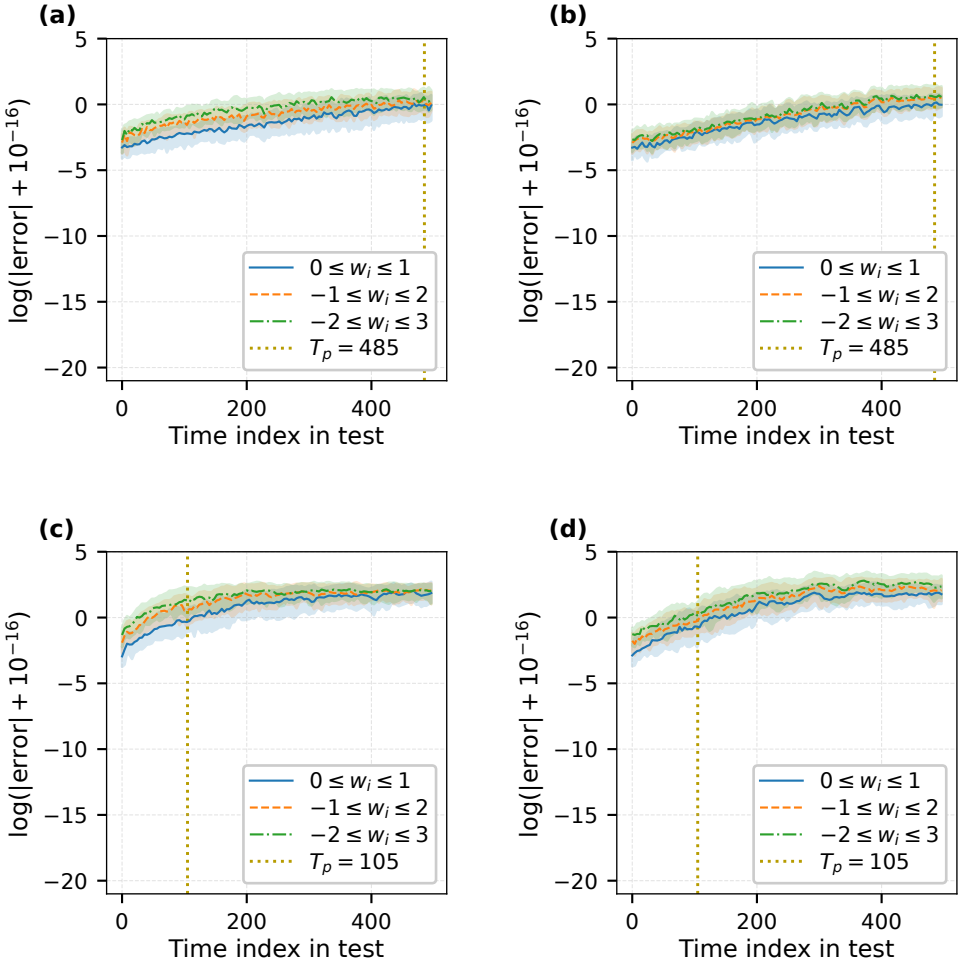


Fig. 17 Prediction errors over the test interval under 1% observation noise. (a) Rössler recursive prediction. (b) Rössler direct prediction. (c) Lorenz'63 recursive prediction. (d) Lorenz'63 direct prediction. Axes, medians, interquartile shading, and dotted evaluation-horizon lines follow Figs. 7 and 9

Table 10 Mean and median improvement rates over all points up to the evaluation horizon under observation noise.

Noise	System	Prediction	$-1 \leq w \leq 2$ mean (%)	$-1 \leq w \leq 2$ median (%)	$-2 \leq w \leq 3$ mean (%)	$-2 \leq w \leq 3$ median (%)
0.1%	Rössler	Recursive	-62.971	22.057	-136.569	-18.172
		Direct	-111.230	-45.521	-174.290	-67.466
	Lorenz'63	Recursive	-178.242	0.116	-308.954	-30.452
		Direct	-136.152	-55.975	-287.012	-104.868
0.5%	Rössler	Recursive	-138.256	-54.906	-246.661	-92.658
		Direct	-125.493	-58.844	-197.031	-67.429
	Lorenz'63	Recursive	-247.465	-131.727	-570.645	-379.804
		Direct	-115.615	-43.269	-274.091	-115.897
1%	Rössler	Recursive	-196.615	-46.697	-347.855	-126.632
		Direct	-85.417	-31.477	-148.872	-52.595
	Lorenz'63	Recursive	-348.539	-156.553	-652.990	-314.734
		Direct	-142.078	-69.417	-328.618	-175.634

4.4 Results on Real Data

We evaluate the proposed method using a multivariate subset derived from the flood-related dataset described by Dawson et al. [12]. All nine variables are used as inputs, with $V8$ as the prediction target. This subset consists of two training series of lengths 728 and 732 and one test series of length 728.

