

Supporting Information
for
Predicting Time Series by Data-Driven Spatiotemporal Information Transformation

Peng Tao^{a,b,#}, Xiaohu Hao^{b,c,#}, Jie Cheng^d and Luonan Chen^{a,b,e,*}

^a Key Laboratory of Systems Health Science of Zhejiang Province, School of Life Science, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou 310024 China

^b Key Laboratory of Systems Biology, Shanghai Institute of Biochemistry and Cell Biology, Center for Excellence in Molecular Cell Science, Chinese Academy of Sciences, Shanghai 200031 China

^c Bioinformatics Core of Excellence Department, GenScript Biotech Corporation, Nanjing 211110, China

^dHUAWEI Technologies Co., Ltd, Shenzhen 518129, China

^eGuangdong Institute of Intelligence Science and Technology, Hengqin, Zhuhai, Guangdong 519031, China

* To whom correspondence may be addressed: Luonan Chen: lnchen@sibs.ac.cn

These authors contributed equally to this work

Contents included in this file:

- *Introduction of the synthetic and real-world datasets*
- *Pseudo code of MT-GPRM*
- *Supplementary Tables: Table S1 – Table S3*
- *Supplementary Figures: Fig. S1 – Fig. S17*

Note: each section starts on a new page.

Introduction of the synthetic and real-world datasets

Coupled Time-Variant Lorentz System

To validate the ability of our method in capturing high-dimensional nonlinear system's dynamics, a 90D time-variant coupled Lorentz system is under consideration. The i -th ($i=1,2,\dots,30$) coupled sub-system is given by

$$\begin{aligned}\dot{x}_i &= \sigma(t)(y_i - x_i) + Cx_{i-1}, \\ \dot{y}_i &= \rho x_i - y_i - x_i z_i, \\ \dot{z}_i &= -\beta z_i + x_i y_i,\end{aligned}$$

where Cx_{i-1} is the coupling term which means the i -th subsystem is coupled with the $(i-1)$ th sub-system via x component. When $i = 1$, we set $i - 1$ as 30 so that the system could be closed. We set ρ , β and C to be typical values, i.e., $\rho = 28$, $\beta = 8/3$ and $C = 0.1$. $\sigma(t) = 10 + 0.01(t | 10)$ is a time-variant parameter function with its value being initially set to be 10 and increased by 0.01 after each 10 time intervals, as a result, the system is time-switching.

We set the initial values to be 0.0001 and the time interval Δt to be 0.02 for generating the dataset. 155 points of the system are chosen in our experiment, they are divided into 21 parts with a sliding window which is set to 5, in each part, 30 points for observation and 25 points for prediction are included.

Coupled Time-Variant Pendulum System

The nonlinear (undamped) pendulum is a classic textbook example for dynamical systems, which is also used for benchmarking deep models. The pendulum system can be expressed as a second-order ODE by

$$\frac{d^2\theta}{dt^2} + \frac{g}{l} \sin \theta = 0$$

where the angular displacement from an equilibrium is denoted by $\theta \in [0, 2\pi]$. Taking l and g to denote the length and gravity, respectively, with $l = 1$ and $g = 9.8$ in practice. With following initial conditions $\theta(0) = \theta_0$ and $\dot{\theta}(0) = 0$ under consideration. The motion of the pendulum is approximately harmonic for a small amplitude of the oscillation $\theta_0 \ll 1$. However, the problem becomes inherently nonlinear for large amplitudes of the oscillations.

With oscillation angles $\theta_0 = 0.8$ on the time interval $t = [0, 20]$. The data are generated using a time step $\Delta t = 0.1$, yielding $T = 200$ equally spaced points $\{x_1, \dots, x_T\} \in \mathbf{R}^2$. In addition, the time series $\{x_t\}$ was mapped to a high-dimensional space via a random orthogonal transformation to obtain the high-dimensional data, i.e., $\mathbf{P} \in \mathbf{R}^{64 \times 2}$, such that $f_t = \mathbf{P}x_t$ for any t . Finally, we use 50 of the time series in our experiment for observation, and make 30 steps prediction accordingly.

Plankton Dataset

This dataset is collected from an optical plankton counter (OPC) and CTD mounted to a ScanFish platform that was towed and undulated behind the R/V Pelican during cruises PE03-NGOMEX, PE04-NGOMEX, PE06-NGOMEX, PE07-NGOMEX, PE09-05, and PE11-06 in the Northern Gulf of Mexico between 2003 and 2010. CTD and MIDAS data were synchronized and merged with simultaneously collected OPC data and aggregated into 1 second time bins. The bottom depth was obtained from the NOAA NCEI coastal relief model. This dataset contains 58 variables and we select the dissolved oxygen concentration to predict. We use 50 time points as observation and make 30 steps prediction accordingly.

Ground Ozone Level Dataset

The eight hour peak set ground ozone level dataset is collected from 1998 to 2004 at the Houston, Galveston and Brazoria area. The dataset contains 72 variables and there are missing values in it. We use the mean of neighbor values to fill these missing values, and select the average temperature to predict. We use 50 time points as observation and make 30 steps prediction accordingly.

Typhoon Eye Trace Dataset

The typhoon eye trace dataset contains satellite cloud images collected by National Institute of Informatics and we selected the typhoon Marcus to predict. The dataset is composed of a series of 241 cloud images (2402 variables) from March 15, 2018 to March 24, 2018 with one image taken per hour. 50 time points were used for observation, and we made the prediction on the next 27 time points.

Wireless Channel Dataset

The wireless channel dataset is originally 192*168 matrix for each time point, which means that the dimension of the dataset is 192*168 high. Considering the physical meaning of the data at each time point, the prediction of the target variable can be made only on the column the variable exists. And each element in the matrix is a complex number that can be computed separately, i.e., the real (amplitude) part and the imaginary (frequency/phase) part. Hence, in our prediction, the dimension of the variables is actually 192. We take 64 time points as observations, making a prediction on the next 24 time points based only on the 64 observations. To be noticed that the prediction results shown here are all manually adjusted according to the experience knowledge in wireless communication: there always exists time delay or advance. In the tested dataset, there exists 5 milliseconds time delay for about every 50 milliseconds (not fixed). So that the final prediction results here are all adjusted by forwarding the computed values by 5 milliseconds in offsetting the time delay.

Pseudo Code

Below we summarize the pseudo code of MT-GPRM.

Algorithm 1 MT-GPRM

```

1   Input
2      $\mathbf{X} = \{X(1), X(2), \dots, X(M)\}$ , size =  $M \times N$ 
3   Output
4     the  $L-1$  steps predicted values of the target variable
5   Initialization
6     Set the parameters:
7        $L, J, B_l, B_u, max\_run$ 
8     Normalization:
9       normalize the data into (0,1) use the min-max scaler
10  Training and Predicting
11    for  $l$  in range( $L$ ):
12       $run\_number = 0$ 
13      while True:
14        optimizing  $\Theta$  and  $\mathbf{K}^f$  according to Equation (8)
15         $run\_number += 1$ 
16        if  $run\_number \geq max\_run$ :
17          break
18        end if
19      end while
20      get the best hyper-parameters  $\Theta$  and  $\mathbf{K}^f$ 
21      get the prediction results for current mappings
22    end for
23    get the average value of the prediction results for each time points

```

The notations are listed as follows: The \mathbf{X} represents the input observed samples of the time series data, whose shape is $M \times N$. It means that there are M observed samples and each sample has N dimension. $L - 1$ is the steps that we are going to predict ahead, and L is the number of mappings constructed in the STI equation. J is the number of the tasks been considered together by GPRs. B_l and B_u represent the initial lower and upper bound of the length scaler used for initializing the kernels in GPR, respectively. run_num and max_run represent the current running number and the maxima allowed running number of the optimization process for getting optimized Θ and \mathbf{K}^f , respectively.

Supplementary Tables

Table S1 Comparison of the prediction results on Lorentz system for all methods

Target	Methods	Noise Free			Noise $\sim N(0, 3)$		
		MAE	RMSE	PCC	MAE	RMSE	PCC
x_{16}	MT-GPRM	0.539	0.6457	0.9987	1.6089	1.7443	0.997
	Informer	10.3511	11.3238	-0.8825	9.4185	10.4936	-0.7859
	ARNN	264.5192	397.2053	-0.7805	15.4995	17.5231	-0.7331
	ALM	1.6677	2.2416	0.9847	2.1625	2.4613	0.9763
	RDE	3.2624	3.9615	0.9923	6.6629	7.2157	0.9384
	LSTM	8.5586	9.3808	-0.103	9.1389	10.250	0.6924
	ARIMA	8.2979	9.7754	-0.696	9.6260	11.190	-0.849
	VAR	21.353	34.576	-0.243	3.6495	4.1923	0.9471
	MA	8.1460	9.0421	-0.971	8.7035	9.4493	-0.949
	SES	10.9401	12.7735	None	10.588	12.269	0.4113
	VARM	7.2013	9.1993	None	7.2013	9.1993	None
	SVR	52717	178105	0.3625	23.343	27.893	-0.961
x_{17}	SVE	7.6023	8.8027	-0.004	7.8758	8.5304	-0.026
	MVE	10.968	12.367	-0.399	10.753	13.553	0.1338
	LASSO	209.38	312.27	-0.772	8.5524	9.7252	0.9452
	ADALASSO	205.34	305.58	-0.773	8.1095	9.1386	0.9260
	MT-GPRM	0.24	0.3079	0.9996	1.5331	1.6618	0.9993
	Informer	10.5158	11.9691	-0.3267	11.5531	12.2954	0.18
	ARNN	601.6276	861.7727	-0.5976	14.9527	19.2562	-0.2158
	ALM	568.2228	2003.7105	0.2661	2.6152	3.1542	0.9784
	RDE	4.8618	5.9409	0.9567	12.651	13.654	0.9366
	LSTM	15.119	16.2577	0.2142	16.629	17.878	0.5285
	ARIMA	10.667	12.302	-0.551	9.4602	11.810	-0.182
x_{18}	VAR	68.789	93.985	-0.165	7.3367	8.2545	0.8171
	MA	12.595	13.270	-0.837	12.145	12.881	-0.788
	SES	22.804	24.689	0.0000	18.953	20.493	None
	VARM	7.748	9.5663	None	7.748	9.5663	None
	SVR	50.042	58.725	0.3283	25.003	27.110	-0.639
	SVE	10.497	11.022	0.3410	10.657	11.386	-0.280
	MVE	14.695	16.027	-0.163	13.528	15.863	0.0184
	LASSO	44.375	48.882	-0.834	39.108	44.129	-0.772
	ADALASSO	153.95	201.29	-0.638	38.827	43.783	-0.773
	MT-GPRM	0.93	1.0606	0.9959	1.5577	1.9505	0.988
	Informer	11.0923	13.139	0.6539	13.4492	15.01	0.672
	ARNN	753.7308	1515.8379	-0.6375	21.5721	24.1102	-0.2175
x_{19}	ALM	3.8133	4.0279	0.9879	4.7853	5.0049	0.9215
	RDE	6.7377	8.1531	0.9830	6.9652	8.9287	0.6822
	LSTM	18.074	19.967	-0.575	17.655	19.589	-0.445
	ARIMA	17.528	18.499	0.7607	17.315	18.162	0.7674
	VAR	43.015	63.232	0.1056	8.8260	10.079	0.8921
	MA	17.960	19.765	0.0379	18.736	20.435	0.2259
	SES	8.3594	10.998	0.0000	10.494	13.352	0.0672

