Supplementary Materials for

Information leverage in interconnected ecosystems: Overcoming the curse of dimensionality

Hao Ye and George Sugihara*

*Corresponding author. Email: gsugihara@ucsd.edu

Published 26 August 2016, *Science* **353**, 922 (2016)
DOI: 10.1126/science.aag0863

This PDF file includes:

Other Supplementary Material for this manuscript includes the following:
(available at www.sciencemag.org/content/353/6302/922/suppl/DC1)

Data S1 (Excel file)

Materials and Methods

Data Sources

We used time series generated from 4 different models (described below) and data collected from a long-term mesocosm experiment using a plankton community isolated from the Baltic Sea *(20, 25-28)*.

Ecosystem Models

Coupled Logistic

We modeled 3 interacting species using the coupled logistic map:

$$\begin{bmatrix} x_{t+1} \\ y_{t+1} \\ z_{t+1} \end{bmatrix} = \boldsymbol{r} \circ \begin{bmatrix} x_t \\ y_t \\ z_t \end{bmatrix} \circ \left(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \boldsymbol{A} \times \begin{bmatrix} x_t \\ y_t \\ z_t \end{bmatrix} \right) \quad \textbf{[1]}$$

where $\boldsymbol{r}$ is a 3×1 vector, $\boldsymbol{A}$ is a 3×3 matrix with $A_{i,i} = 1$, and ∘ is the Schur product (entrywise product). We used initial conditions $\begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} = \begin{bmatrix} 0.2 \\ 0.2 \\ 0.2 \end{bmatrix}$, $\boldsymbol{r} = \begin{bmatrix} 3.6 \\ 3.0 \\ 3.0 \end{bmatrix}$, and $\boldsymbol{A} = \begin{bmatrix} 1 & 0.2 & 0.2 \\ 0.2 & 1 & -0.2 \\ 0.2 & -0.2 & 1 \end{bmatrix}$.

3-species Food Chain

We modeled a 3-species food chain following Hastings & Powell *(20)*:

$$\begin{aligned} dx/dt &= x(1-x) - f_1(x)y \\ dy/dt &= f_1(x)y - f_2(y)z - d_1 y \\ dz/dt &= f_2(y)z - d_2 z \end{aligned} \quad \textbf{[2]}$$

where

$$f_i(u) = a_i u / (1 + b_i u).$$

We used the parameterization $a_1 = 2.5$, $b_1 = 3.2$, $a_2 = 0.1$, $b_2 = 2.0$, $d_1 = 0.2$, and $d_2 = 0.015$; and initial conditions $x_0 = 0.8$, $y_0 = 0.2$, and $z_0 = 8$.

Flour Beetle Model

We modeled 3 life stages (larvae, pupae, and adults) of the flour beetle, *Tribolium castaneum* following Dennis et al. *(25)*:

$$\begin{aligned} L_{t+1} &= bA_t \exp\left(-c_{\text{el}} L_t - c_{\text{ea}} A_t\right) \\ P_{t+1} &= L_t\left(1-\mu_{\text{l}}\right) \\ A_{t+1} &= P_t \exp\left(-c_{\text{pa}} A_t\right) + A_t\left(1-\mu_{\text{a}}\right) \end{aligned} \qquad \textbf{[3]}$$

using the parameterization $b = 10.67$, $\mu_{\text{l}} = 0.1955$, $\mu_{\text{a}} = 0.96$, $c_{\text{el}} = 0.01647$, $c_{\text{ea}} = 0.01313$, and $c_{\text{pa}} = 0.35$, which are the maximum-likelihood estimates of the model parameters from real data, but with c_{pa} adjusted to give chaotic dynamics. Initial conditions were $L_1 = 250$, $P_1 = 5$, and $A_1 = 100$.

12-species Resource Competition Model

We modeled a 12-species system that competes on 5 shared resources, following Huisman & Weissing (*26*):

$$\begin{aligned} dN_i/dt &= N_i\left(\mu_i\left(R_1,\ldots,R_k\right) - m_i\right) \\ dR_j/dt &= D\left(S_j - R_j\right) - \sum_{i=1}^{n} c_{ij}\mu_i\left(R_1,\ldots,R_k\right)N_i \end{aligned} \qquad \textbf{[4]}$$

where

$$\mu\left(R_1,\ldots,R_k\right) = \min\left(\frac{r_i R_1}{K_{1i}+R_1},\ldots,\frac{r_i R_k}{K_{ki}+R_k}\right)$$

We used the parameterization $r_i = 1$, $m_i = D = 0.25$, $\text{S}_1 = 6$, $\text{S}_2 = 10$, $\text{S}_3 = 14$, $\text{S}_4 = 4$, $\text{S}_5 = 9$,

$$K = \begin{bmatrix} 0.39 & 0.34 & 0.30 & 0.24 & 0.23 & 0.41 & 0.20 & 0.45 & 0.14 & 0.15 & 0.38 & 0.28 \\ 0.22 & 0.39 & 0.34 & 0.30 & 0.27 & 0.16 & 0.15 & 0.05 & 0.38 & 0.29 & 0.37 & 0.31 \\ 0.27 & 0.22 & 0.39 & 0.34 & 0.30 & 0.07 & 0.11 & 0.05 & 0.38 & 0.41 & 0.24 & 0.25 \\ 0.30 & 0.24 & 0.22 & 0.39 & 0.34 & 0.28 & 0.12 & 0.13 & 0.27 & 0.33 & 0.04 & 0.41 \\ 0.34 & 0.30 & 0.22 & 0.20 & 0.39 & 0.40 & 0.50 & 0.26 & 0.12 & 0.29 & 0.09 & 0.16 \end{bmatrix}, \text{ and}$$

$$C = \begin{bmatrix} 0.04 & 0.04 & 0.07 & 0.04 & 0.04 & 0.22 & 0.10 & 0.08 & 0.02 & 0.17 & 0.25 & 0.03 \\ 0.08 & 0.08 & 0.08 & 0.10 & 0.08 & 0.14 & 0.22 & 0.04 & 0.18 & 0.06 & 0.20 & 0.04 \\ 0.10 & 0.10 & 0.10 & 0.10 & 0.14 & 0.22 & 0.24 & 0.12 & 0.03 & 0.24 & 0.17 & 0.01 \\ 0.05 & 0.03 & 0.03 & 0.03 & 0.03 & 0.09 & 0.07 & 0.06 & 0.03 & 0.03 & 0.11 & 0.05 \\ 0.07 & 0.09 & 0.07 & 0.07 & 0.07 & 0.05 & 0.24 & 0.05 & 0.08 & 0.10 & 0.02 & 0.04 \end{bmatrix}$$; with

initial conditions $R_j = S_j$, $N_i = 0.1 + i/100$ for species 1-5. Species 6-8 were introduced at time t = 1,000 with $N_i = 0.1$; species 9 and 10 were introduced at time t = 3,000 with $N_i = 0.1$; and species 11 and 12 were introduced at time t = 5,000 with $N_i = 0.1$.

Time Series Generation

The coupled logistic map and flour beetle model were run forward in time following equations 1 and 3, respectively. The 3-species food chain model was run by solving equation 2 using the classical Runge-Kutta method with a time step of 0.01 and downsampling by a factor of 800. The 12-species resource competition model was run by solving equation 4 using the classical Runge-Kutta method with a time step of 0.01 and downsampling by a factor of 500. For each model, we generated time series with 3000 points (starting at t = 5,001 for the 12-species resource competition model and t = 1 otherwise), and rescaled so that each time series had mean = 0 and variance = 1. Observation error was simulated by adding i.i.d. Gaussian noise with mean = 1 and variance = 0 to 0.3 in increments of 0.05.

Plankton Community

This experiment has previously been described in numerous publications as an example of chaotic dynamics. We used the raw data from the supplement of Benincà et al. *(28)* and focused on forecasting the abundances of rotifers (mainly *Brachionus plicatilis*) and the calanoid copepod *Eurytemora affinis*. Benincà et al. *(28)* noted that the primary food items of the rotifers and calanoid copepods were picocyanobacteria and nanoflagellates (Fig. 4A). We used convergent cross mapping *(5)* to verify that causal information about the prey abundances were

present in the predator abundances (Fig. 4B). Thus multivariate attractor reconstructions using the prey time series are suitable for forecasting the predators (rotifers and calanoid copepods).

Instead of interpolating the data (which can pollute the dynamics by combining information from multiple sampling times), we instead extracted all possible segments of 15 points or more and where the lag was 6-8 days (~ 1 week). This procedure resulted in 17 segments, comprising 725 data points. Next, we applied a fourth root transformation as suggested in Benincà et al. *(28)* to suppress sharp peaks that distort the attractor reconstruction (especially when searching for nearest neighbors) before applying the rescaling described above. Forecast skill was computed after back-transforming into the same scale as the raw abundance data.

Attractor Reconstructions

For each system, we generated all possible 3-dimensional attractor reconstructions where each coordinate could be any of the time series variables with a lag of 0, τ, or 2τ. Then we kept only the reconstructions where at least one coordinate had no lag, resulting in 64 reconstructions for the simple model systems and 5116 reconstructions for the 12-species system. For all models, we used $\tau = 1$. Note that for the 3-species food chain model, τ is effectively 8, because the model was run using a time step of 0.01 and downsampled by 800. For the 12-species resource competition model, τ is effectively 5, because the model was run using a time step of 0.01 and downsampled by 500. For the mesocosm data $\tau = 1$ corresponds to ~1 week.

In the general case with n time series variables, l possible lags for each variable, and embedding dimension E, the number of possible attractor reconstructions is given by

$$m := \binom{nl}{E} - \binom{n(l-1)}{E}$$

where $\binom{nl}{E}$ is the number of attractors formed by choosing E of the nl possible coordinates, and $\binom{n(l-1)}{E}$ is the number of "invalid" attractors where all E coordinates are chosen from among the $n(l\text{-}1)$ possible coordinates with positive (i.e. nonzero) lag.

Multiview Embedding

Using the method of simplex projection *(13)* and leave-one-out cross-validation, we computed the performance of each attractor reconstruction when forecasting 1 time step ahead on an in-sample portion of the data, called the library. The attractors were ranked based on ϱ, the correlation coefficient between observations and predictions. The top attractors were then used to produce forecasts for the out-of sample portion of the data. The multiview embedding forecast is defined to be the arithmetic mean of the forecasts from the top attractors:

$$\hat{y}_{t+p} = \frac{1}{k}\sum_{i=1}^{k} y_{nn^i(t)+p}$$

where k is the number of attractors to average over, $nn^i(t)$ is the time index of the nearest neighbor in the i-th attractor at time t, and $p = 1$ is the prediction horizon. Following *(21)*, we used the simple heuristic of averaging the top $k = \sqrt{m}$ reconstructions for MVE where m is the number of candidate reconstructions.

Conventional simplex projection *(13)* uses $b = E + 1$ nearest neighbors from a single attractor reconstruction, and each of these neighbors represents a unique historical state. These neighbors are weighted by their distance, which is influenced by observational error, so forecasts can be quite sensitive to noise. However, multiview embedding uses the single nearest neighbor from k attractor reconstructions. As these points potentially represent the same historical state and have different observation errors, averaging over them will give a larger weighting towards

the points that are consistently closer to the actual system state, in a manner similar to the weighted average used in simplex projection.

Performance Metrics

Forecast skill was measured using 3 different metrics: ϱ, the correlation coefficient between observations and predictions; MAE, mean absolute error; and RMSE, root mean square error. ϱ captures forecast accuracy, how well the predictions line up with observations. Both MAE and RMSE capture forecast error, the "average" deviation of predictions from observations.

For the model data, we sampled 100 libraries from the generated time series by choosing the beginning of each library from the uniform distribution {501, 2001}. (Note that we exclude the first 500 points from each time series, so as to reduce the effect of transient behavior.) Out-of-sample forecasts were then produced for last 500 points of the time series {2501-3000}. The libraries ranged in size from 25 to 100 vectors (25, 50, 75, 100); thus, there was no overlap between the library and forecasts, because the last possible point in the library is at time 2100 and the first point to be forecast is at time 2501. For the mesocosm data, we used a 4-fold cross-validation scheme where 1/4 of the data was held out of sample and the remaining 3/4 was the "library". Here, the 17 segments were divided into 4 approximately equal sized subsets. The forecasts for each quarter of the data were then combined before computing statistics. (Note that although the statistics are computed once over the whole time series, the attractors were ranked separate for each 1/4 of the data, so that all forecasts are out-of-sample.)