Using only the two training series, the mutual-information procedure estimated $\tau = 27$ (6.75 days), but neither FNS nor the supplementary conventional FNN analysis identified an embedding dimension in $m = 1, \ldots, 12$. We therefore did not regard a minimum embedding dimension as identified and did not perform a parameter-selected primary prediction analysis for this dataset. For continuity with the previous benchmark, the results below retain the legacy parameters $m = 3$ and $\tau = 5$ and are presented only as an exploratory analysis. We use recursive prediction as the prediction scheme. For each point in the test data, we perform recursive prediction for lead times $H = 1, 2, 3, 4$. Figures 18, 19, 20, and 21 show the prediction results for all evaluation points at each lead time. Table 11 reports MAE improvement rates for the single flood test series, using Eq. (31) with MAEs computed over all evaluation points or local extreme values.

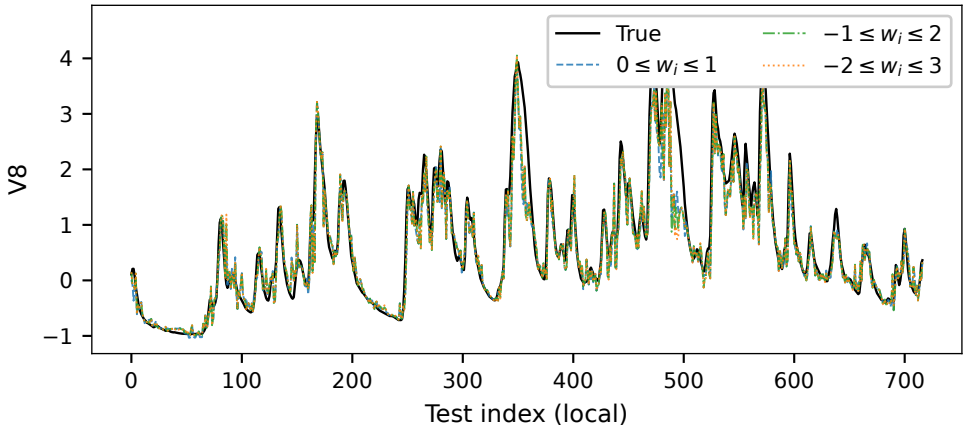


Fig. 18 Exploratory prediction results for all evaluation points in the flood data with the legacy setting $m = 3$, $\tau = 5$, and $H = 1$. The horizontal axis represents the time index in the test data. The vertical axis represents the value of the target variable $V8$. The curve labeled True represents the observed values. The other curves represent the predicted values obtained with the convex combination $0 \le w \le 1$, the affine combination $-1 \le w \le 2$, and the affine combination $-2 \le w \le 3$. This figure compares the observed values and the predicted values over all evaluation points for the 1-step-ahead prediction

Table 11 shows positive MAE improvements for both evaluation sets at $H = 1$, with larger improvements at local extreme values. For $H = 2$, $H = 3$, and $H = 4$, the MAE improvements for all evaluation points become negative for both affine weight