VARM	29.1664	30.3124	None	29.1664	30.3124	None
SVR	176.31	228.86	- 0.890	19.626	21.932	- 0.120
SVE	17.436	19.337	0.0177	18.363	20.210	- 0.186
MVE	16.159	19.731	- 0.296	19.839	24.186	- 0.268
LASSO	1879.0	3336.0	- 0.700	1753.6	3092.8	- 0.704
ADALASSO	1753.6	3092.8	- 0.704	16.881	18.536	0.4583

Table S2 Prediction results of MT-GPRM($J=5$) and GPRM($J=1$) on Lorentz system

Target	Index	Performance with Different Level of Noise			
		Noise Free	Noise~ $N(0,1)$	Noise~ $N(0,2)$	Noise~ $N(0,3)$
$J=1$ GPRM	MAE	0.1711	0.1839	0.1804	0.1504
	RMSE	0.2141	0.2261	0.2246	0.1917
	PCC	0.9389	0.9447	0.9255	0.9409
	VAR	0.4399	0.3169	0.2585	0.3008
	MAE	0.1196	0.1182	0.1182	0.1235
	RMSE	0.1927	0.189	0.1906	0.2010
	PCC	0.9188	0.9244	0.9278	0.9231
	VAR	0.3297	0.3345	0.3443	0.3561
	MAE	0.3022	0.4207	0.4381	0.4235
	RMSE	0.3226	0.4513	0.4683	0.4565
	PCC	0.8729	0.7242	0.7084	0.6625
	VAR	0.2634	0.1768	0.1778	0.1772
$J=5$ MT-GPRM	MAE	0.0247	0.0673	0.1749	0.1619
	RMSE	0.03	0.0722	0.1955	0.1833
	PCC	0.9983	0.999	0.9692	0.9712
	VAR	0.2992	0.2722	0.3066	0.2245
	MAE	0.007	0.0143	0.0194	0.0281
	RMSE	0.0089	0.019	0.0231	0.0350
	PCC	0.9996	0.9994	0.9974	0.9961
	VAR	0.2104	0.203	0.2111	0.2054
	MAE	0.0287	0.0617	0.1113	0.1724
	RMSE	0.0323	0.0711	0.1239	0.1830
	PCC	0.9931	0.9861	0.9733	0.9667
	VAR	0.4548	0.4743	0.4191	0.3549

Table S3 Comparison of the prediction results on real-world datasets for all methods

Methods	Typhoon – longitude			Typhoon – Latitude			Oxygen Content			Average Temperature		
	MAE	RMSE	PCC	MAE	RMSE	PCC	MAE	RMSE	PCC	MAE	RMSE	PCC
MT-GPRM	0.1297	0.1612	0.9982	0.0928	0.108	0.9909	0.1209	0.1563	0.8564	0.9950	1.2291	0.9401
GPRM	0.3119	0.3443	0.9968	0.2608	0.2932	0.9799	0.2354	0.2698	0.7384	2.2495	2.9229	0.7953
Informer	1.5585	1.8363	-0.2355	1.2517	1.3521	0.7003	0.8427	0.9167	-0.3205	5.4292	6.3606	0.3508
ARNN	1.1009	1.3429	0.9906	1.2039	1.3617	-0.2381	0.5502	0.7141	-0.1526	9.7963	10.9444	-0.5144
ALM	0.6839	1.1311	0.8023	0.1497	0.2084	0.9898	0.2937	0.3972	0.5022	4.0483	5.5051	0.2833
RDE	6.7518	6.8818	0.4251	0.8495	1.0179	0.4350	2.7831	3.3177	0.2618	7.3420	8.4775	0.4202
LSTM	7.1229	8.1907	0.7565	0.7870	1.0111	-0.315	1.1305	1.1874	-0.085	3.9242	4.9044	0.4251
ARIMA	11.527	11.535	0.9922	0.6593	0.8147	-0.848	1.8324	1.8589	-0.043	9.4527	10.997	0.0579
VAR	6.6266	7.1684	-0.500	0.6639	0.8432	0.5680	11240	30773	-0.406	67.938	104.54	0.2280
MA	7.6307	7.6699	0.9993	0.7111	0.9436	-0.885	1.4660	1.4950	0.2520	8.2600	9.3741	0.3622
SES	2.8148	3.1620	0.0000	0.8519	1.0431	0.0000	0.2357	0.3176	0.0000	10.190	11.237	0.0000
VARM	103.88	105.67	0.4363	12.399	12.639	-0.352	4.9725	6.4143	0.4084	8.5357	9.6458	0.2264
SVR	5.5426	5.7693	0.4325	0.9676	1.1362	0.0486	2.9091	3.3681	0.1221	9.7993	10.774	0.5705
SVE	9.0268	9.1274	0.3477	0.5665	0.7919	-0.254	1.5415	1.6244	-0.553	7.9239	9.2436	-0.178
MVE	128.39	135.72	0.1175	254.13	259.10	-0.114	0.6531	0.7738	0.0758	9.5075	10.738	-0.176
LASSO	0.8320	1.1023	0.9947	0.8474	1.0174	0.9042	2.9140	3.5985	0.3551	8.5279	9.6406	0.3277
ADALASSO	0.8219	1.0900	0.9947	1.1218	1.3182	-0.966	6.3040	8.3069	0.4317	8.5278	0.6406	0.3291

Supplementary Figures

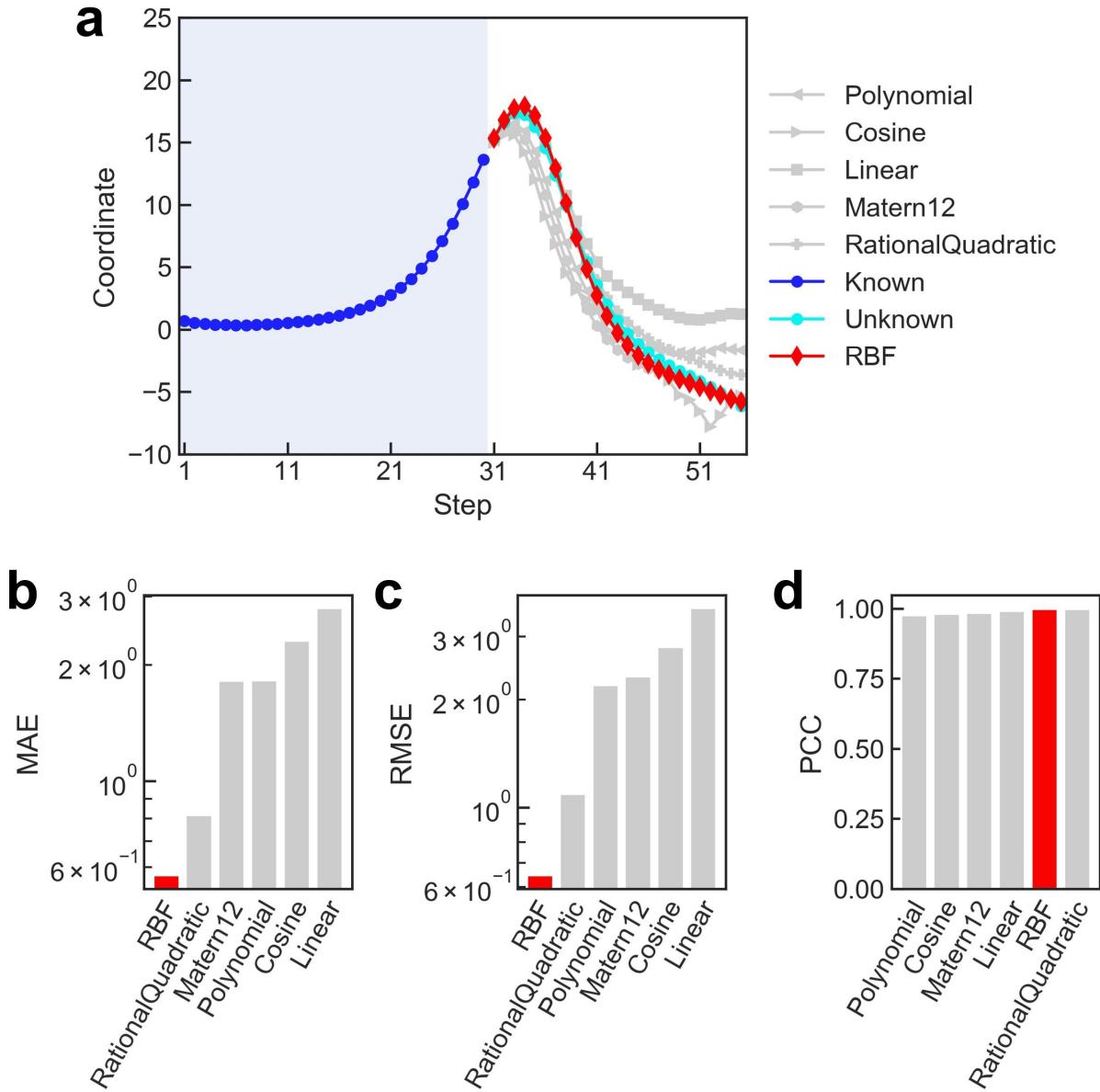


Fig. S1. Comparison of different kernel functions on a 90-dimension time-variant coupled Lorentz system (target variable is x_{16}). **(a)** is the predicted results for different kernel functions. The blue, cyan and red lines represent the known (training, 30 time points), unknown (test, 25 time points) and predicted values (MT-GPRM) of the target variable, respectively. For the convenience of visualization, the predicted values of the other 5 kernel function are displayed in gray. The corresponding prediction performances, including MAE, RMSE and PCC, are shown in **(b-d)**, respectively.