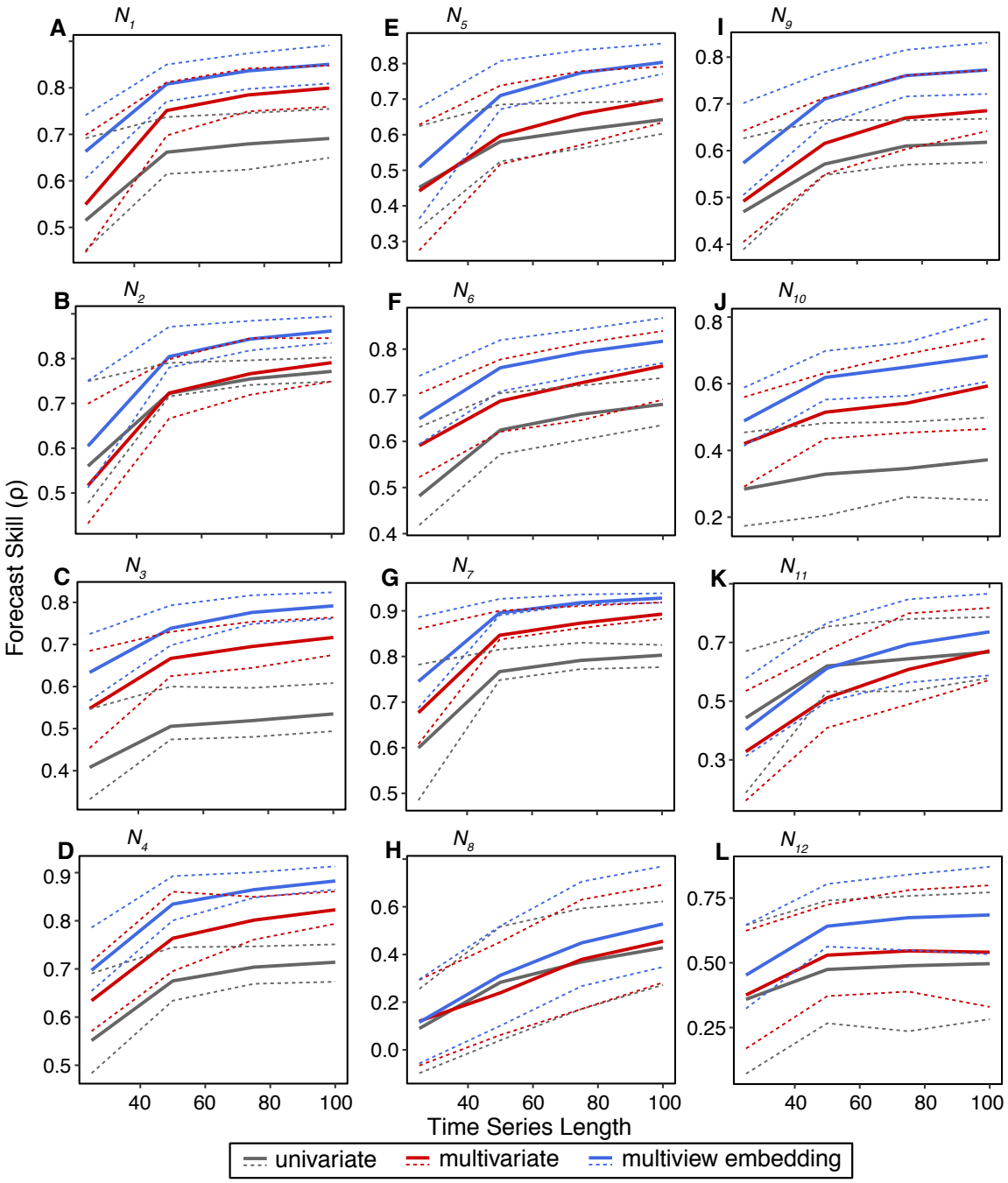

Fig. S1. Comparison of forecast skill for univariate, multivariate, and MVE methods for the 12-species resource competition model (*26*) with 30% observational error. **(A-L)** Forecast accuracy (ϱ, correlation between observations and predictions) vs. time series length for variables N_1 – N_{12}. Solid lines indicate average values over 100 randomly sampled libraries and dashed lines denote upper and lower quartiles.

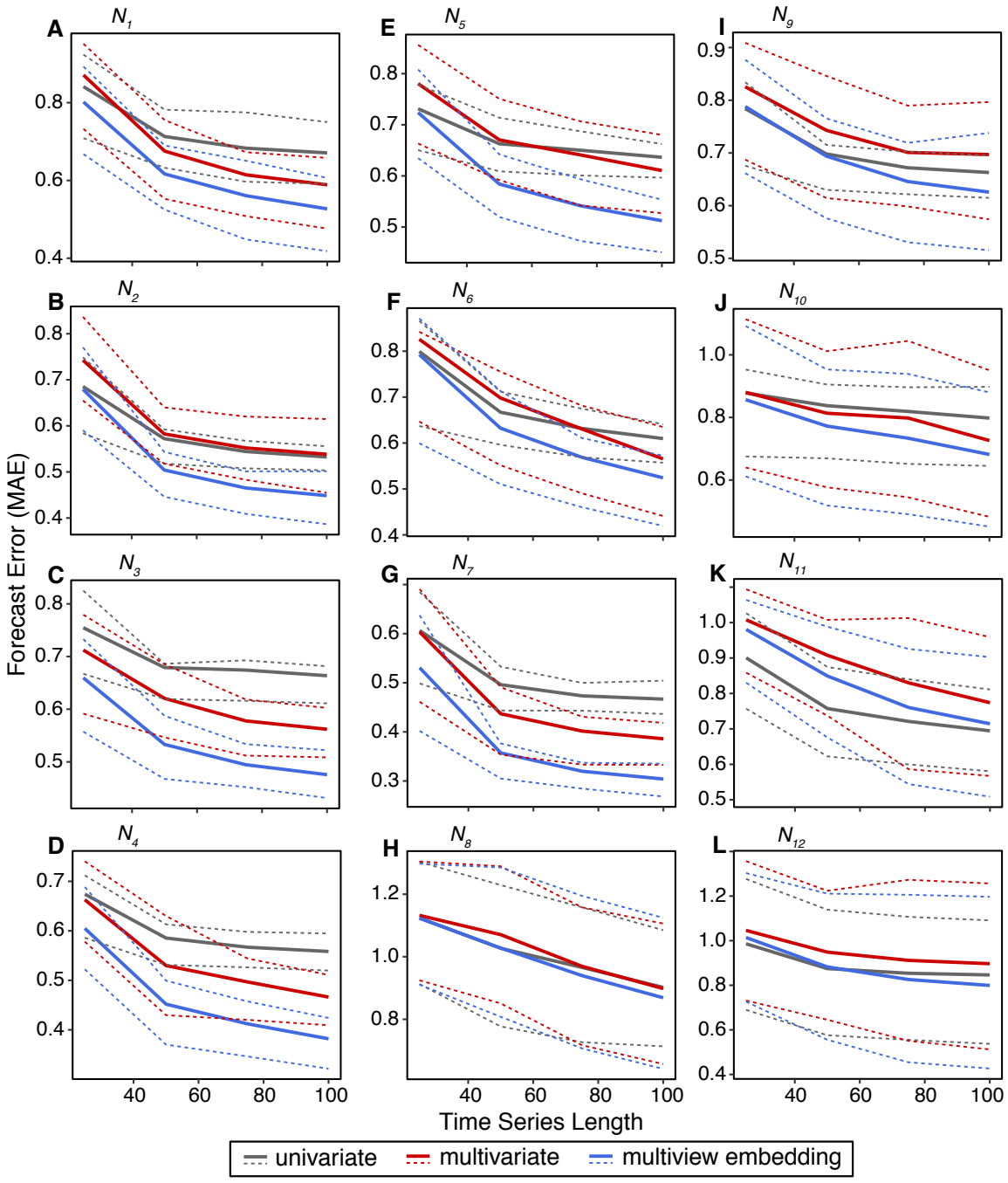

Fig. S2. Comparison of forecast skill for univariate, multivariate, and MVE methods for the 12-species resource competition model (*26*) with 30% observational error. **(A-L)** Forecast error (MAE, mean absolute error) vs. time series length for variables $N_1 - N_{12}$. Solid lines indicate average values over 100 randomly sampled libraries and dashed lines denote upper and lower quartiles.

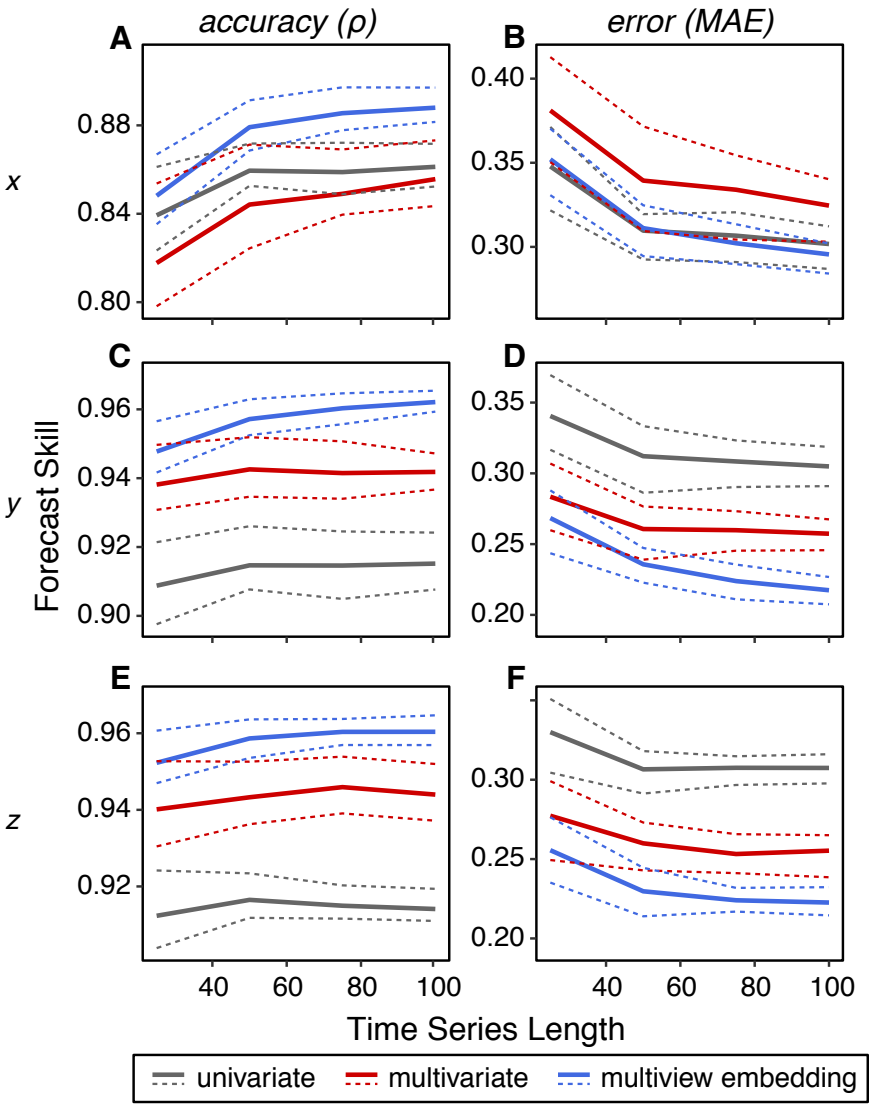

Fig. S3. Comparison of forecast performance for the 3-species logistic map with 10% observational error. Multiview embedding generally produces more precise forecasts than the multivariate and univariate methods. **(A-B)** Forecast accuracy (ϱ, correlation between observations and predictions), and forecast errors (MAE, mean absolute error) vs. library size when predicting variable *x*. Solid lines indicate average values over 100 randomly sampled libraries and dashed lines denote upper and lower quartiles. **(C-D)** Same as (A-B), but for variable *y*. **(E-F)** Same as (A-B), but for variable *z*.