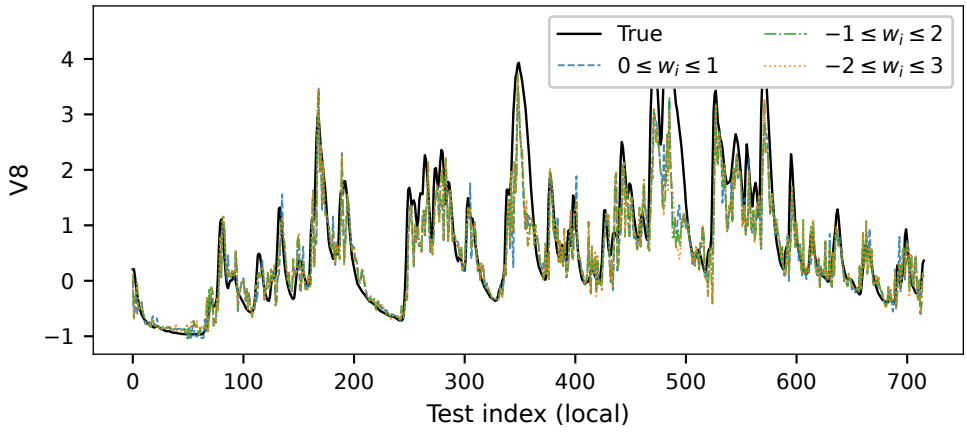


Fig. 19 Exploratory prediction results for all evaluation points in the flood data with the legacy setting $m = 3$, $\tau = 5$, and $H = 2$. The horizontal axis represents the time index in the test data. The vertical axis represents the value of the target variable $V8$. The curve labeled True represents the observed values. The other curves represent the predicted values obtained with the convex combination $0 \le w \le 1$, the affine combination $-1 \le w \le 2$, and the affine combination $-2 \le w \le 3$. This figure compares the observed values and the predicted values over all evaluation points for the 2-step-ahead prediction

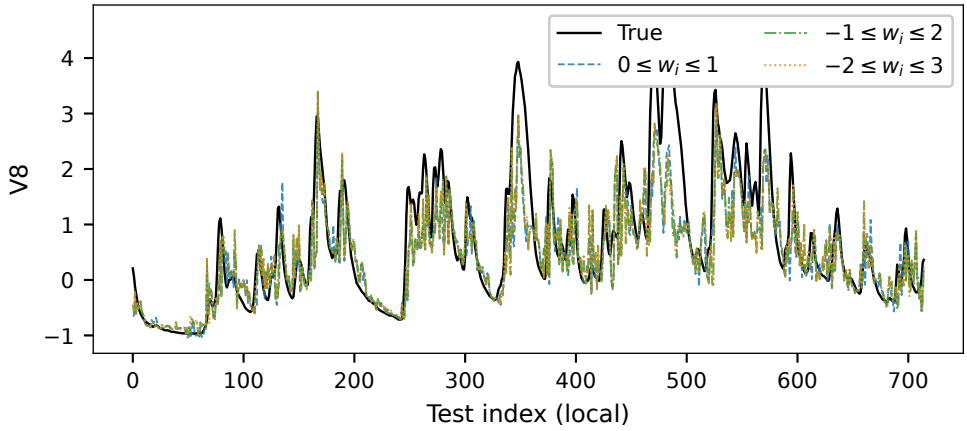


Fig. 20 Exploratory prediction results for all evaluation points in the flood data with the legacy setting $m = 3$, $\tau = 5$, and $H = 3$. The horizontal axis represents the time index in the test data. The vertical axis represents the value of the target variable $V8$. The curve labeled True represents the observed values. The other curves represent the predicted values obtained with the convex combination $0 \le w \le 1$, the affine combination $-1 \le w \le 2$, and the affine combination $-2 \le w \le 3$. This figure compares the observed values and the predicted values over all evaluation points for the 3-step-ahead prediction