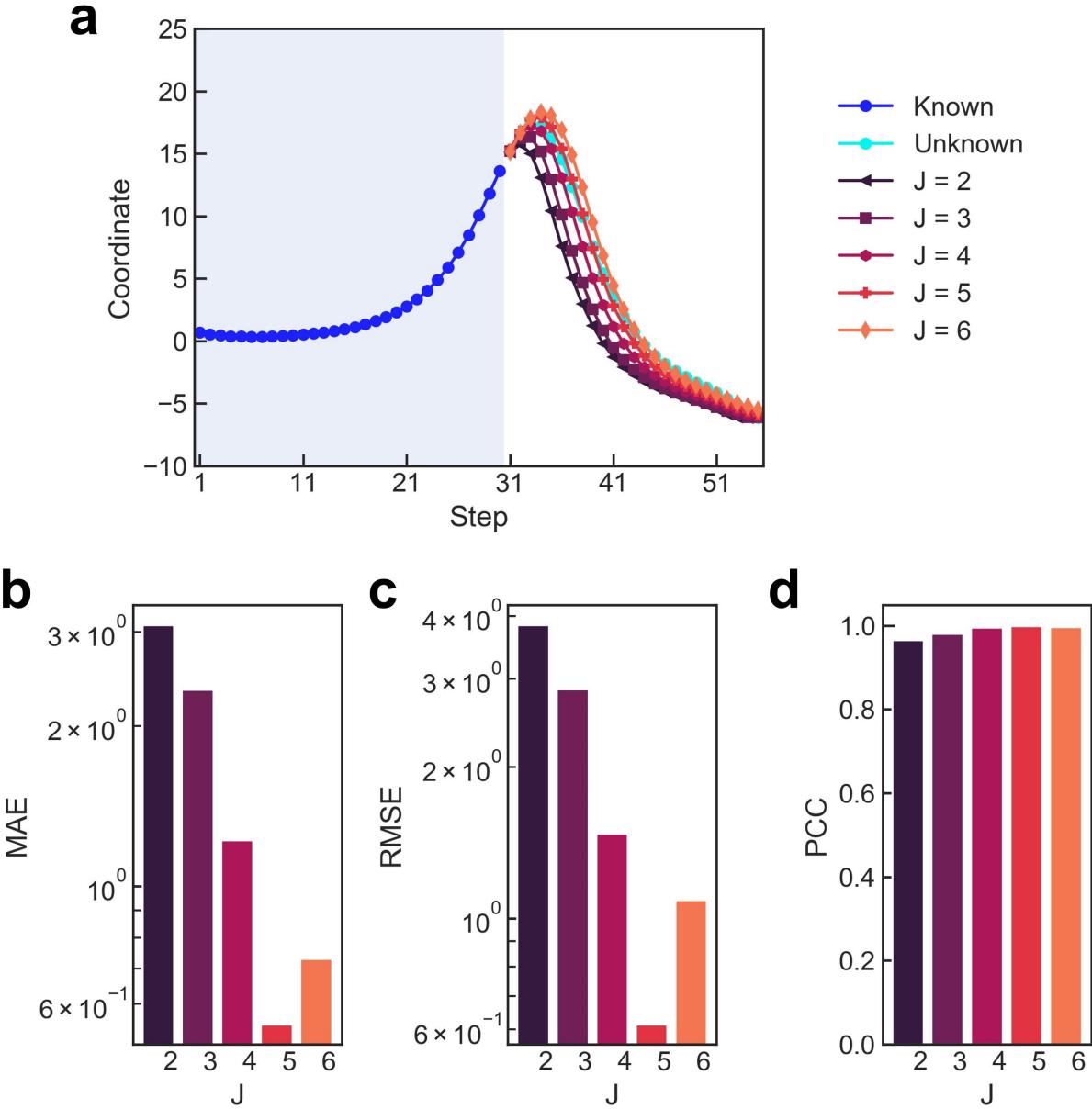


Fig. S2. Comparison of different task numbers on a 90-dimension time-variant coupled Lorentz system (target variable is x_{16}). (a) is the predicted results for different J . The blue, cyan and red lines represent the known (training, 30 time points), unknown (test, 25 time points) and predicted values (MT-GPRM) of the target variable, respectively. The corresponding prediction performances, including MAE, RMSE and PCC, are shown in (b-d), respectively.

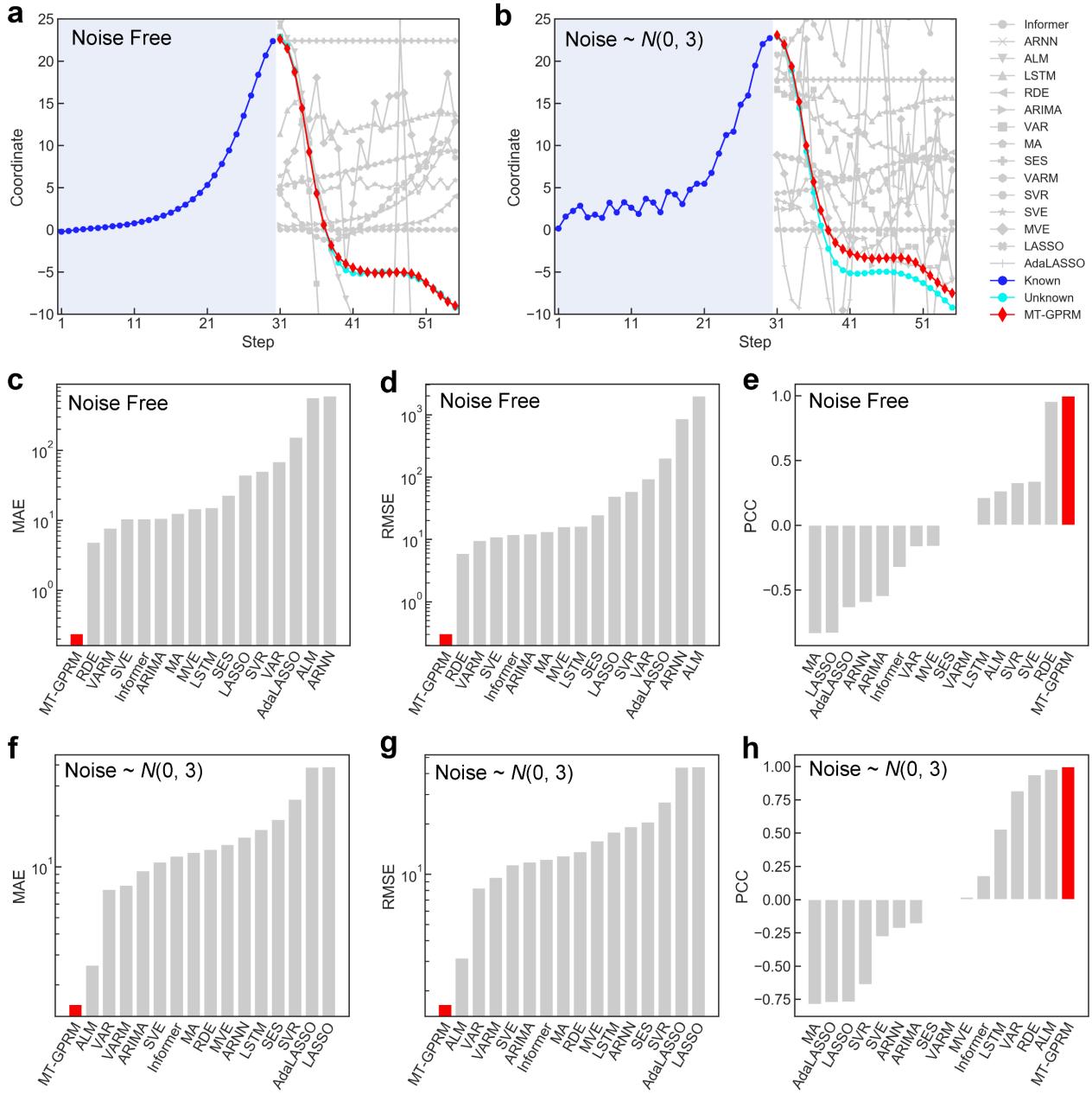


Fig. S3. Test results on a 90-dimension time-variant coupled Lorentz system (target variable is x_{17}). (a and b) are the predicted results for the noise-free and noise-polluted (Gaussian noise $\sim N(0, 3)$) cases, respectively. The blue, cyan and red lines represent the known (training, 30 time points), unknown (test, 25 time points) and predicted values (MT-GPRM) of the target variable, respectively. For the convenience of visualization, the predicted values of the other 15 methods are displayed in gray. The corresponding prediction performances, including MAE, RMSE and PCC, are shown in (c-e) for noise-free case, (f-h) for noise-polluted case, respectively.