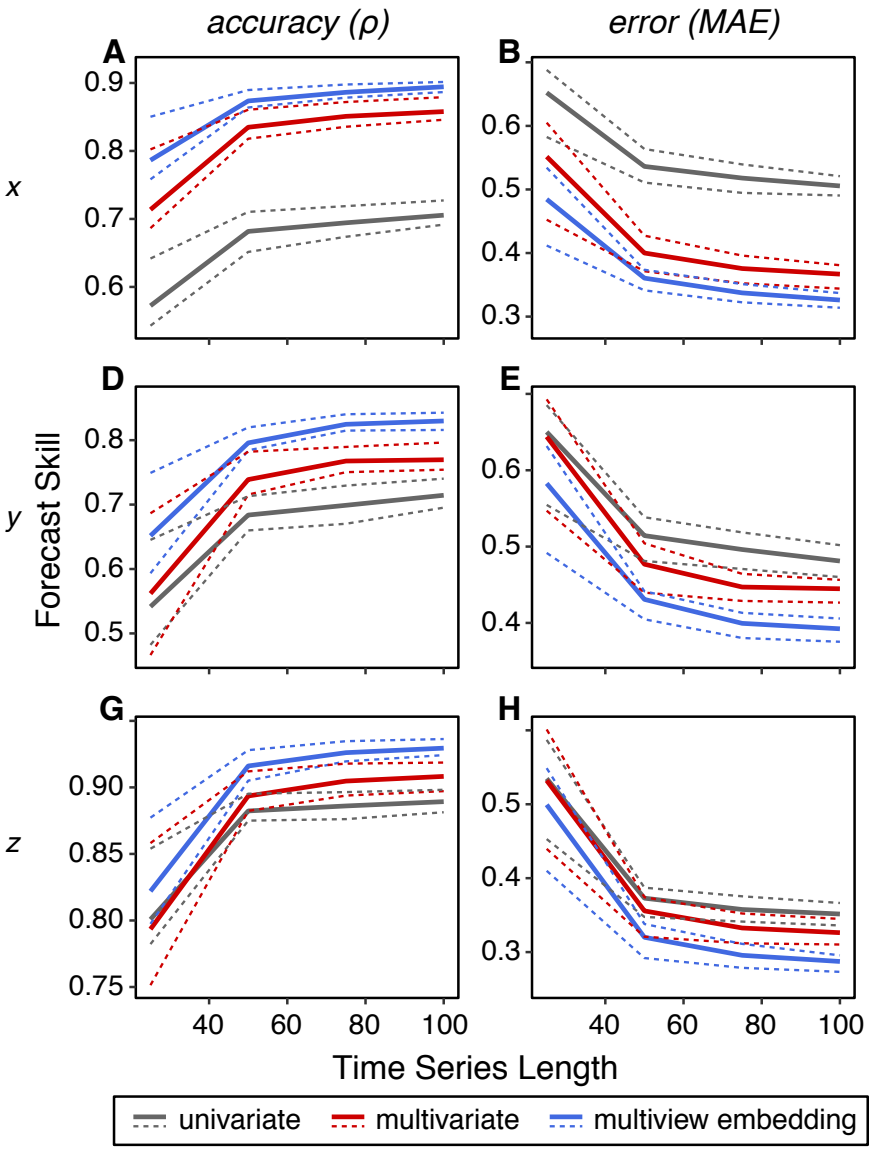

Fig. S4. Comparison of forecast performance for the 3-species food chain model (*20*) with 10% observational error. Multiview embedding generally produces more precise forecasts than the multivariate and univariate methods. **(A-B)** Forecast accuracy (ϱ, correlation between observations and predictions), and forecast errors (MAE, mean absolute error) vs. library size when predicting variable *x*. Solid lines indicate average values over 100 randomly sampled libraries and dashed lines denote upper and lower quartiles. **(C-D)** Same as (A-B), but for variable *y*. **(E-F)** Same as (A-B), but for variable *z*.

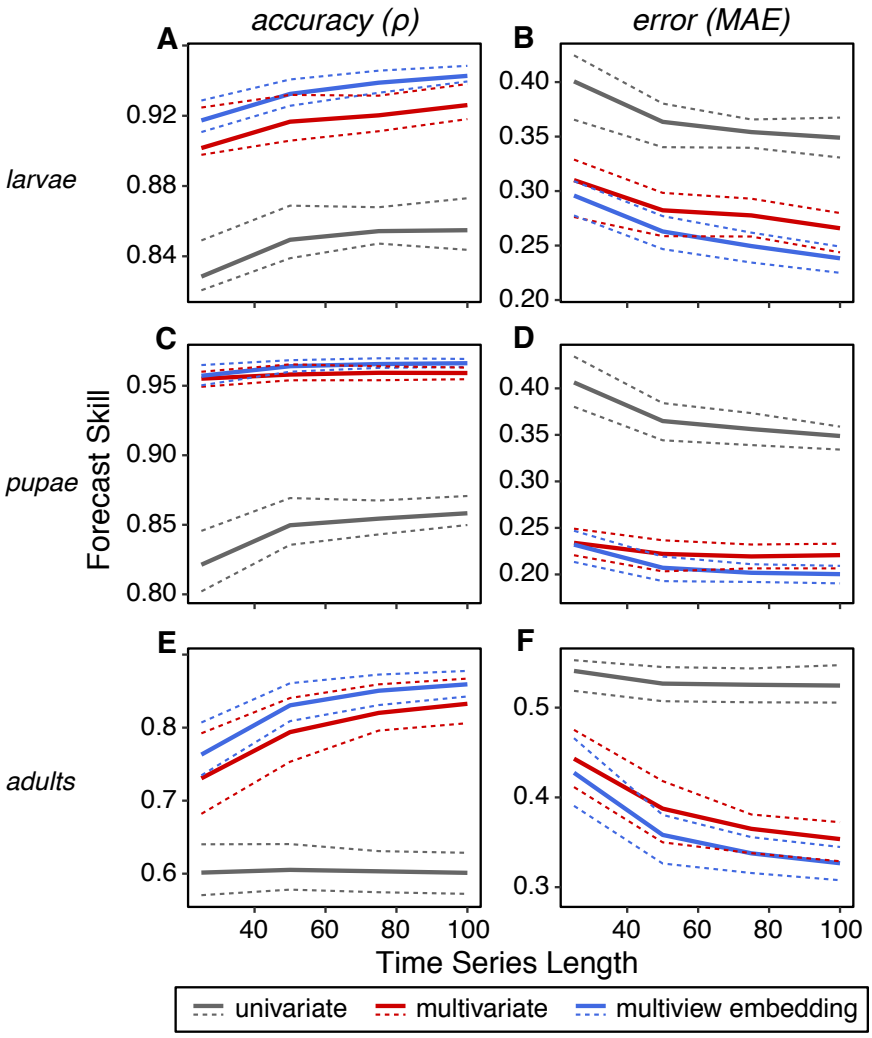

Fig. S5. Comparison of forecast performance for the flour beetle model (*25*) with 10% observational error. Multiview embedding generally produces more precise forecasts than the multivariate and univariate methods. **(A-B)** Forecast accuracy (ϱ, correlation between observations and predictions), and forecast errors (MAE, mean absolute error) vs. library size when predicting larvae. Solid lines indicate average values over 100 randomly sampled libraries and dashed lines denote upper and lower quartiles. **(C-D)** Same as (A-B), but for pupae. **(E-F)** Same as (A-B), but for adults.

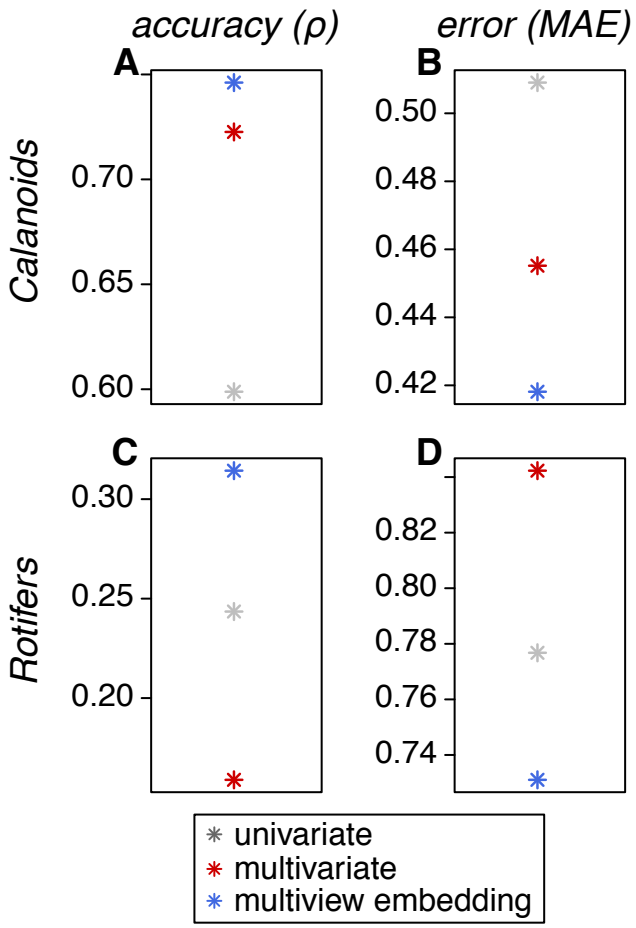

Fig. S6. Comparison of forecast performance for the long-term mesocosm experiment (*27, 28*). Multiview embedding produces more precise forecasts than the multivariate and univariate methods. **(A-B)** Forecast accuracy (ϱ, correlation between observations and predictions) and forecast errors (MAE, mean absolute error) when predicting calanoid copepods. **(C-D)** Same as (A-B), but for rotifers.

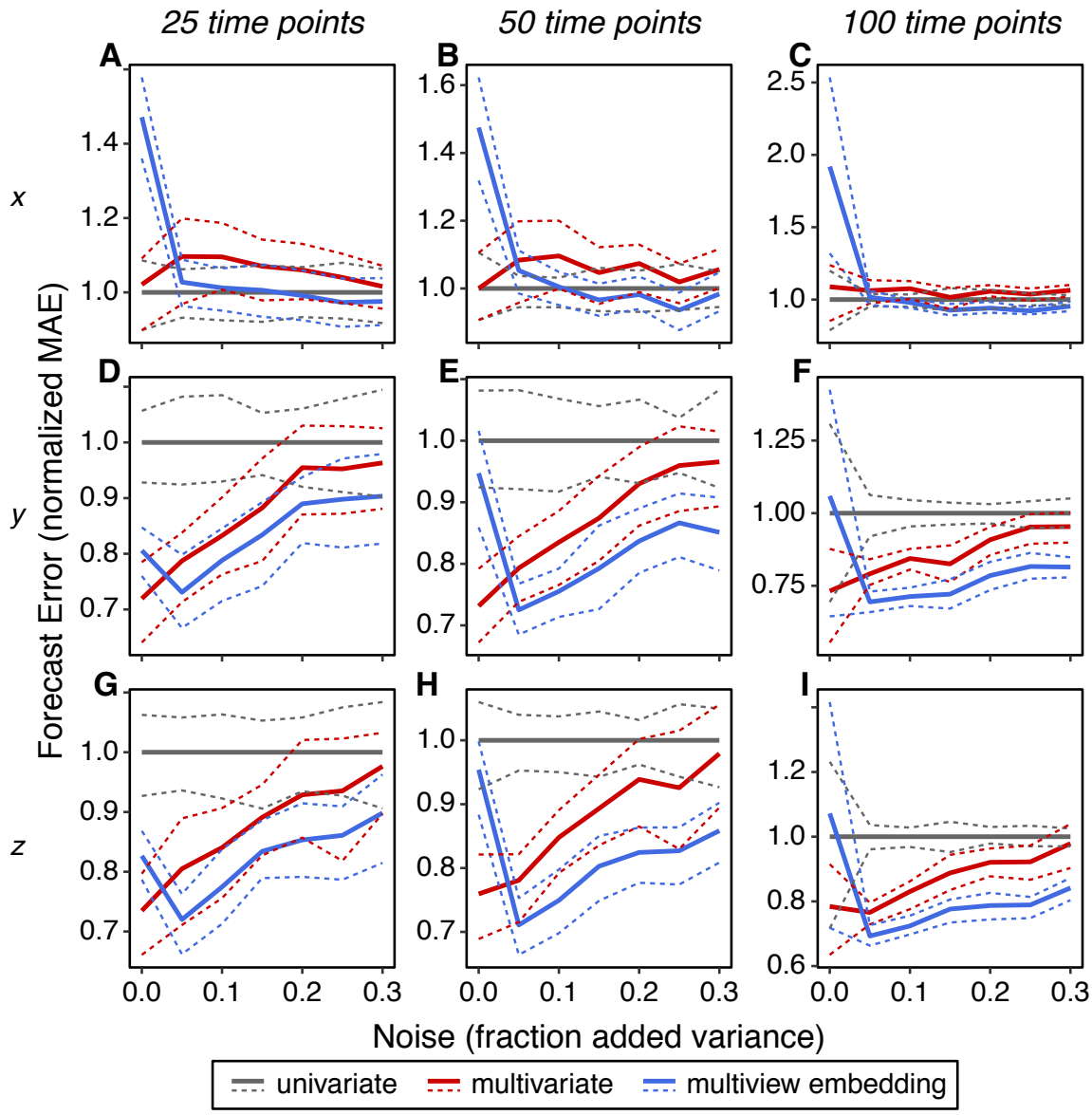

Fig. S7. Normalized forecast error vs. degree of noise for the 3-species logistic map. Other than the case with no observational noise, forecast errors for multiview embedding are generally better (i.e. lower) than errors for the multivariate and univariate methods. **(A-C)** Normalized MAE (i.e. scaled so that mean univariate MAE = 1) vs. degree of noise when predicting variable *x* using 25, 50, or 100 time points. Solid lines indicate average values over 100 randomly sampled libraries and dashed lines denote upper and lower quartiles. **(D-F)** Same as (A-C), but for variable *y*. **(G-I)** Same as (A-C), but for variable *z*.