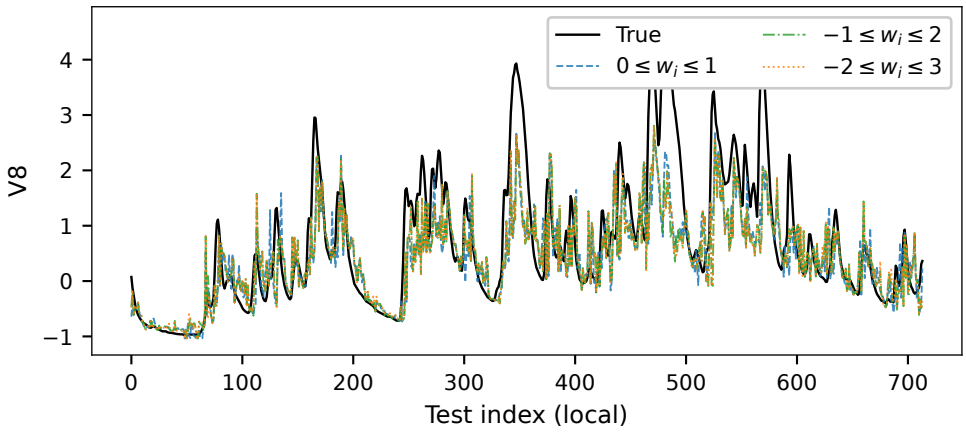


Fig. 21 Exploratory prediction results for all evaluation points in the flood data with the legacy setting $m = 3$, $\tau = 5$, and $H = 4$. The horizontal axis represents the time index in the test data. The vertical axis represents the value of the target variable $V8$. The curve labeled True represents the observed values. The other curves represent the predicted values obtained with the convex combination $0 \le w \le 1$, the affine combination $-1 \le w \le 2$, and the affine combination $-2 \le w \le 3$. This figure compares the observed values and the predicted values over all evaluation points for the 4-step-ahead prediction

Table 11 MAE improvement rates for the single flood test series, evaluated over all points or local extreme values with the exploratory legacy setting $m = 3$ and $\tau = 5$.

Lead time	Weight range	MAE improvement for all evaluation points (%)	MAE improvement for local extreme values (%)
$H = 1$	$-1 \le w \le 2$	2.635	6.007
	$-2 \le w \le 3$	2.911	4.960
$H = 2$	$-1 \le w \le 2$	-1.412	4.765
	$-2 \le w \le 3$	-1.301	4.778
$H = 3$	$-1 \le w \le 2$	-2.076	1.991
	$-2 \le w \le 3$	-1.960	1.005
$H = 4$	$-1 \le w \le 2$	-1.761	11.872
	$-2 \le w \le 3$	-2.075	11.856

ranges, whereas the MAE improvements for local extreme values remain positive for about 5%, 1%, and 12%, respectively.

The wider affine range, $-2 \le w \le 3$, does not consistently outperform $-1 \le w \le 2$. Within this exploratory legacy benchmark, the results show improvement for all evaluation points only at $H = 1$, whereas the local-extreme evaluation remains positive for all examined lead times. Because the prespecified parameter-selection procedure did not identify m, these results are not used as primary evidence for the method's effectiveness.

5 Discussion

The experimental results suggest that affine combinations may be useful when the prediction target requires extrapolation, particularly in terms of the median prediction error at local extreme values. For the clean Rössler data and the exploratory Lorenz'63 setting, the affine combinations reduce the median error at recursive local extremes, although the mean improvement is negative in most comparisons. Direct prediction is beneficial only in selected short-term Rössler cases and is otherwise unstable. The extrapolation experiment further demonstrates that affine combinations can predict values outside the observed training range, whereas convex combinations are restricted to interpolation within the convex hull.

These findings are consistent with the properties of the proposed method discussed in Section 3. Remark 1 shows that the predicted values remain bounded when both the training data and the affine weights are bounded. Thus, the proposed method permits extrapolation while preventing unbounded prediction values and can therefore be interpreted as a controlled extrapolation method.

The benefit depends on the prediction step p and the prediction scheme: recursive prediction shows positive median improvements at selected local extremes, whereas direct prediction is unreliable except in selected short-term Rössler cases. The restriction of the direct-prediction benefit to short-term cases is qualitatively consistent with the horizon effect in Subsection 3.5, where the reconstruction gain can be offset by error amplification and Taylor remainders as p increases. The dependence is not uniformly monotonic; for example, the median improvement in recursive Lorenz'63 prediction is larger at the last local extreme value than at the first.

The difference between local extreme values and all evaluation points further clarifies the role of affine weights. Local extremes are likely to occur near boundary-like regions of the reconstructed state space, where extrapolation can be beneficial. In contrast, all evaluation points include many points in the interior of the attractor, where interpolation by convex combinations may already be adequate. This distinction may explain why affine combinations improve predictions at local extremes more consistently than they improve average performance over all evaluation points.

The experiments with observation noise reveal an important limitation of affine combinations. All reported mean improvements are negative for the noisy toy-model data, while positive median improvements remain only under isolated conditions at 0.1% noise. At 0.5% and 1% noise, all reported median improvements are negative. The degradation is neither monotonic nor uniform because observation noise can alter not only the positions of reconstructed points but also the selection and relative arrangement of their nearest neighbors. These changes can destabilize the estimated affine weights, particularly for wider ranges such as $-2 \leq w \leq 3$. Careful selection of the weight range, regularization, or noise reduction may therefore be necessary when the method is applied to noisy real-world data.