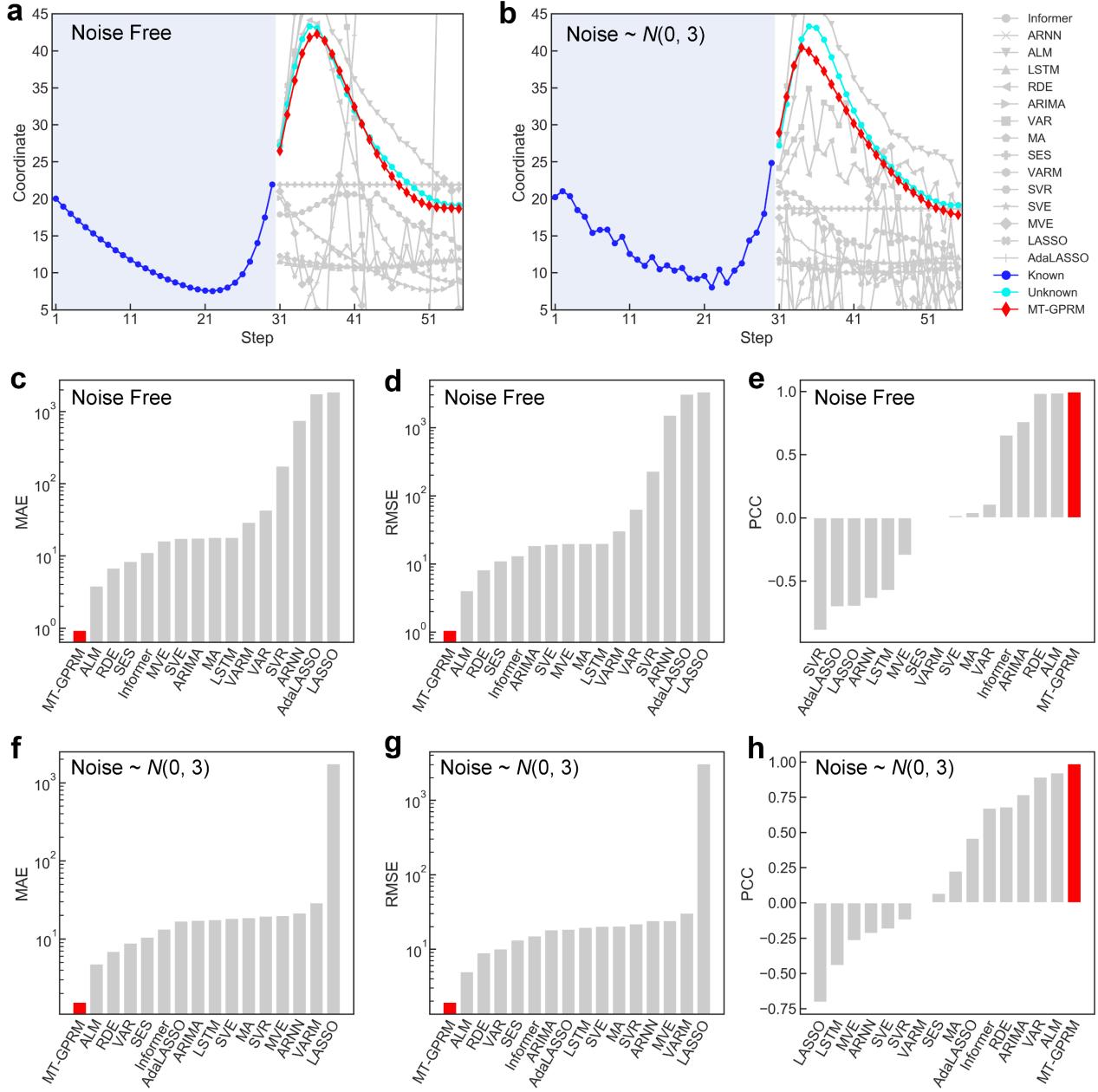


Fig. S4. Test results on a 90-dimension time-variant coupled Lorentz system (target variable is x_{18}). **(a and b)** are the predicted results for the noise-free and noise-polluted (Gaussian noise $\sim N(0, 3)$) cases, respectively. The blue, cyan and red lines represent the known (training, 30 time points), unknown (test, 25 time points) and predicted values (MT-GPRM) of the target variable, respectively. For the convenience of visualization, the predicted values of the other 15 methods are displayed in gray. The corresponding prediction performances, including MAE, RMSE and PCC, are shown in **(c-e)** for noise-free case, **(f-h)** for noise-polluted case, respectively.

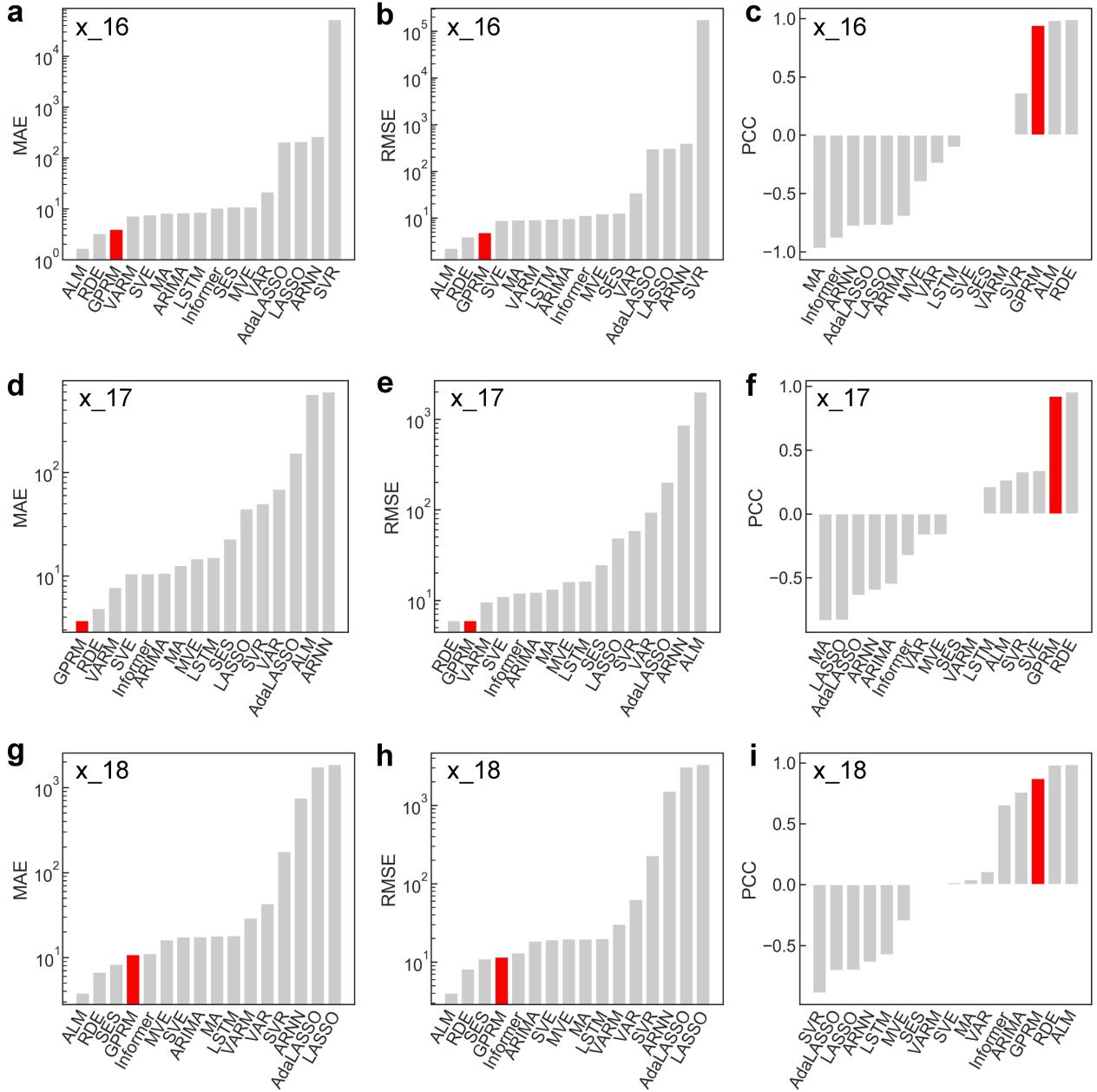


Fig. S5. Comparison of prediction performances between GPRM ($J=1$) and other 15 methods for noise-free case. The three rows from top to bottom are the prediction performances of x_{16} , x_{17} and x_{18} , and the three columns from left to right are the results of RMSE, MAE and PCC, respectively.

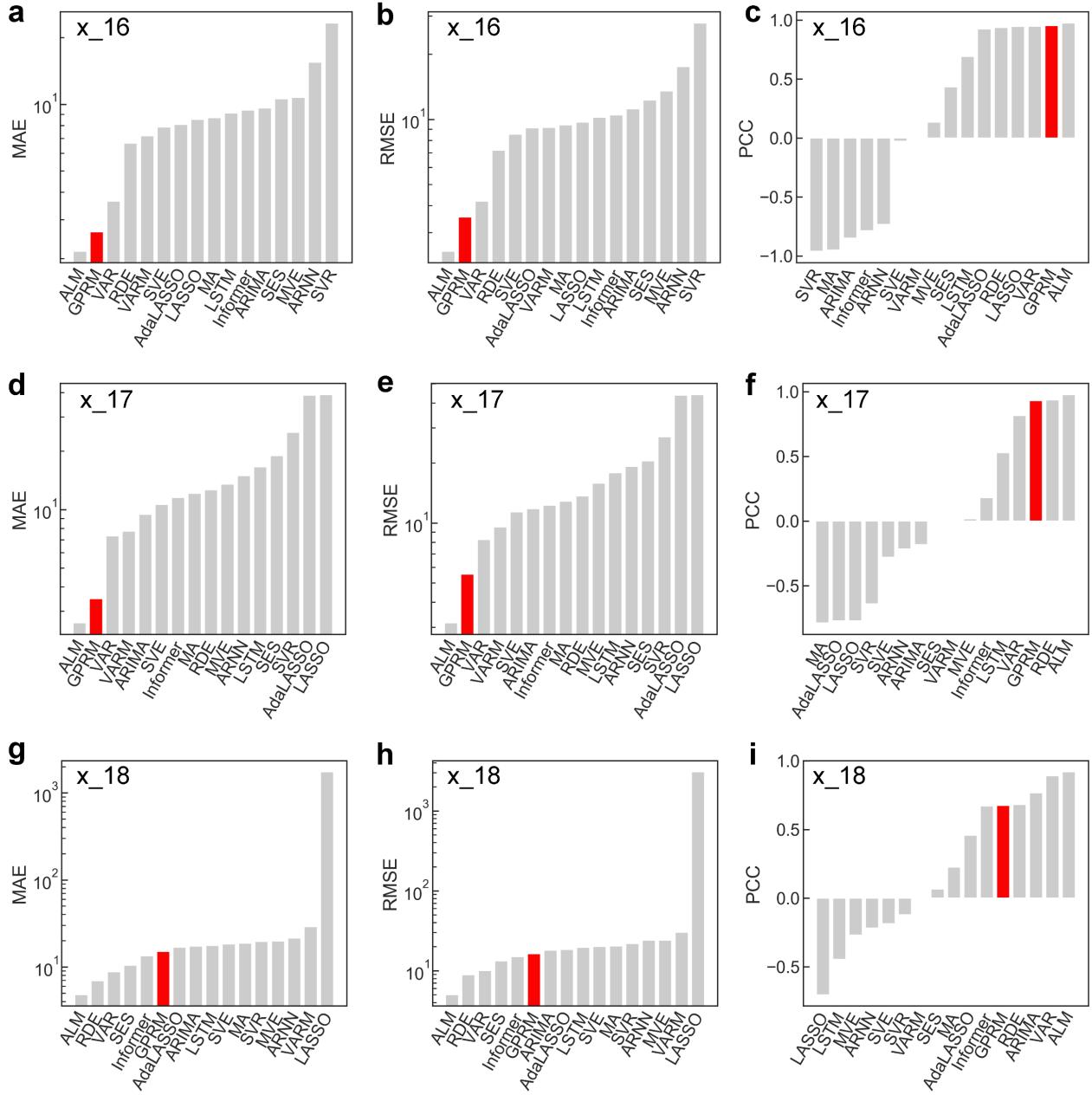


Fig. S6. Comparison of prediction performances between GPRM ($J=1$) and other 15 methods for noise polluted (Gaussian noise $\sim N(0, 3)$) case. The three rows from top to bottom are the prediction performances of x_{16} , x_{17} and x_{18} , and the three columns from left to right are the results of RMSE, MAE and PCC, respectively.