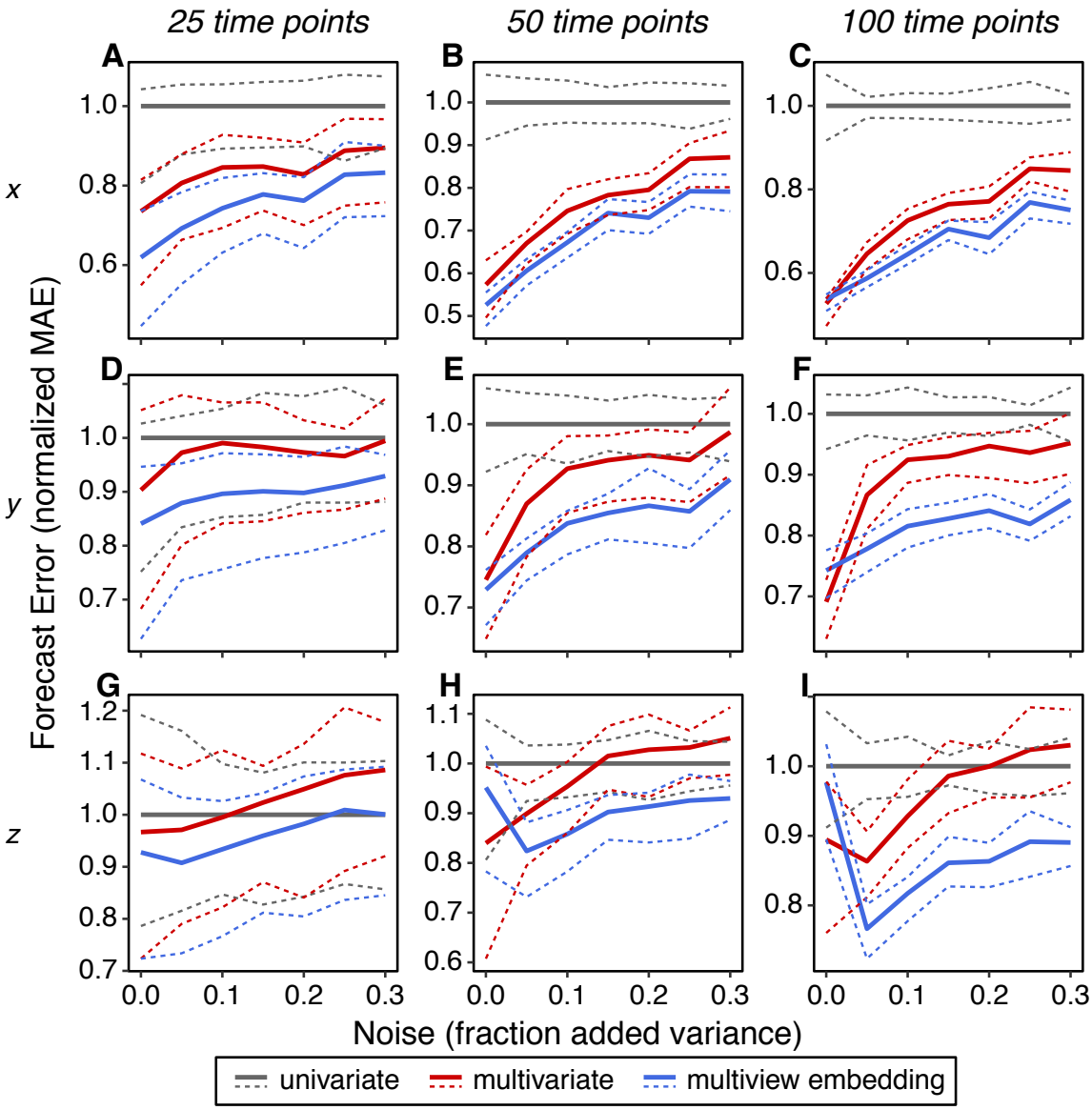

Fig. S8. Normalized forecast error vs. degree of noise for the 3-species food chain model (*20*). Forecast errors for multiview embedding are nearly always better (i.e. lower) than errors for the multivariate and univariate methods. **(A-C)** Normalized MAE (i.e. scaled so that mean univariate MAE = 1) vs. degree of noise when predicting variable *x* using 25, 50, or 100 time points. Solid lines indicate average values over 100 randomly sampled libraries and dashed lines denote upper and lower quartiles. **(D-F)** Same as (A-C), but for variable *y*. **(G-I)** Same as (A-C), but for variable *z*.

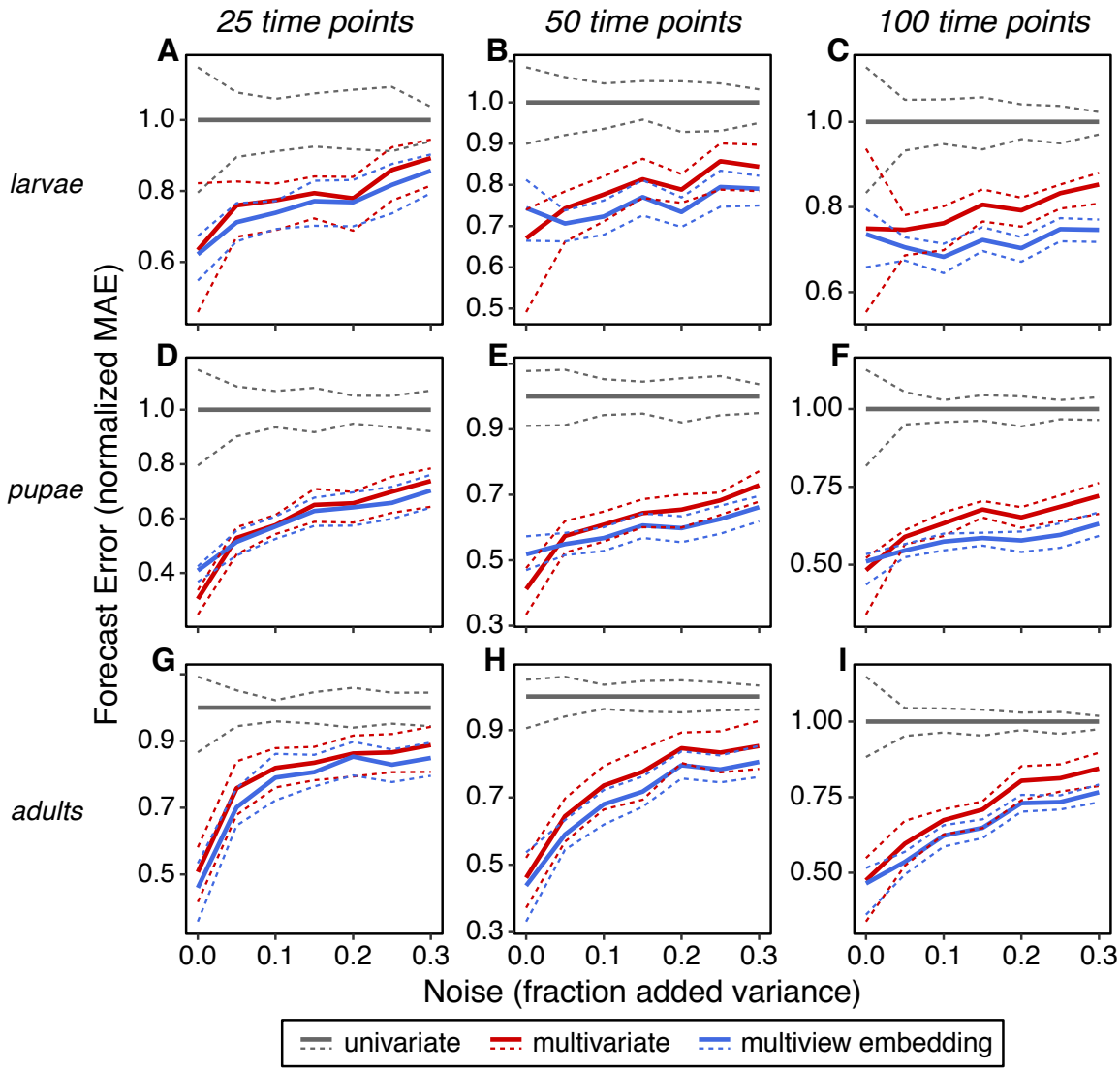

Fig. S9. Normalized forecast error vs. degree of noise for the flour beetle model (*25*). Forecast errors for multiview embedding are nearly always better (i.e. lower) than errors for the multivariate and univariate methods. **(A-C)** Normalized MAE (i.e. scaled so that mean univariate MAE = 1) vs. degree of noise when predicting larvae using 25, 50, or 100 time points. Solid lines indicate average values over 100 randomly sampled libraries and dashed lines denote upper and lower quartiles. **(D-F)** Normalized MAE vs. degree of noise when predicting pupae. **(G-I)** Normalized MAE vs. degree of noise when predicting adults.

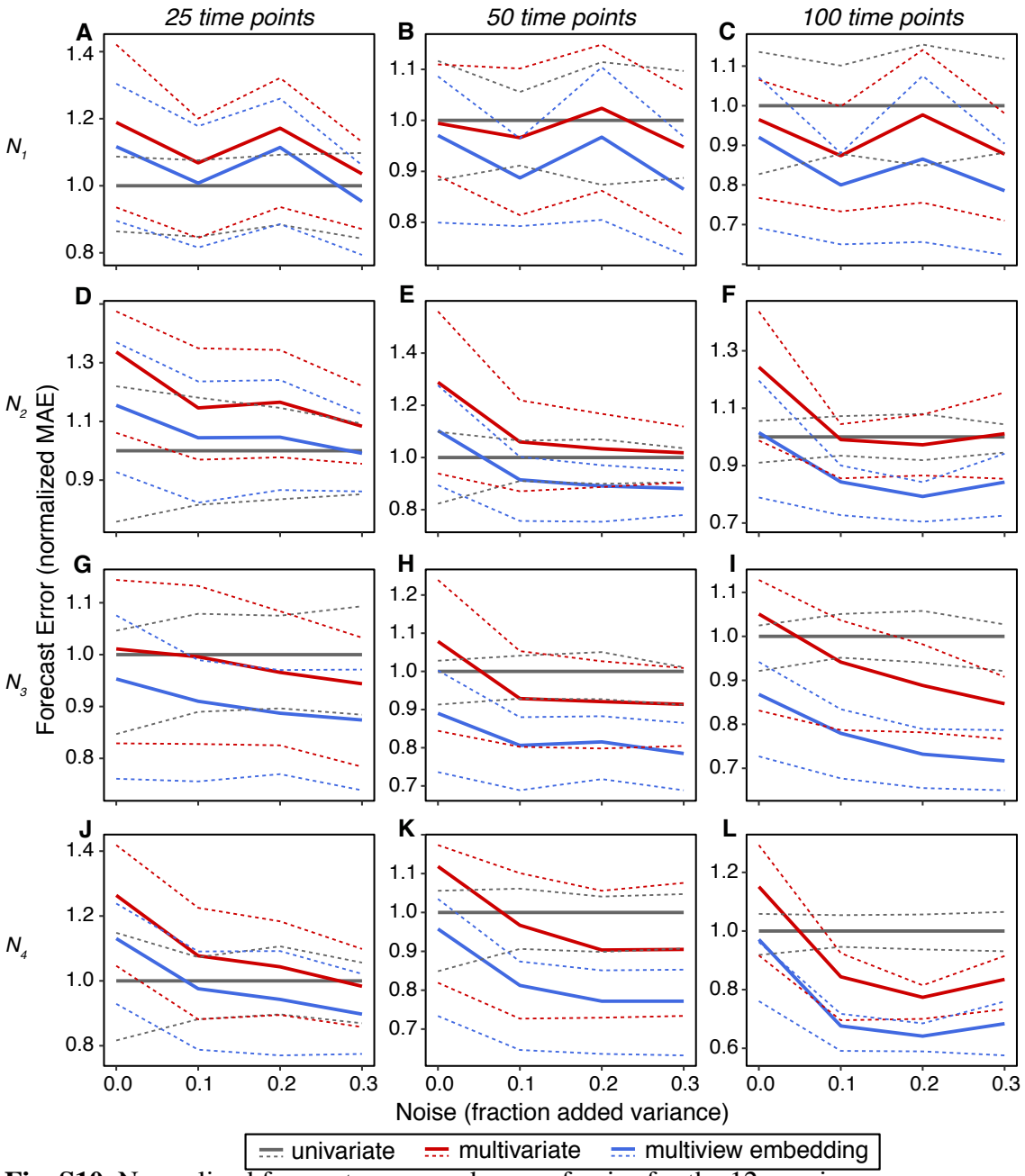

Fig. S10. Normalized forecast error vs. degree of noise for the 12-species resource competition model (*26*). In most cases, forecast errors for multiview embedding are better (i.e. lower) than errors for the multivariate and univariate methods. **(A-C)** Normalized MAE (i.e. scaled so that mean univariate MAE = 1) vs. degree of noise when predicting N_1 using 25, 50, or 100 time points. Solid lines indicate average values over 100 randomly sampled libraries and dashed lines denote upper and lower quartiles. **(D-F)** Same as (A-C), but for N_2. **(G-I)** Same as (A-C), but for N_3. **(J-L)** Same as (A-C), but for N_4.