The embedding dimension also affects the local neighborhood used for prediction. As shown in Subsection 4.2, the prespecified FNS criterion selected $m = 4$ for Rössler but did not identify a dimension for Lorenz'63 or the flood data within $m = 1, \ldots, 12$. The value $m = 3$ for Lorenz'63 was obtained using conventional FNN and was used only for exploratory analysis. Consequently, the conclusions for each dataset must

be interpreted according to whether the embedding dimension was selected by the primary criterion.

The flood benchmark is likewise exploratory because it uses the legacy setting $m = 3$ and $\tau = 5$. Under this setting, affine combinations improve predictions at local extreme values for the examined lead times, whereas their improvement over all evaluation points is limited and becomes negative at several lead times. Because the primary procedure estimated $\tau = 27$ but did not identify an embedding dimension, these results do not constitute evidence from the parameter-selected primary analysis. Nevertheless, their pattern is consistent with the observed sensitivity of affine combinations to noise and neighborhood perturbations, suggesting that additional stabilization may be required for real-world applications.

6 Conclusion

We proposed a nonlinear time-series prediction method based on affine combinations of neighboring points in the reconstructed state space. By allowing bounded negative weights and weights larger than one, the method extends conventional convex-combination-based prediction and can represent states outside the convex hull of the neighbors. Theoretical analysis establishes bounded predictions for bounded training data and weights and derives a conditional lower bound on the ratio of convex to affine prediction-error norms that accounts for the Taylor remainders. The dependence on the evaluation horizon, prediction scheme, weight range, and observation noise is evaluated empirically.

Experiments on clean Rössler data and an exploratory Lorenz'63 setting showed improved median accuracy at local extremes in selected recursive cases. However, mean improvements were often negative, and direct prediction was generally unstable except in selected short-term Rössler cases. The extrapolation experiment confirmed that the method can predict values outside the observed training range, suggesting potential benefits near boundary-like regions of the reconstructed state space.

The results also revealed substantial limitations. Under observation noise, all mean improvements were negative, and at noise levels of 0.5% and 1%, all median improvements were negative. The wider weight range $-2 \leq w \leq 3$ was generally less stable than $-1 \leq w \leq 2$. For the flood data, neither FNS nor supplementary FNN identified an embedding dimension within $m = 1, \ldots, 12$; therefore, the legacy results with $m = 3$ and $\tau = 5$ should be regarded only as an exploratory benchmark.

Overall, affine-combination-based prediction is not a uniformly superior alternative to convex-combination-based prediction. It may be useful when extrapolation is required, particularly for local-extreme prediction over moderate horizons with reliable neighborhood reconstruction and appropriately constrained weights. Future work should address automatic weight-range selection, prediction-oriented embedding validation, regularization against observation noise, and further evaluation of how reconstructed-state-space geometry affects performance on real-world datasets.

References

[1] Ding, D., Zhang, M., Pan, X., Yang, M., He, X.: Modeling extreme events in

time series prediction. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 1114–1122 (2019) https://doi.org/10.1145/3292500.3330896

[2] Paleologos, E.K., Skitzi, I., Katsifarakis, K., Darivianakis, N.: Neural network simulation of spring flow in karst environments. Stochastic Environmental Research and Risk Assessment **27**(8), 1829–1837 (2013) https://doi.org/10.1007/s00477-013-0717-y

[3] Hirata, Y., Shiro, M.: Improving time series prediction accuracy for the maxima of a flow by reconstructions using local cross sections. Chaos **32**(6), 063103 (2022) https://doi.org/10.1063/5.0092433

[4] Liu, Z., Hirata, Y.: Predicting local maxima of nonlinear time series with a neural network and edit distance. Physica A: Statistical Mechanics and its Applications **677**, 130935 (2025) https://doi.org/10.1016/j.physa.2025.130935

[5] Hirata, Y., Sukegawa, N.: Two efficient calculations of edit distance between marked point processes. Chaos **29**(10), 101107 (2019) https://doi.org/10.1063/1.5125651

[6] Hirata, Y., Shiro, M., Takahashi, N., Aihara, K., Suzuki, H., Mas, P.: Approximating high-dimensional dynamics by barycentric coordinates with linear programming. Chaos **25**(1), 013114 (2015) https://doi.org/10.1063/1.4906746

[7] Mees, A.I.: Dynamical systems and tesselations: Detecting determinism in data. International Journal of Bifurcation and Chaos **1**(4), 777–794 (1991) https://doi.org/10.1142/S0218127491000579