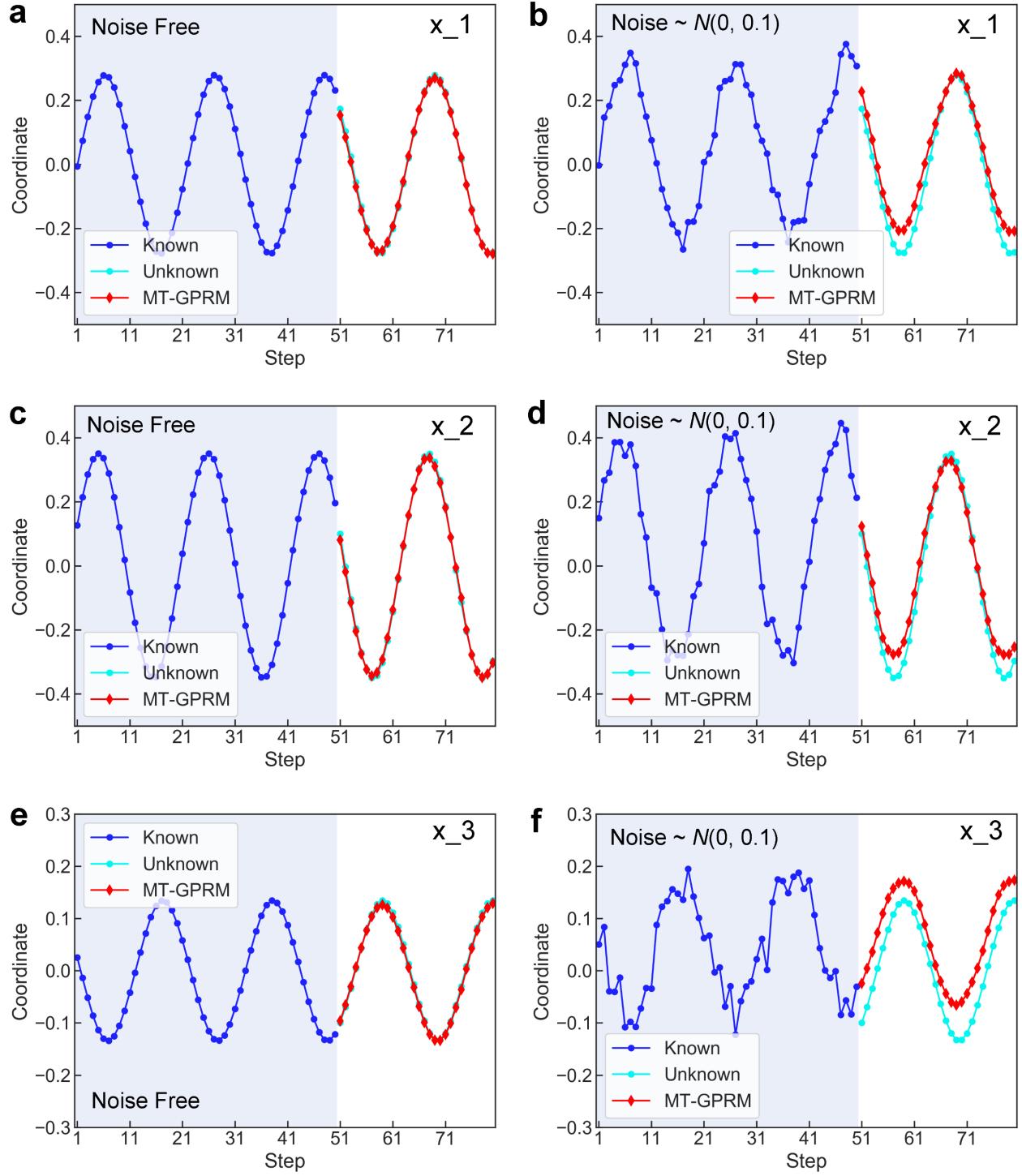


Fig. S7. Test results on a 64-dimension time-variant coupled pendulum system. The three rows from top to bottom are the prediction performances of x_1 , x_2 and x_3 , and the two columns from left to right are the results of the noise-free and noise polluted ($\text{Gaussian noise} \sim N(0, 0.1)$) cases, respectively. The blue, cyan and red lines represent the known (training, 50 time points), unknown (test, 30 time points) and predicted values (MT-GPRM) of the target variables, respectively.

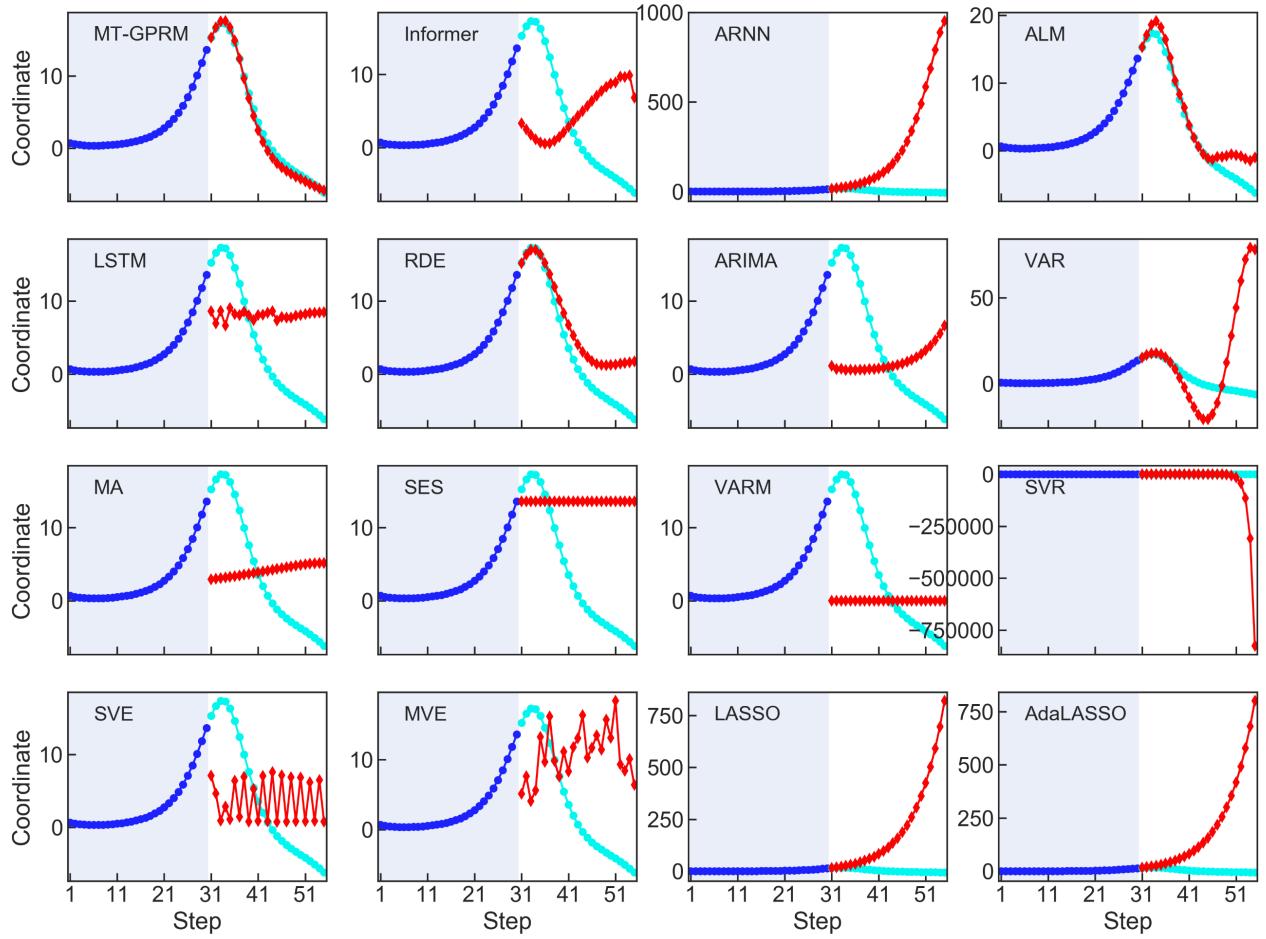


Fig. S8. Predicted results of all 16 methods on a 90-dimension time-variant coupled Lorentz system (target variable is $x_{_16}$).

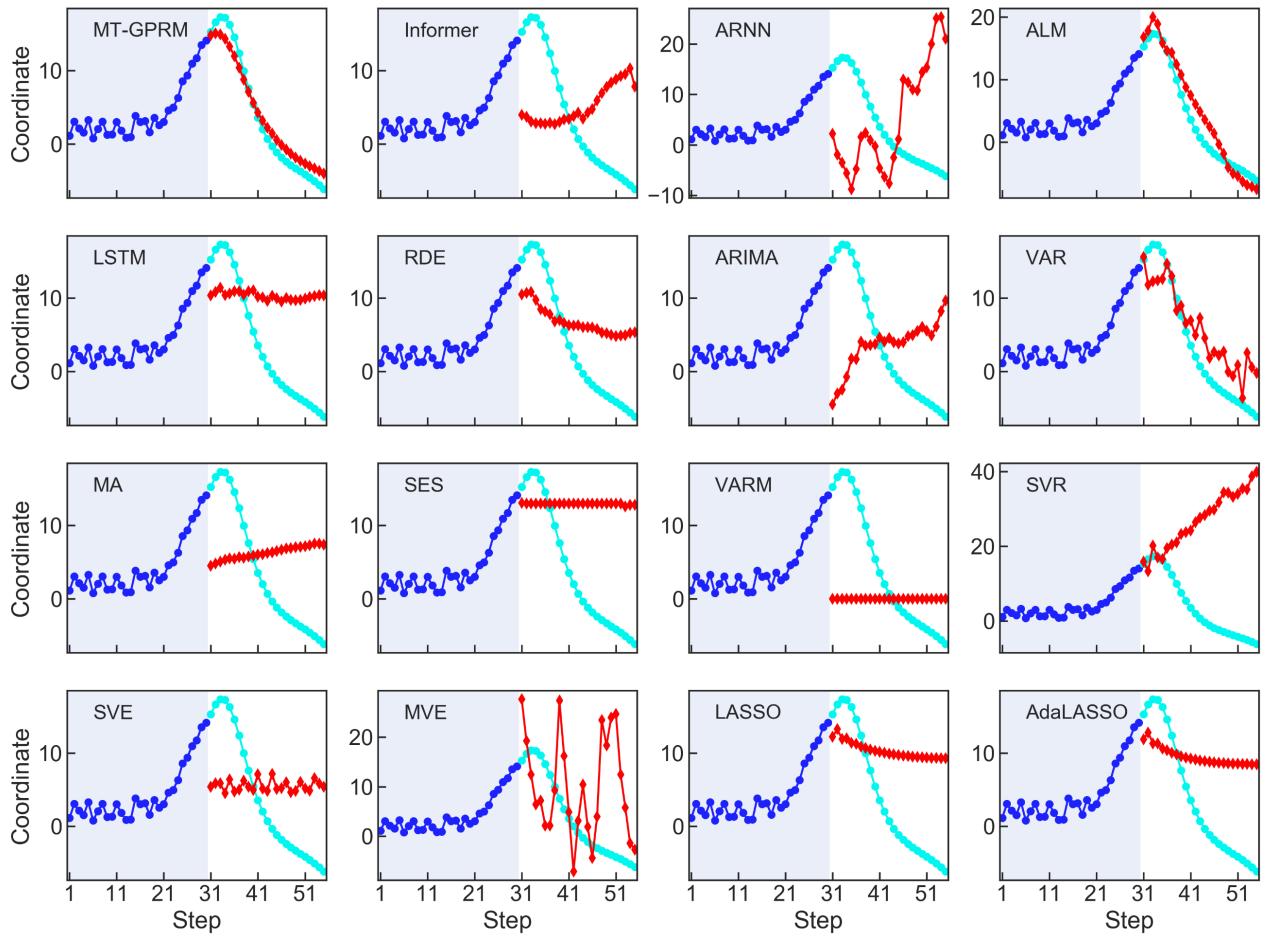


Fig. S9. Predicted results of all 16 methods on a 90-dimension time-variant coupled Lorentz system (target variable is $x_{_16}$ and Gaussian noise $\sim N(0, 3)$).