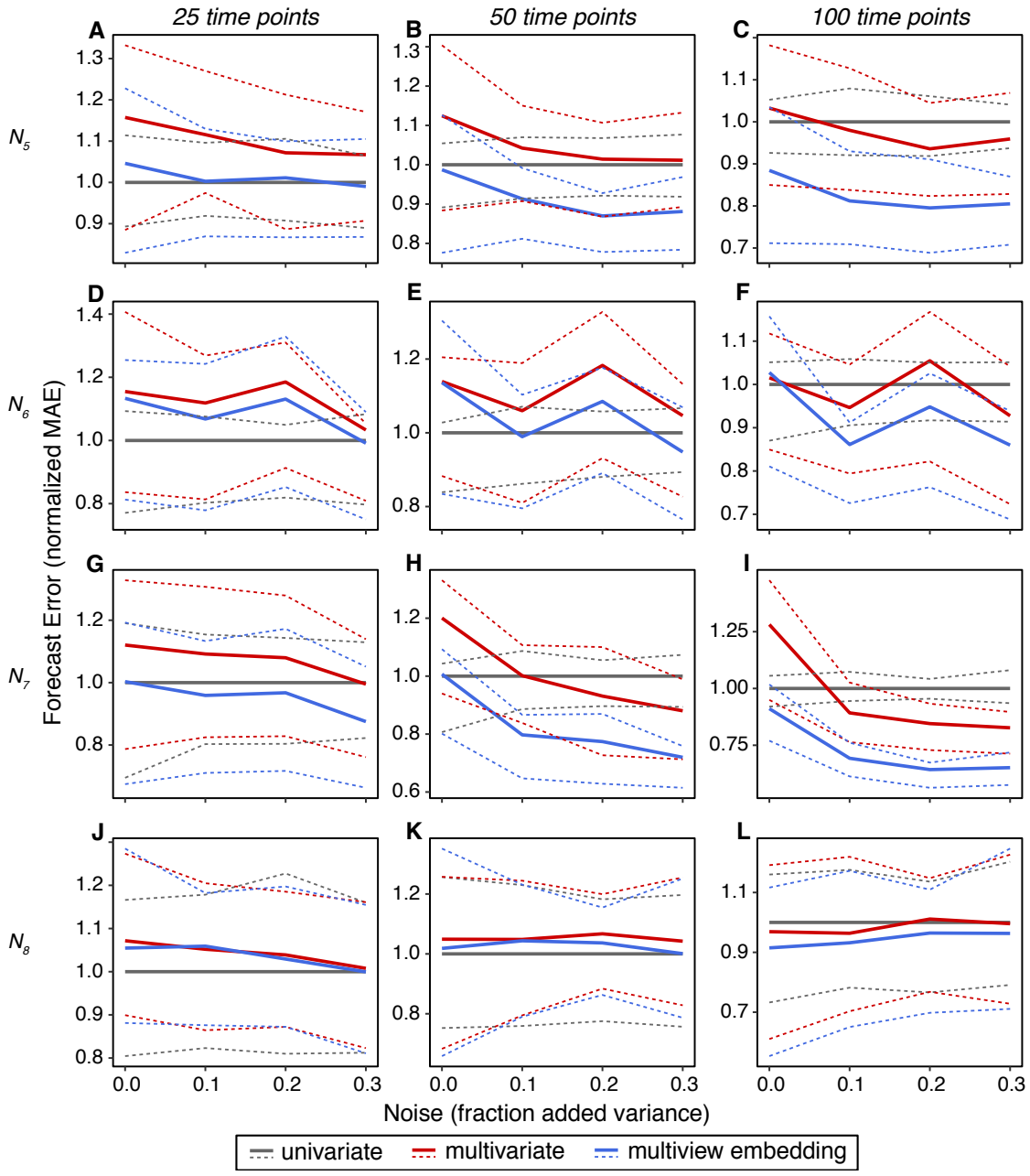

Fig. S11. Normalized forecast error vs. degree of noise for the 12-species resource competition model (*26*). In most cases, forecast errors for multiview embedding are better (i.e. lower) than errors for the multivariate and univariate methods. **(A-C)** Normalized MAE (i.e. scaled so that mean univariate MAE = 1) vs. degree of noise when predicting N_5 using 25, 50, or 100 time points. Solid lines indicate average values over 100 randomly sampled libraries and dashed lines denote upper and lower quartiles. **(D-F)** Same as (A-C), but for N_6. **(G-I)** Same as (A-C), but for N_7. **(J-L)** Same as (A-C), but for N_8.

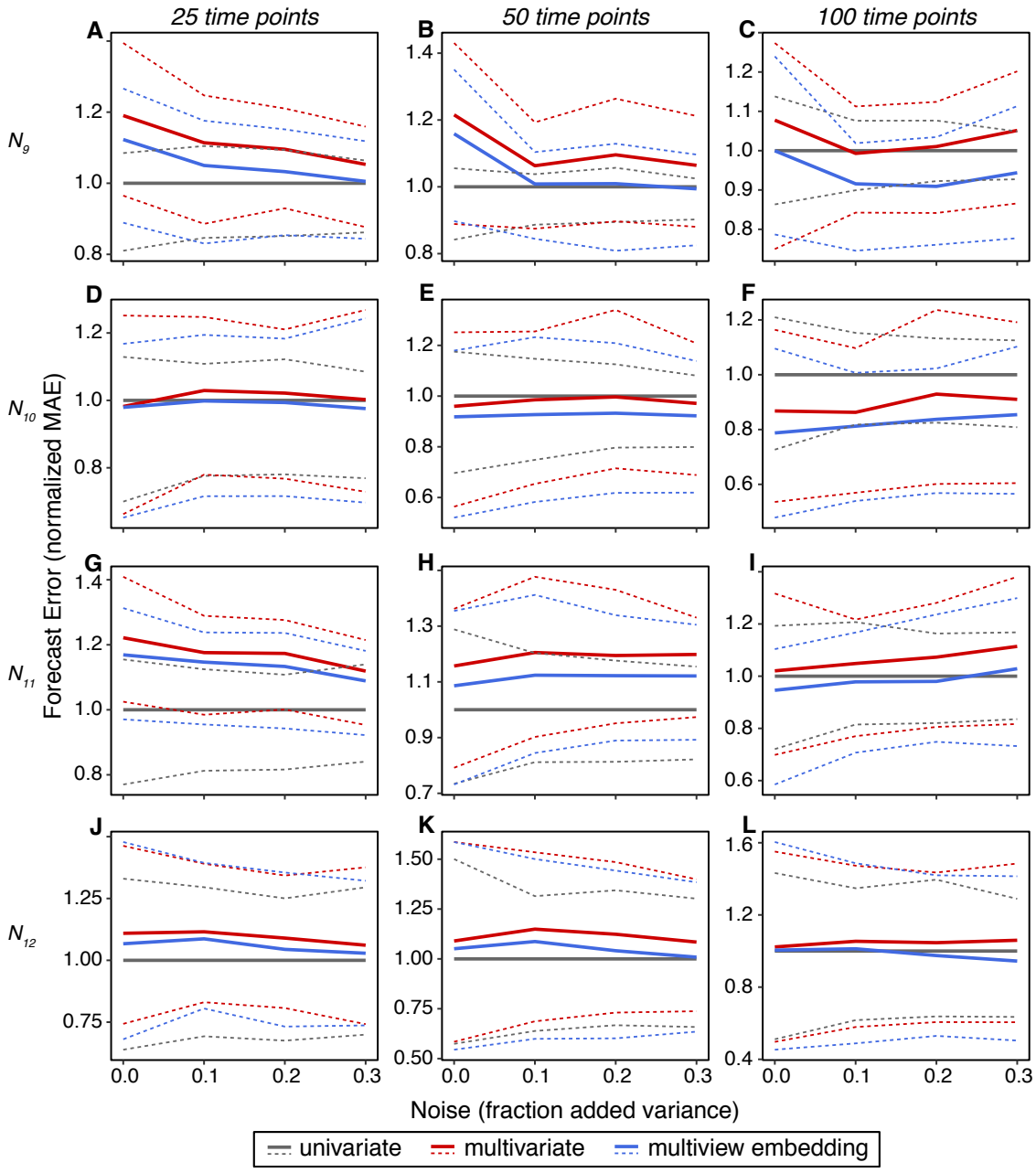

Fig. S12. Normalized forecast error vs. degree of noise for the 12-species resource competition model (*26*). In most cases, forecast errors for multiview embedding are better (i.e. lower) than errors for the multivariate and univariate methods. **(A-C)** Normalized MAE (i.e. scaled so that mean univariate MAE = 1) vs. degree of noise when predicting N_9 using 25, 50, or 100 time points. Solid lines indicate average values over 100 randomly sampled libraries and dashed lines denote upper and lower quartiles. **(D-F)** Same as (A-C), but for N_{10}. **(G-I)** Same as (A-C), but for N_{11}. **(J-L)** Same as (A-C), but for N_{12}.

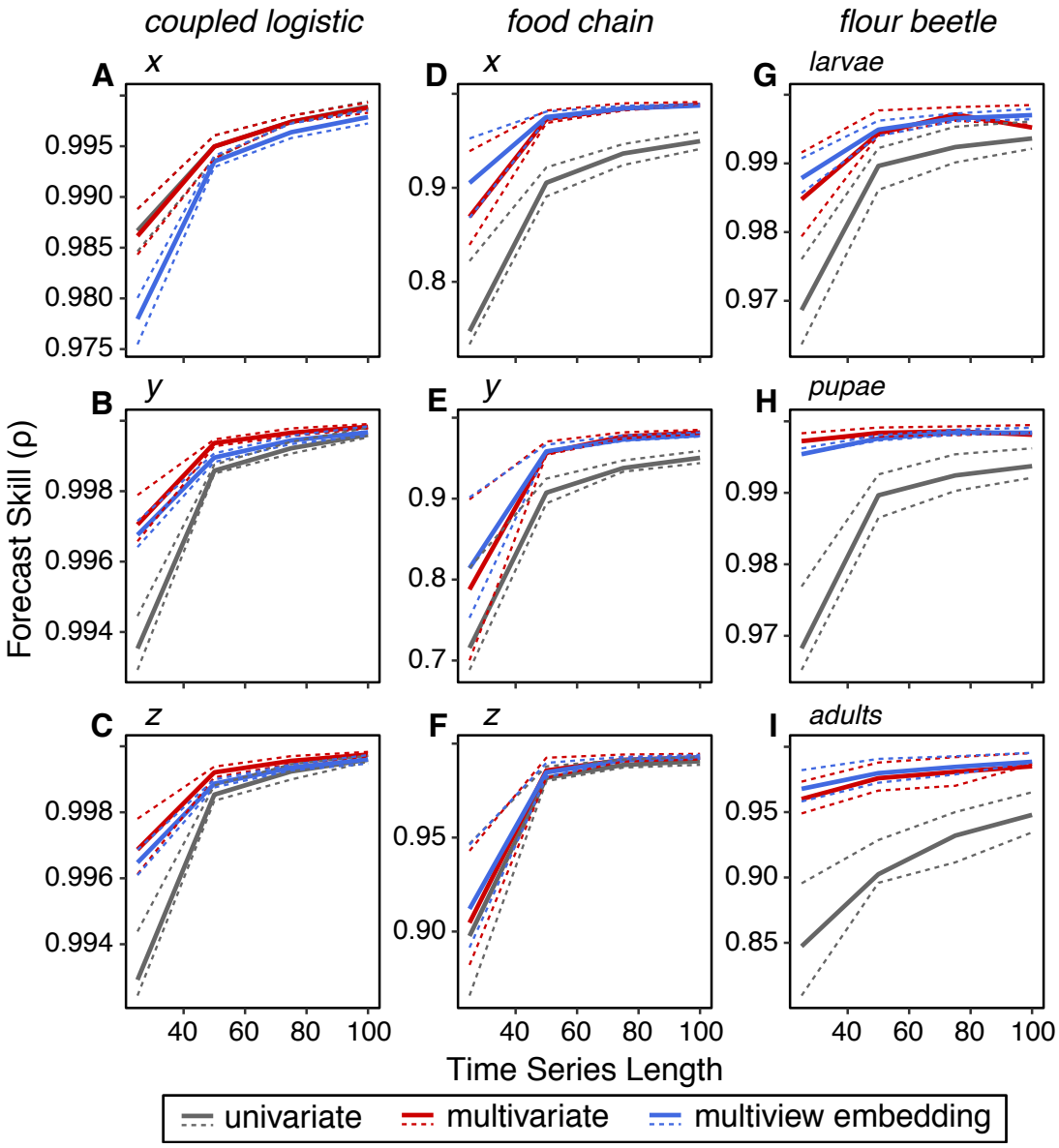

Fig. S13. Comparison of forecast skill for univariate, multivariate, and MVE methods on data from the simple model systems (*20*, *25*) with no observational error. In the absence of observation error, the multivariate method and MVE perform similarly and both generally perform better than the univariate method. **(A-C)** Forecast skill (ϱ, correlation between observations and predictions) vs. library size for variables *x*, *y*, and *z* in the 3-species coupled logistic. Solid lines indicate average values over 100 randomly sampled libraries and dashed lines denote upper and lower quartiles. **(D-F)** Same as **(A-C)**, but for the 3-species food chain model (*20*). **(G-I)** Same as **(A-C)**, but for the flour beetle model (*25*).