[8] Takens, F.: Detecting strange attractors in turbulence. In: Rand, D.A., Young, L.-S. (eds.) Dynamical Systems and Turbulence, Warwick 1980, Lecture Notes in Mathematics, vol. 898, pp. 366–381. Springer, Berlin, Heidelberg (1981). https://doi.org/10.1007/BFb0091924

[9] Lyapunov, A.M.: The General Problem of the Stability of Motion (in Russian). Zilberberg Printing House, Kharkov (1892). Doctoral dissertation, University of Kharkov; published by the Kharkov Mathematical Society

[10] Rössler, O.E.: An equation for continuous chaos. Physics Letters A **57**(5), 397–398 (1976) https://doi.org/10.1016/0375-9601(76)90101-8

[11] Lorenz, E.N.: Deterministic nonperiodic flow. Journal of the Atmospheric Sciences **20**(2), 130–141 (1963) https://doi.org/10.1175/1520-0469(1963)020⟨0130:DNF⟩2.0.CO;2

[12] Dawson, C.W., See, L.M., Abrahart, R.J., Wilby, R.L., Shamseldin, A.Y., Anctil, F., Belbachir, A.N., Bowden, G., Dandy, G., Lauzon, N., Maier, H.: A

comparative study of artificial neural network techniques for river stage forecasting. In: Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005, vol. 4, pp. 2666–2670. IEEE, Montreal, Quebec, Canada (2005). https://doi.org/10.1109/IJCNN.2005.1556324

[13] Harris, C.R., Millman, K.J., van der Walt, S.J., *et al.*: Array programming with NumPy. Nature **585**(7825), 357–362 (2020) https://doi.org/10.1038/s41586-020-2649-2

[14] Virtanen, P., Gommers, R., Oliphant, T.E., *et al.*: SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature Methods **17**(3), 261–272 (2020) https://doi.org/10.1038/s41592-019-0686-2

[15] Dormand, J.R., Prince, P.J.: A family of embedded Runge–Kutta formulae. Journal of Computational and Applied Mathematics **6**(1), 19–26 (1980) https://doi.org/10.1016/0771-050X(80)90013-3

[16] Kantz, H.: A robust method to estimate the maximal lyapunov exponent of a time series. Physics Letters A **185**(1), 77–87 (1994) https://doi.org/10.1016/0375-9601(94)90991-1

[17] Fraser, A.M., Swinney, H.L.: Independent coordinates for strange attractors from mutual information. Physical Review A **33**(2), 1134–1140 (1986) https://doi.org/10.1103/PhysRevA.33.1134

[18] Kraskov, A., Stögbauer, H., Grassberger, P.: Estimating mutual information. Physical Review E **69**(6), 066138 (2004) https://doi.org/10.1103/PhysRevE.69.066138

[19] Kennel, M.B., Abarbanel, H.D.I.: False neighbors and false strands: A reliable minimum embedding dimension algorithm. Physical Review E **66**(2), 026209 (2002) https://doi.org/10.1103/PhysRevE.66.026209

[20] Kennel, M.B., Brown, R., Abarbanel, H.D.I.: Determining embedding dimension for phase-space reconstruction using a geometrical construction. Physical Review A **45**(6), 3403–3411 (1992) https://doi.org/10.1103/PhysRevA.45.3403

Statements and Declarations

Funding

This research is supported by Japan Society for the Promotion of Science (JSPS) KAKENHI Grant No. JP26K00013. We also thank Katsuhiko Hirata and Eiko Hirata for their financial donations.

Competing Interests

The authors declare that they have no conflict of interest related to this paper.

Author Contributions

N.F. and Y.H. conceived and designed the study. N.F. wrote the code and performed the numerical experiments. T.K. provided important feedback. All the authors interpreted the results of the numerical experiments. N.F. wrote the initial draft. All the authors reviewed/revised the draft and agreed to submit the final version of the draft.

Data Availability

The fixed synthetic datasets and source code used in this study are provided in Online Resource 1. The flood observations described by Dawson et al. [12] are not included because the authors do not have permission to redistribute them. Enquiries about access should be directed to the data provider, Christian W. Dawson; access is subject to the provider's permission.

Online Resource 1 (ESM_1.zip). Python source code, fixed synthetic inputs, parameter settings, numerical reference results, and a README for the forecasting experiments. Flood observations are not included.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ESM1.zip