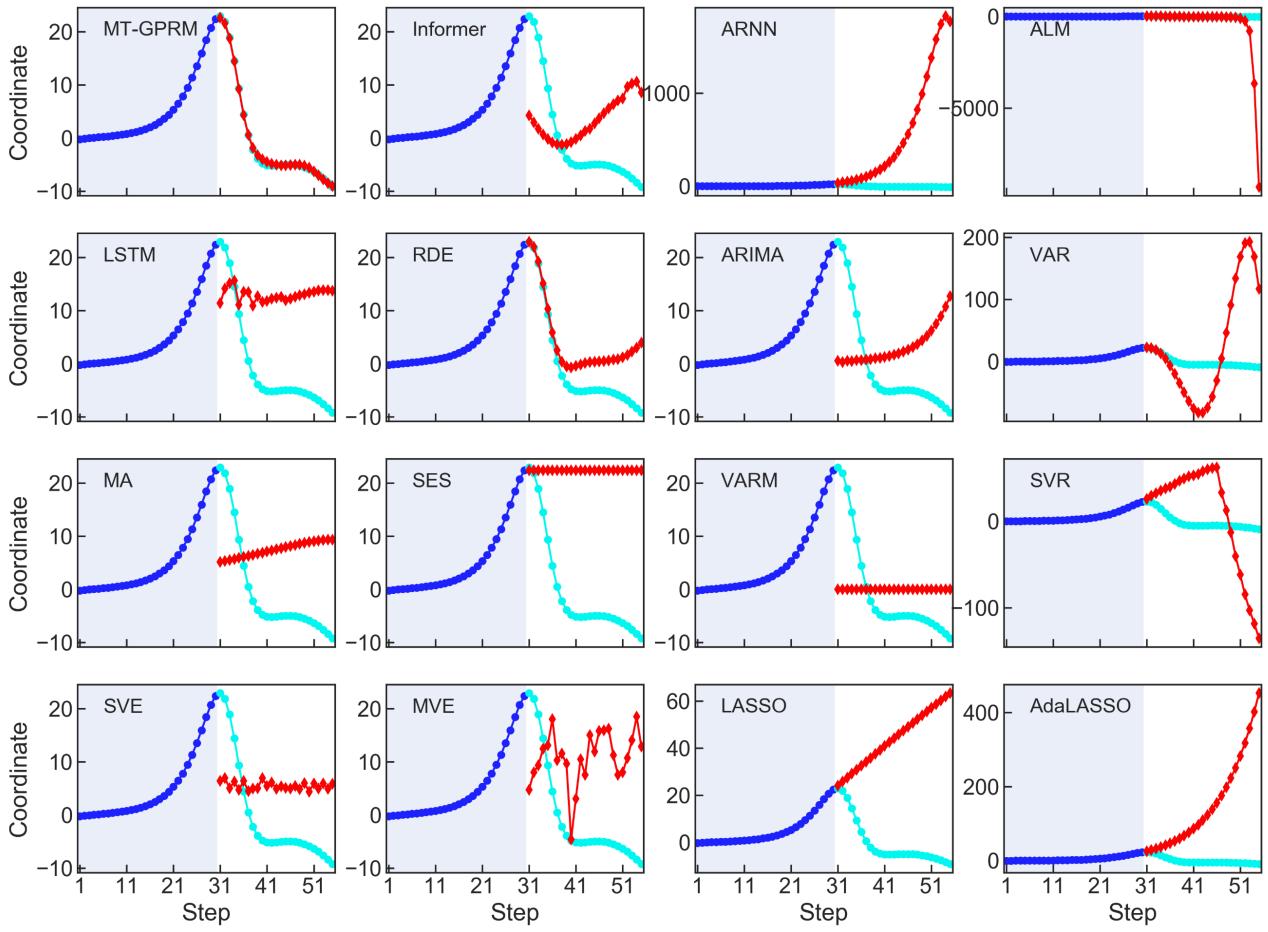


Fig. S10. Predicted results of all 16 methods on a 90-dimension time-variant coupled Lorentz system (target variable is x_{-17}).

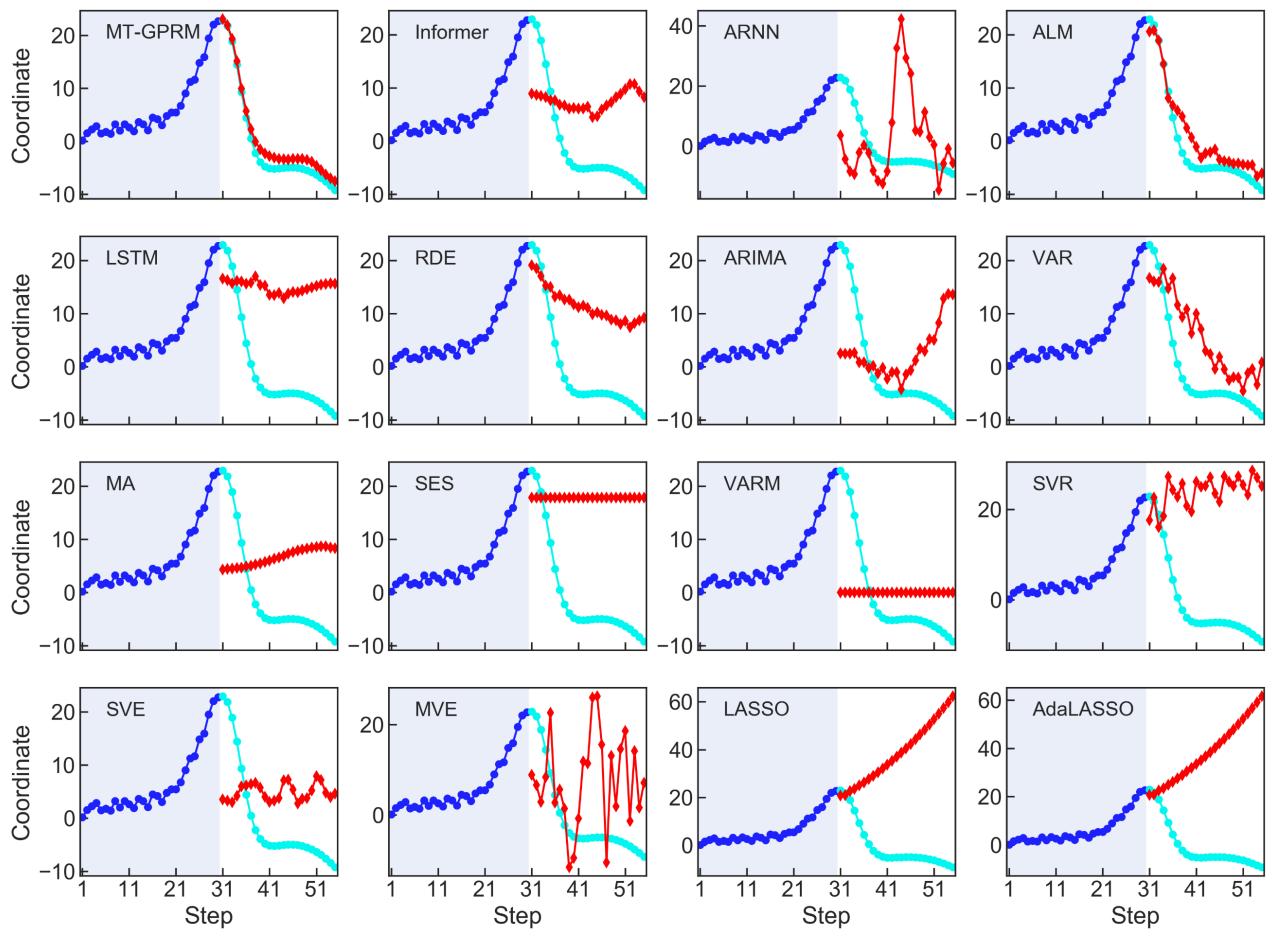


Fig. S11. Predicted results of all 16 methods on a 90-dimension time-variant coupled Lorentz system (target variable is x_{17} and Gaussian noise $\sim N(0, 3)$).