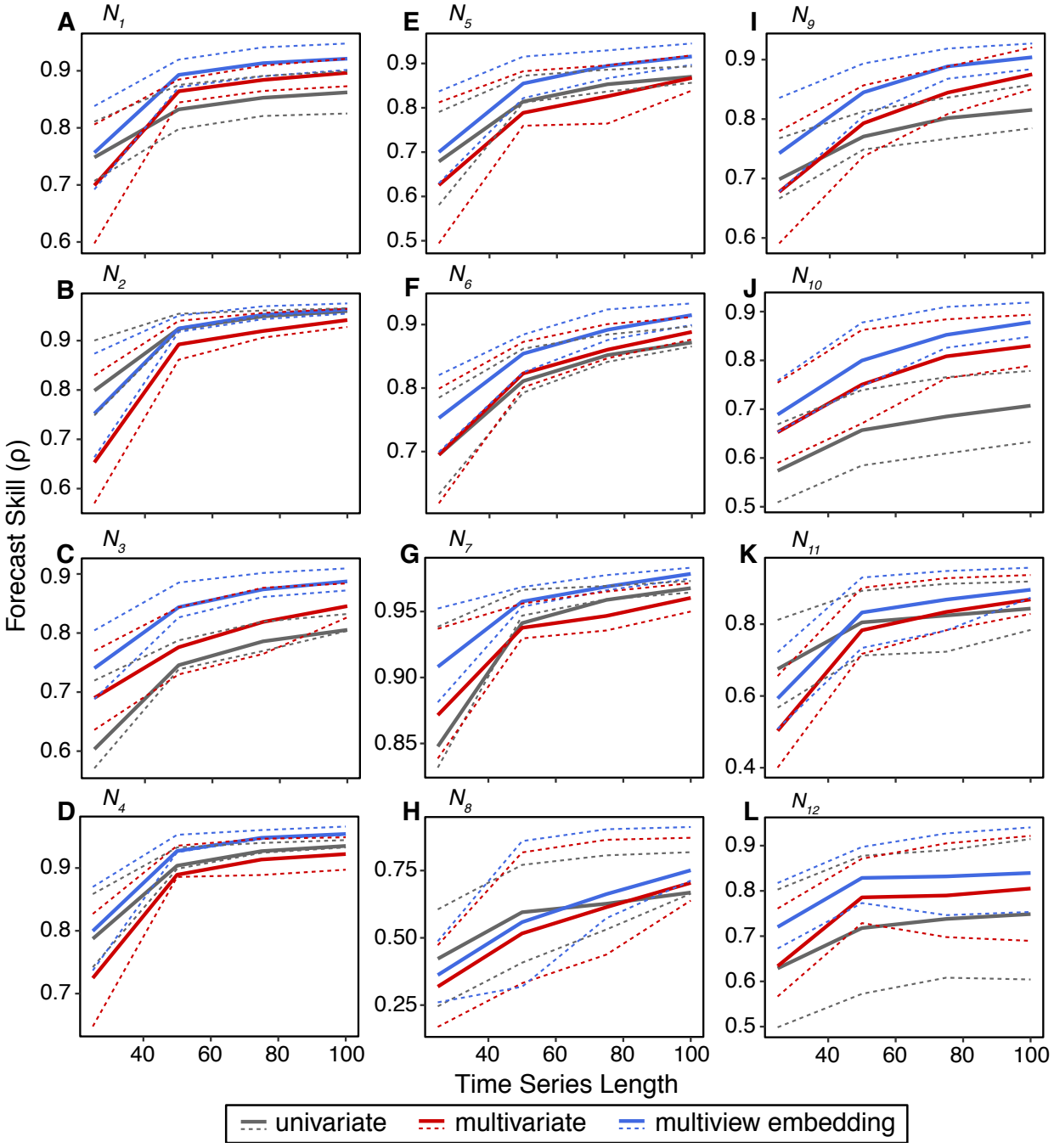

Fig. S14. Comparison of forecast skill for univariate, multivariate, and MVE methods on data from the 12-species resource competition model (*26*) with no observational error. In the absence of observation error, the performance of all methods primarily depends on time series length, although the MVE method generally outperforms the multivariate and univariate methods. **(A-L)** Forecast skill (ϱ, correlation between observations and predictions) vs. library size for variables $N_1 - N_{12}$.

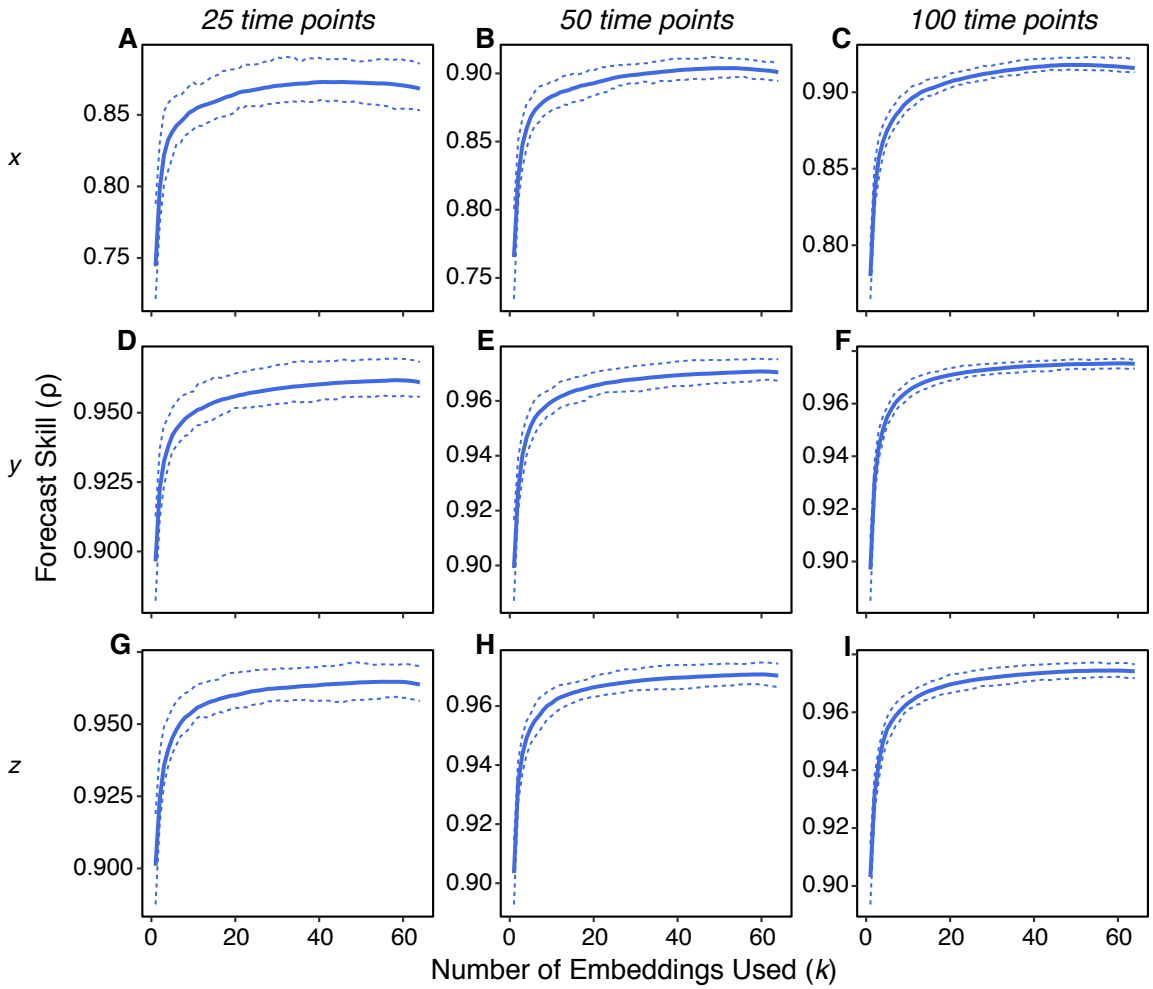

Fig. S15. MVE forecast skill vs. number of embeddings for the 3-species coupled logistic with 10% observational error. Multiview embedding forecast skill varies based on the number of embeddings used for MVE. **(A-C)** Forecast accuracy (ϱ, correlation between observations and predictions) vs. number of embeddings when predicting variable *x* and with different library sizes. Solid lines indicate average values over 100 randomly sampled libraries and dashed lines denote upper and lower quartiles. **(D-F)** Same as (A-C), but for variable *y*. **(G-I)** Same as (A-C), but for variable *z*.

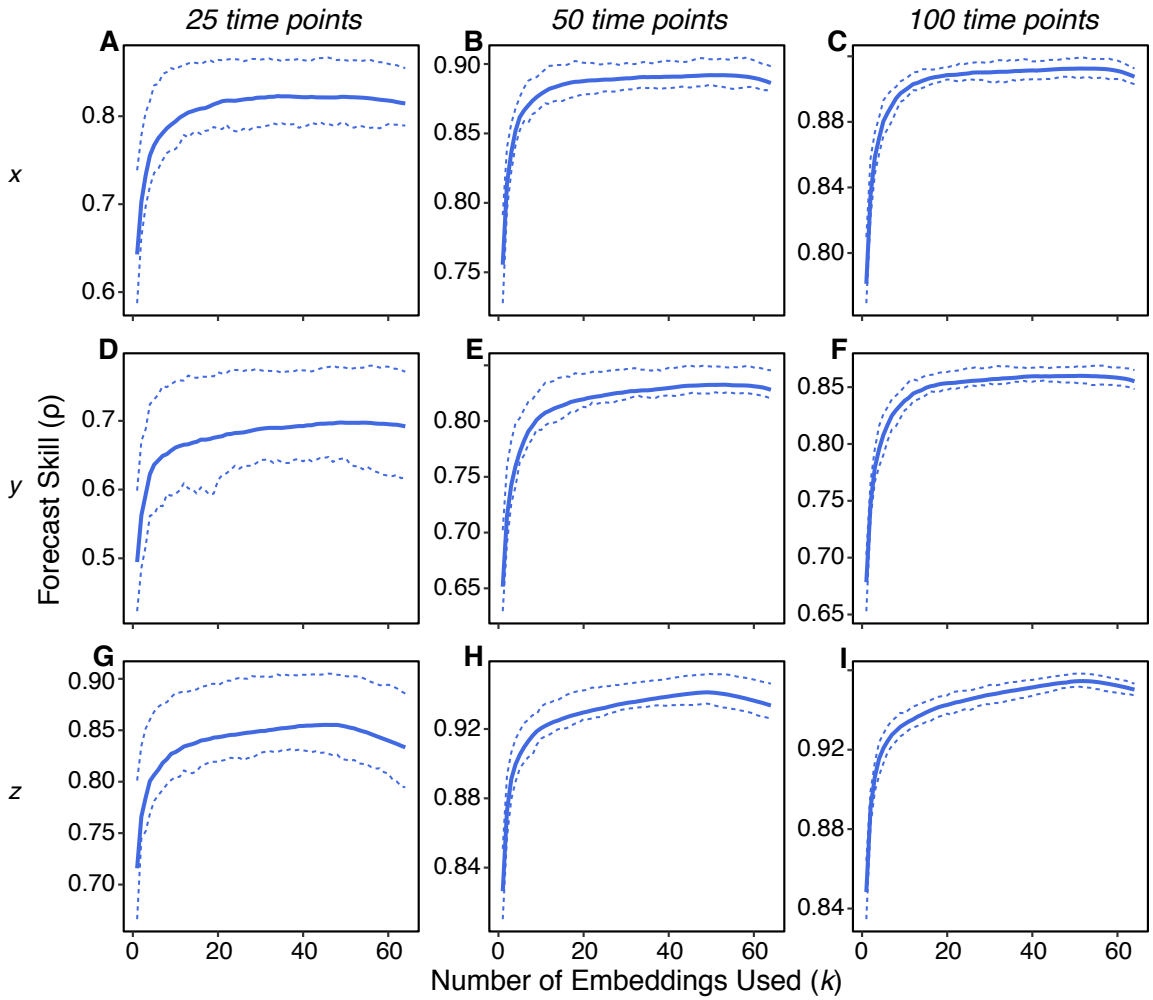

Fig. S16. MVE forecast skill vs. number of embeddings for the 3-species food chain model (*20*) with 10% observational error. Multiview embedding forecast skill varies based on the number of embeddings used for MVE. **(A-C)** Forecast accuracy (ϱ, correlation between observations and predictions) vs. number of embeddings when predicting variable *x* and with different library sizes. Solid lines indicate average values over 100 randomly sampled libraries and dashed lines denote upper and lower quartiles. **(D-F)** Same as (A-C), but for variable *y*. **(G-I)** Same as (A-C), but for variable *z*.