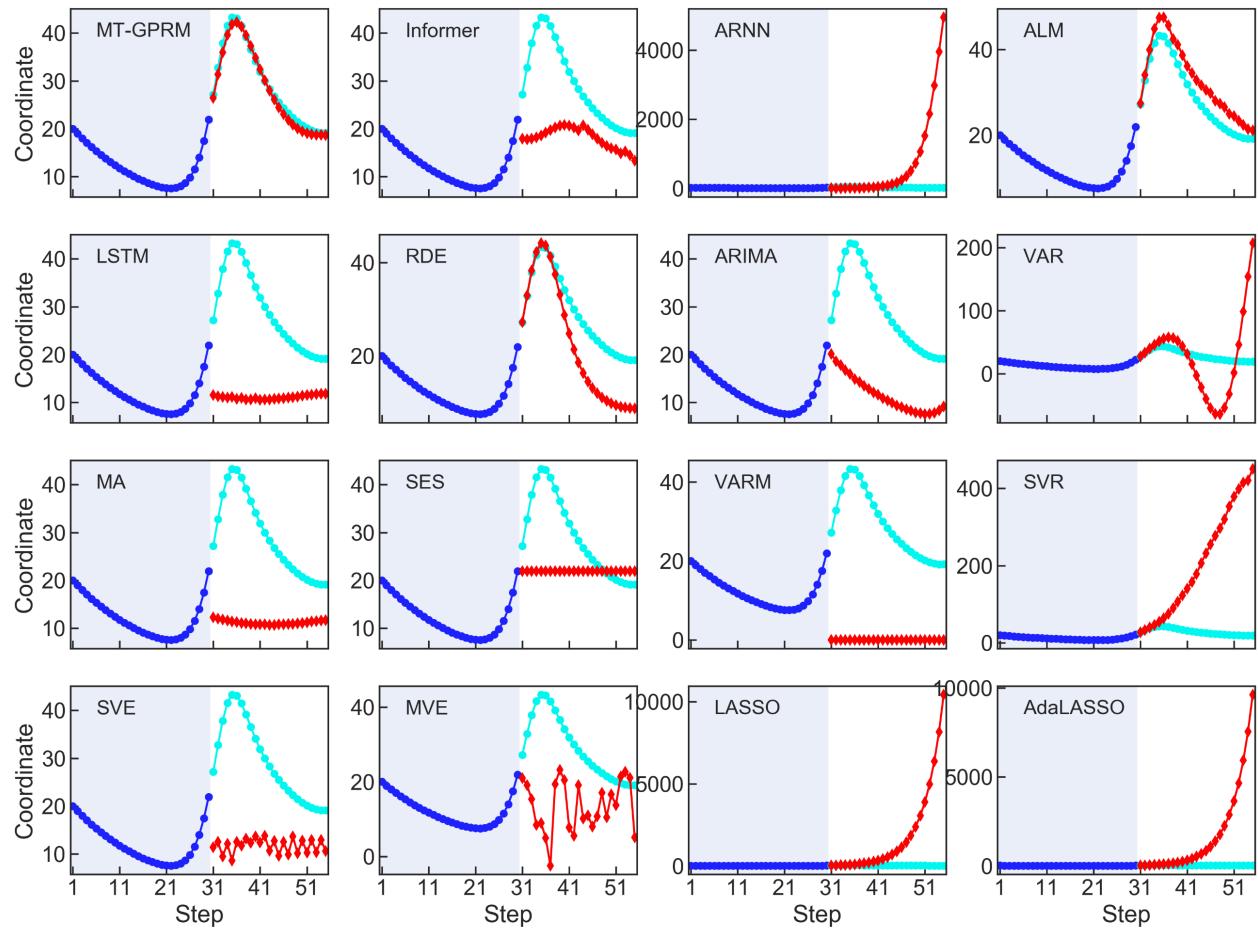


Fig. S12. Predicted results of all 16 methods on a 90-dimension time-variant coupled Lorentz system (target variable is x_{-18}).

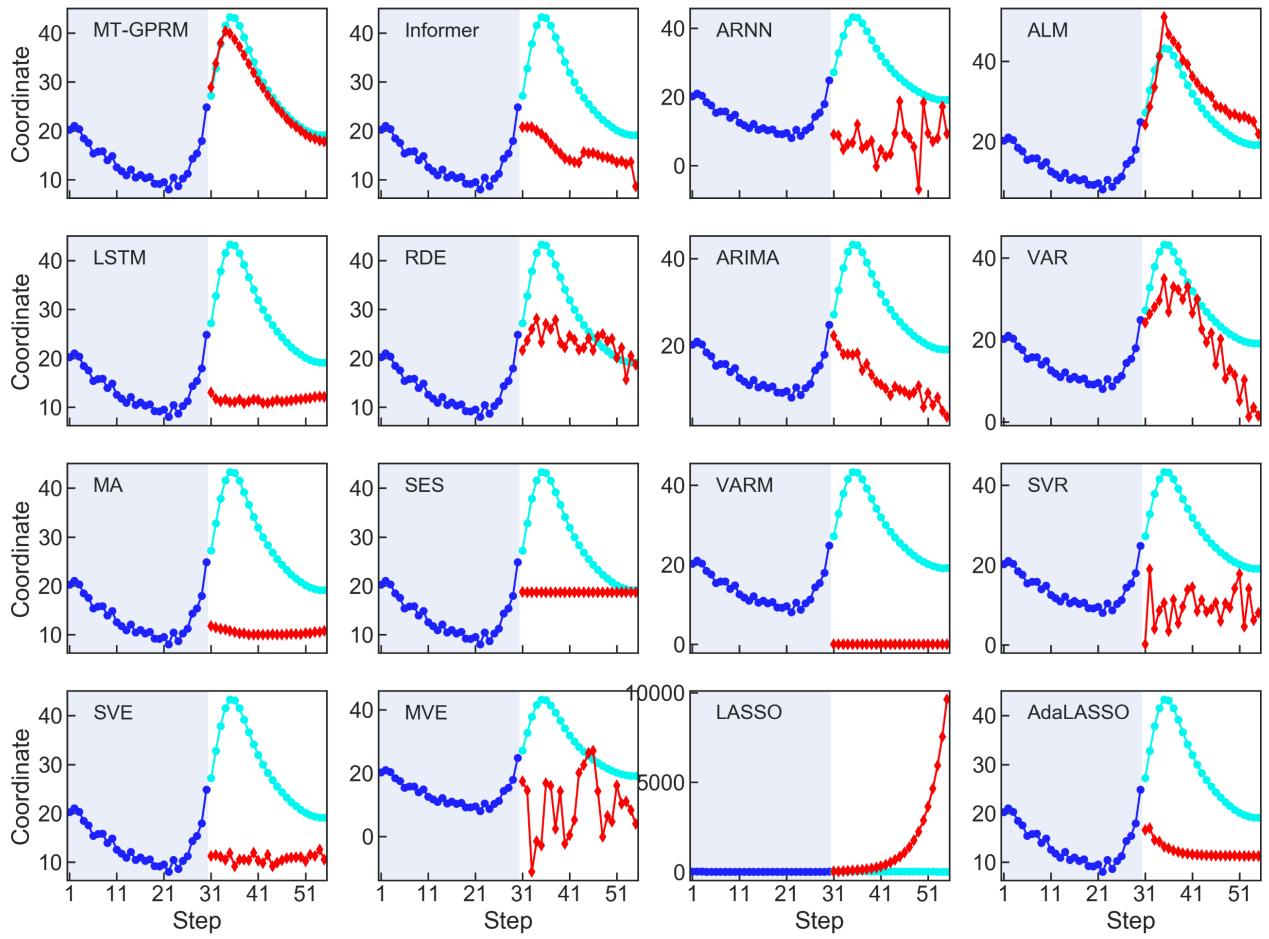


Fig. S13. Predicted results of all 16 methods on a 90-dimension time-variant coupled Lorentz system (target variable is x_{18} and Gaussian noise $\sim N(0, 3)$).

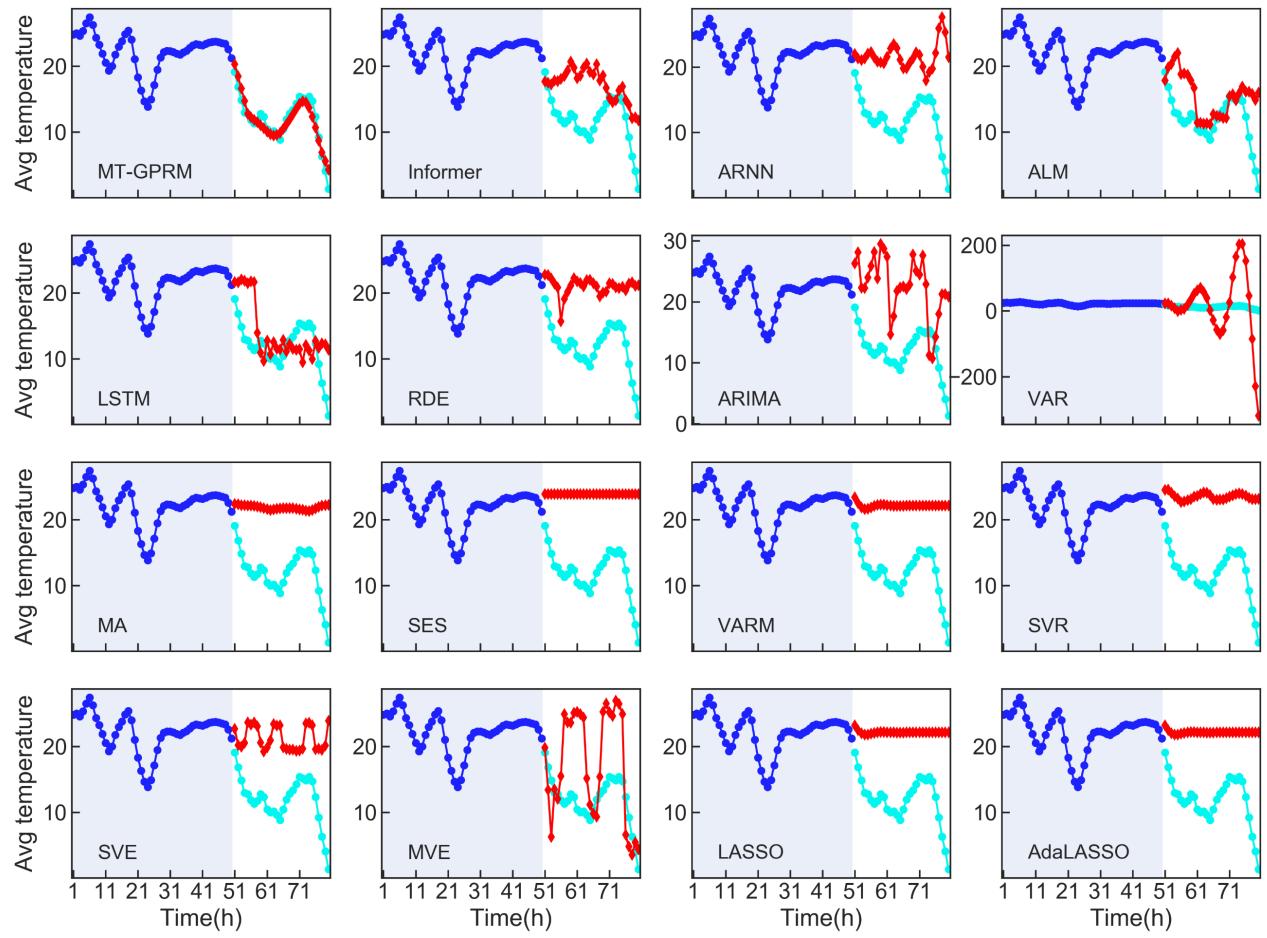


Fig. S14. Predicted results of all 16 methods on the plankton dataset.