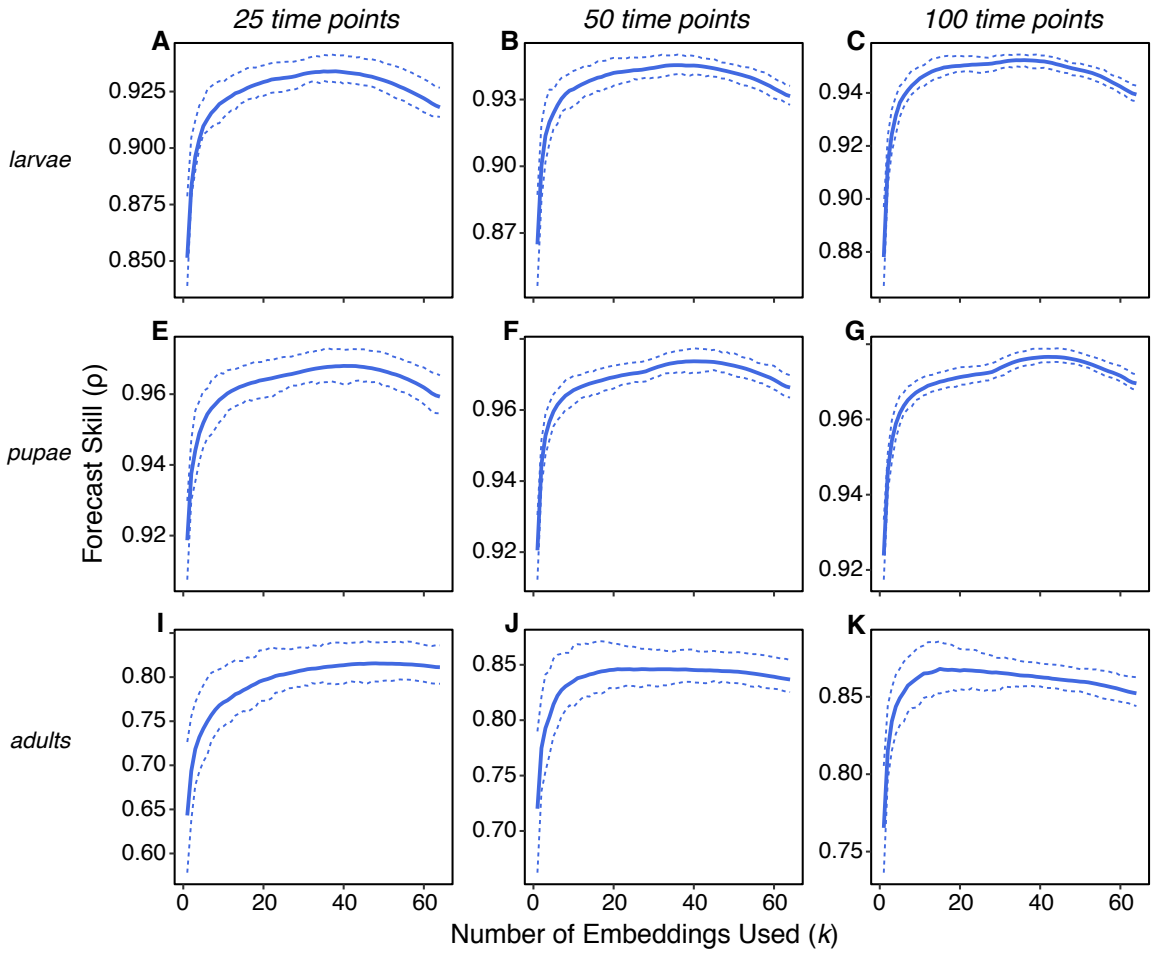

Fig. S17. MVE forecast skill vs. number of embeddings for the flour beetle model (*25*) with 10% observational error. Multiview embedding forecast skill varies based on the number of embeddings used for MVE. **(A-C)** Forecast accuracy (ϱ, correlation between observations and predictions) vs. number of embeddings when predicting larvae and with different library sizes. Solid lines indicate average values over 100 randomly sampled libraries and dashed lines denote upper and lower quartiles. **(D-F)** Same as (A-C), but for pupae. **(G-I)** Same as (A-C), but for adults.

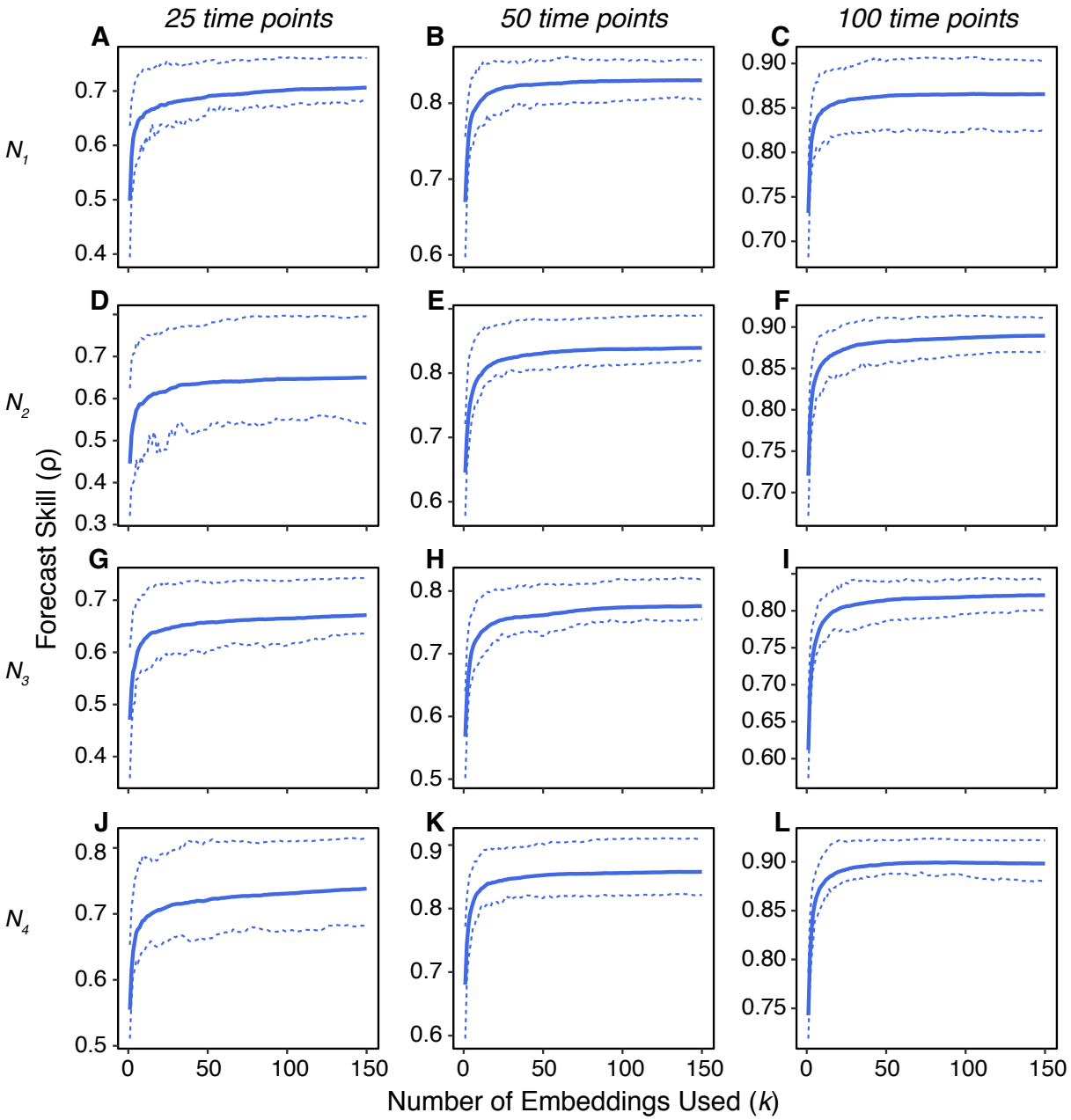

Fig. S18. MVE forecast skill vs. number of embeddings for the 12-species resource competition model (*26*) with 30% observational error. Multiview embedding forecast skill varies based on the number of embeddings used for MVE. **(A-C)** Forecast accuracy (ϱ, correlation between observations and predictions) vs. number of embeddings when predicting N_1 and with different library sizes. Solid lines indicate average values over 100 randomly sampled libraries and dashed lines denote upper and lower quartiles. **(D-F)** Same as (A-C), but for N_2. **(G-I)** Same as (A-C), but for N_3. **(J-L)** Same as (A-C), but for N_4.

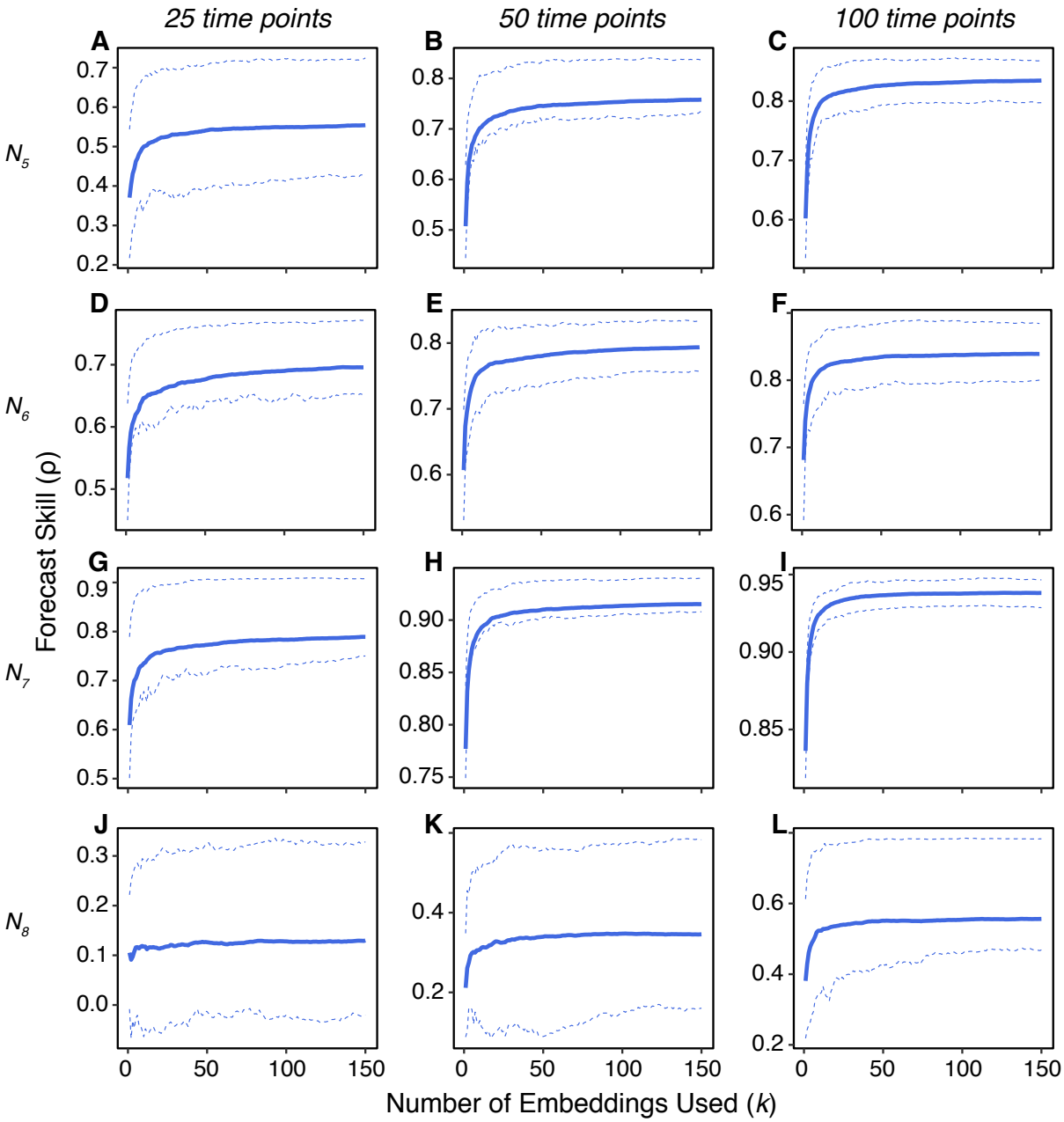

Fig. S19. MVE forecast skill vs. number of embeddings for the 12-species resource competition model (*26*) with 30% observational error. Multiview embedding forecast skill varies based on the number of embeddings used for MVE. **(A-C)** Forecast accuracy (ϱ, correlation between observations and predictions) vs. number of embeddings when predicting N_5 and with different library sizes. Solid lines indicate average values over 100 randomly sampled libraries and dashed lines denote upper and lower quartiles. **(D-F)** Same as (A-C), but for N_6. **(G-I)** Same as (A-C), but for N_7. **(J-L)** Same as (A-C), but for N_8.

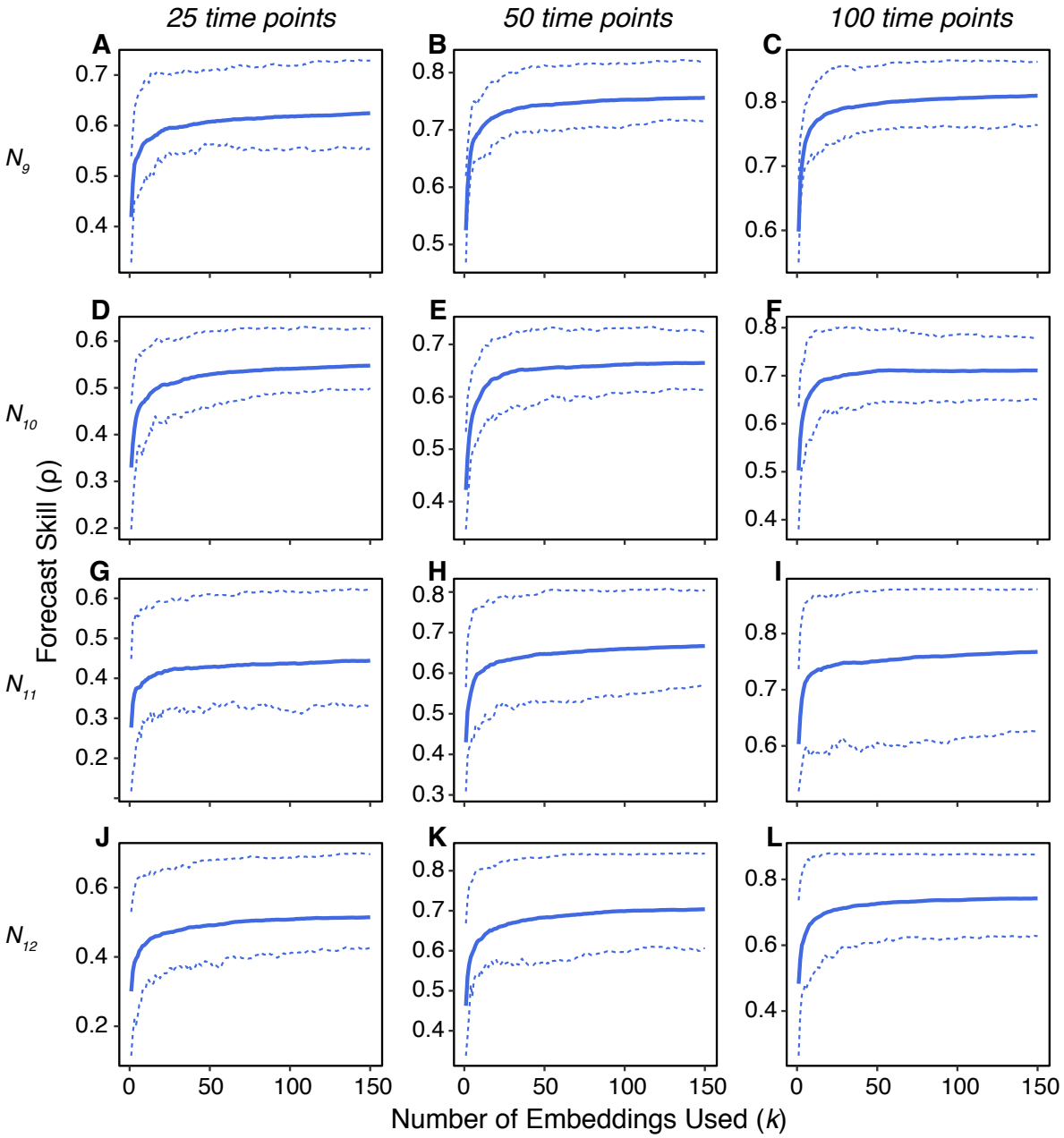

Fig. S20. MVE forecast skill vs. number of embeddings for the 12-species resource competition model (*26*) with 30% observational error. Multiview embedding forecast skill varies based on the number of embeddings used for MVE. **(A-C)** Forecast accuracy (ϱ, correlation between observations and predictions) vs. number of embeddings when predicting N_9 and with different library sizes. Solid lines indicate average values over 100 randomly sampled libraries and dashed lines denote upper and lower quartiles. **(D-F)** Same as (A-C), but for N_{10}. **(G-I)** Same as (A-C), but for N_{11}. **(J-L)** Same as (A-C), but for N_{12}.

External Database S1. This Excel file contains the model simulation data. Sheet "Coupled Logistic" contains data for the coupled logistic model; sheet "3-species Food Chain" contains data for the 3-species food chain model (*20*); sheet "Flour Beetle Model" contains data for the flour beetle model (*25*); sheet "12-species Resource Competition" contains data for the 12-species resource competition model (*26*).

References and Notes

1. R. M. May, S. A. Levin, G. Sugihara, Complex systems: Ecology for bankers. *Nature* **451**, 893–895 (2008).doi:10.1038/451893a Medline

2. D. Boyd, K. Crawford, Critical questions for big data. *Inf. Commun. Soc.* **15**, 662–679 (2012). doi:10.1080/1369118X.2012.678878

3. J. C. McBride, X. Zhao, N. B. Munro, G. A. Jicha, F. A. Schmitt, R. J. Kryscio, C. D. Smith, Y. Jiang, Sugihara causality analysis of scalp EEG for detection of early Alzheimer's disease. *Neuroimage Clin.* **7**, 258–265 (2014).doi:10.1016/j.nicl.2014.12.005 Medline

4. M. G. M. Olde Rikkert *et al*., *Crit. Care Med.* **44**, 601–606 (2016).

5. G. Sugihara, R. May, H. Ye, C. H. Hsieh, E. Deyle, M. Fogarty, S. Munch, Detecting causality in complex ecosystems. *Science* **338**, 496–500 (2012).doi:10.1126/science.1227079 Medline

6. J. Fan, F. Han, H. Liu, Challenges of big data analysis. *Natl. Sci. Rev.* **1**, 293–314 (2014).doi:10.1093/nsr/nwt032 Medline

7. D. Lazer, R. Kennedy, G. King, A. Vespignani, The parable of Google Flu: Traps in big data analysis. *Science* **343**, 1203–1205 (2014).doi:10.1126/science.1248506 Medline

8. E. A. Fulton, thesis, University of Tasmania (2001).

9. D. L. Donoho, "High-dimensional data analysis: The curses and blessings of dimensionality," presented at the American Mathematical Society Math Challenges of the 21st Century conference, Los Angeles, CA, 7 to 12 August 2000.

10. D. L. DeAngelis, S. Yurek, Equation-free modeling unravels the behavior of complex ecological systems. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 3856–3857 (2015).doi:10.1073/pnas.1503154112 Medline

11. C.-H. Hsieh, C. Anderson, G. Sugihara, Extending nonlinear analysis to short ecological time series. *Am. Nat.* **171**, 71–80 (2008).doi:10.1086/524202 Medline

12. T. Clark, H. Ye, F. Isbell, E. R. Deyle, J. Cowles, G. D. Tilman, G. Sugihara, Spatial convergent cross mapping to detect causal relationships from short time series. *Ecology* **96**, 1174–1181 (2015).doi:10.1890/14-1479.1 Medline

13. G. Sugihara, R. M. May, Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature* **344**, 734–741 (1990).doi:10.1038/344734a0 Medline

14. G. Sugihara, Nonlinear forecasting for the classification of natural time series. *Philos. Trans. Phys. Sci. Eng.* **348**, 477–495 (1994). doi:10.1098/rsta.1994.0106

15. P. A. Dixon, M. J. Milicich, G. Sugihara, Episodic fluctuations in larval supply. *Science* **283**, 1528–1530 (1999).doi:10.1126/science.283.5407.1528 Medline

16. F. Takens, *Dyn. Syst. Turbul. Lect. Notes Math.* **898**, 366–381 (1981).

17. T. Sauer, J. A. Yorke, M. Casdagli, Embedology. *J. Stat. Phys.* **65**, 579–616 (1991). doi:10.1007/BF01053745

18. E. R. Deyle, G. Sugihara, Generalized theorems for nonlinear state space reconstruction. *PLOS ONE* **6**, e18295 (2011).doi:10.1371/journal.pone.0018295 Medline

19. M. Casdagli, S. Eubank, J. D. Farmer, J. Gibson, State space reconstruction in the presence of noise. *Physica D* **51**, 52–98 (1991). doi:10.1016/0167-2789(91)90222-U

20. A. Hastings, T. Powell, Chaos in a three-species food chain. *Ecology* **72**, 896–903 (1991). doi:10.2307/1940591

21. R. O. Duda, P. E. Hart, D. G. Stork, *Pattern Classification* (Wiley, 2012).

22. For nearest-neighbor methods, asymptotic convergence requires only that the selection of *k* satisfies Stone's consistency theorem (*23*). The square root was chosen for simplicity and for its use in the machine learning literature.

23. C. J. Stone, Consistent nonparametric regression. *Ann. Stat.* **5**, 595–620 (1977). doi:10.1214/aos/1176343886

24. See supplementary materials on *Science* Online.

25. B. Dennis, R. A. Desharnais, J. M. Cushing, S. M. Henson, R. F. Costantino, Estimating chaos and complex dynamics in an insect population. *Ecol. Monogr.* **71**, 277–303 (2001). doi:10.1890/0012-9615(2001)071[0277:ECACDI]2.0.CO;2

26. J. Huisman, F. J. Weissing, Biodiversity of plankton by species oscillations and chaos. *Nature* **402**, 407–410 (1999). doi:10.1038/46540

27. R. Heerkloss, G. Klinkenberg, *Verhandlungen - Int. Vereinigung für Theor. und Angew. Limnol.* **26**, 1952–1956 (1998).

28. E. Benincà, K. D. Jöhnk, R. Heerkloss, J. Huisman, Coupled predator-prey oscillations in a chaotic food web. *Ecol. Lett.* **12**, 1367–1378 (2009).doi:10.1111/j.1461-0248.2009.01391.x Medline

29. T. Sauer, A noise reduction method for signals from nonlinear systems. *Physica D* **58**, 193–201 (1992). doi:10.1016/0167-2789(92)90108-Y

30. R. E. Kalman, A new approach to linear filtering and prediction problems. *J. Basic Eng.* **82**, 35–45 (1960). doi:10.1115/1.3662552

31. T. L. Carroll, F. J. Rachford, Using filtering effects to identify objects. *Chaos* **22**, 023107 (2012).doi:10.1063/1.3702566 Medline

32. E. Ott, C. Grebogi, J. A. Yorke, Controlling chaos. *Phys. Rev. Lett.* **64**, 1196–1199 (1990).doi:10.1103/PhysRevLett.64.1196 Medline