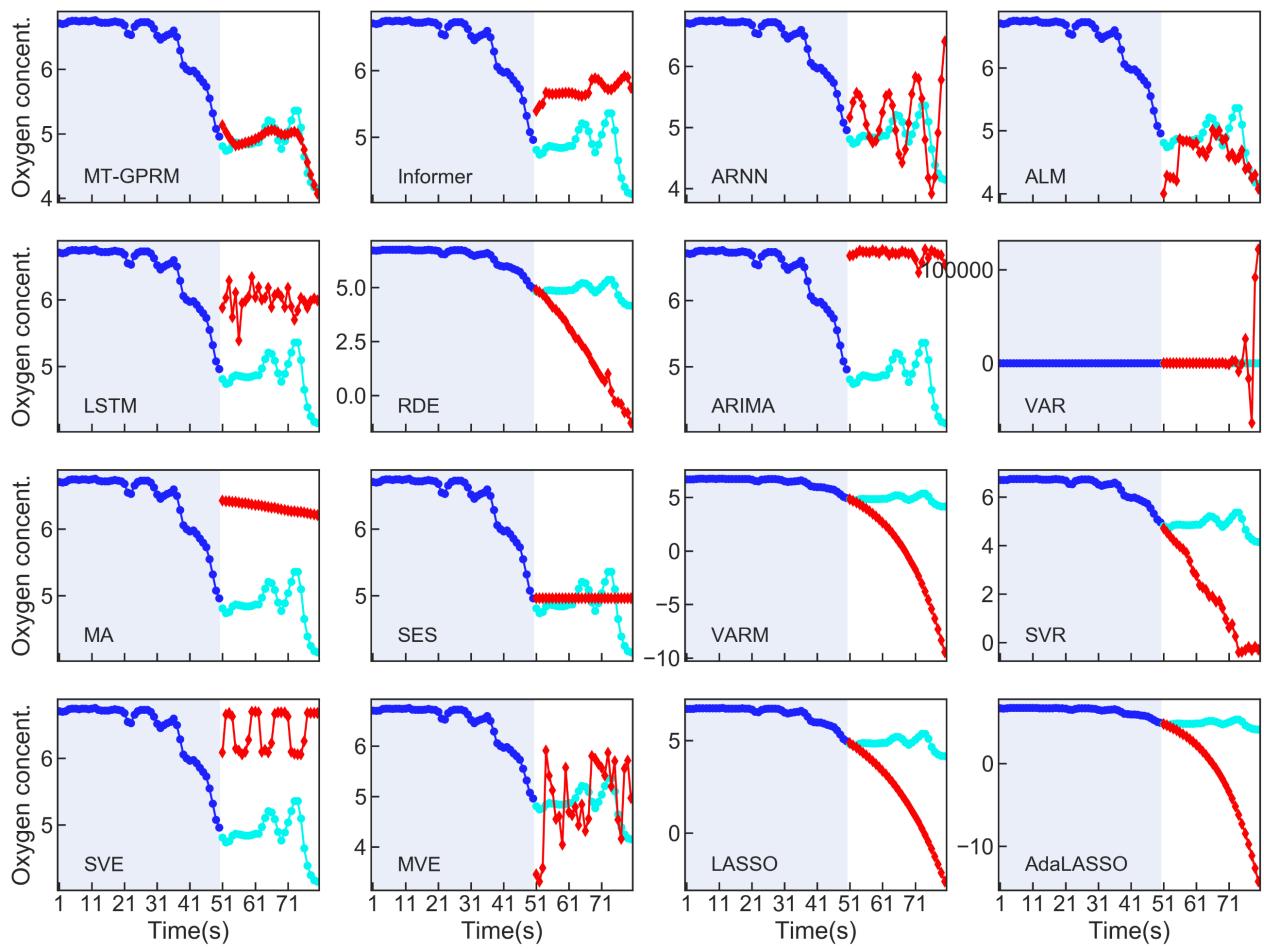


Fig. S15. Predicted results of all 16 methods on the ozone level dataset.

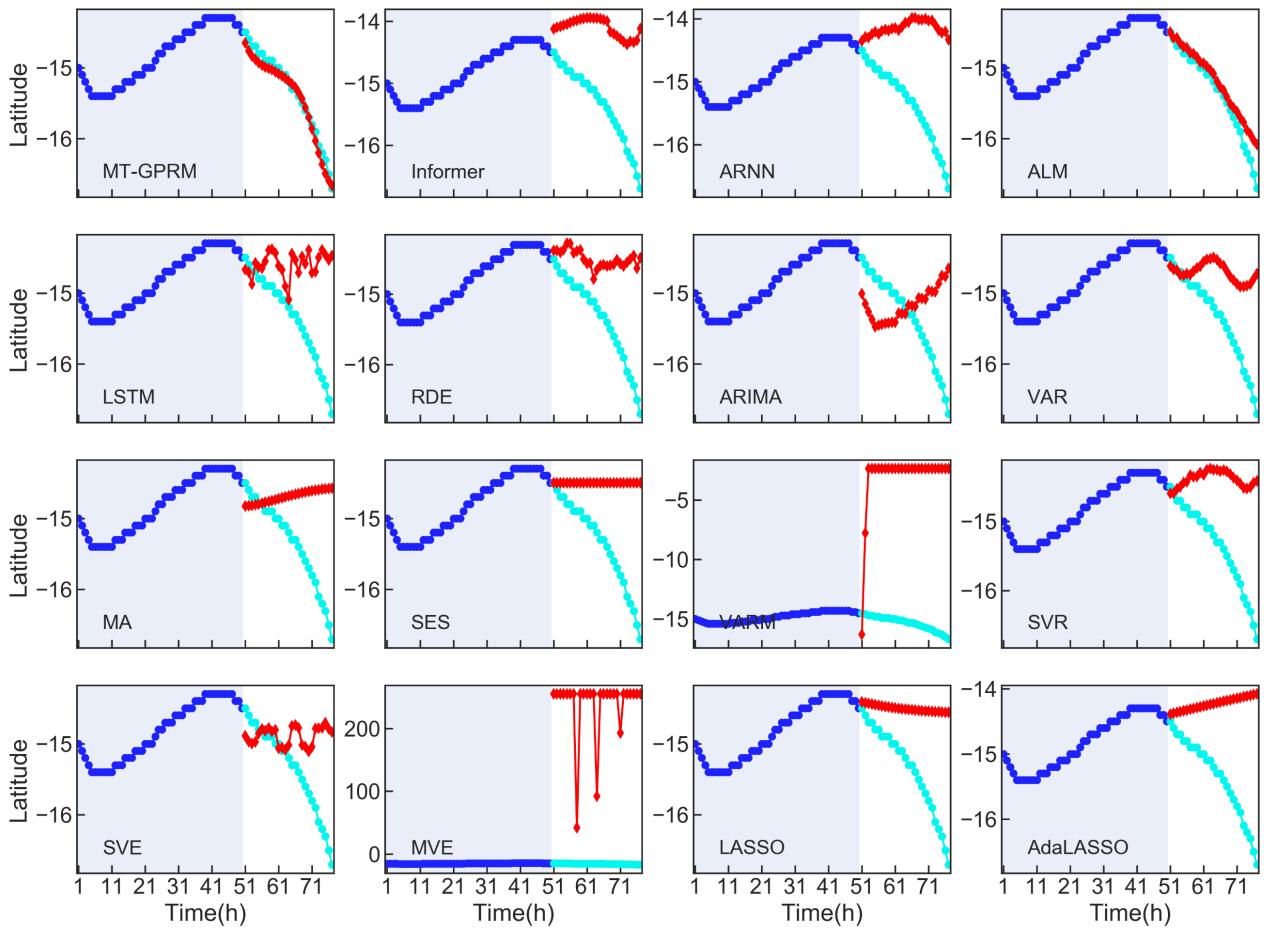


Fig. S16. Predicted results of all 16 methods on the typhoon eye trace dataset (the latitude part).

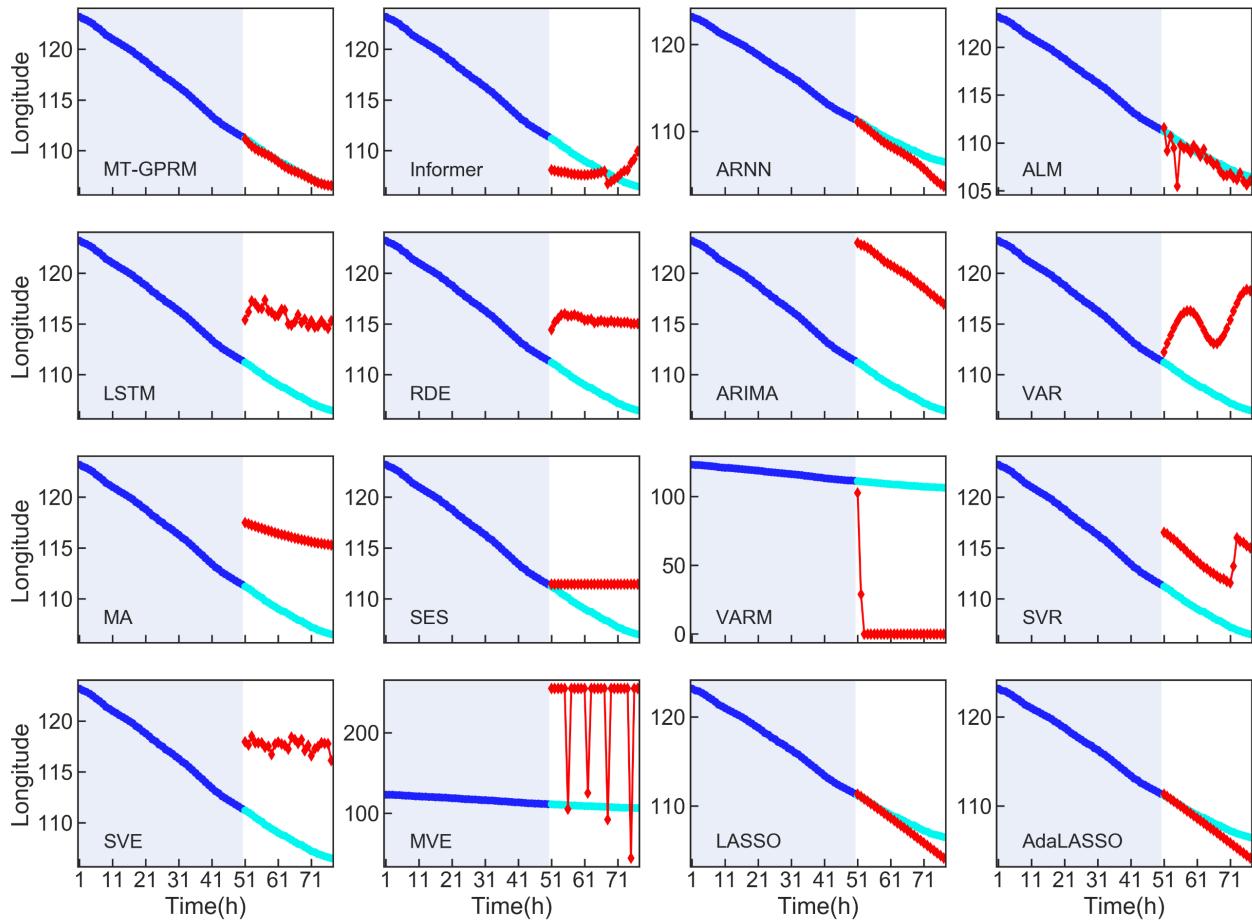


Fig. S17. Predicted results of all 16 methods on the typhoon eye trace dataset (the longitude part).