A APPENDIX

A.1 ABLATION STUDIES ON PREDEFINED MEASUREMENT FUNCTIONS

Table 5 shows the results from the ablation study of measurement function on M4-Yearly data. G represents generic, purely data-driven. P is polynomial function, E is the exponential function, S is the sine function. The integers in the first row are the number of each type of function. Basically, we gradually add more and different types of predefined measurement functions to purely data-drive model and observe improvements brought by the polynomial, exponential function, sine functions.

KNF	G	P1	P2	P3	P2+E1	P2+E2	P2+E3	P2+E1+S1	P2+E1+S2	P2+E1+S3	P2+E1+S4
sMAPE	14.66	14.65	14.56	14.79	14.48	14.63	14.66	14.35	14.16	14.01	14.15

Table 5: Ablation study of measurement function on M4-Yearly data. G represents generic, purely data-driven. P is polynomial function, E is the exponential function, S is the sine function. The integers in the first row are the number of each type of function.

A.2 SIMPLE NONLINEAR SYSTEM WITH FINITE KOOPMAN SPACE

We consider the following simple nonlinear system with discrete spectrum.

$$\dot{x_1} = \mu x_1$$
$$\dot{x_2} = \lambda(x_2 - x_1^2)$$

This system has a minimal Koopman invariant subspace spanned by $\{x_1, x_2, x_1^2\}$:

$$\frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \\ x_1^2 \end{bmatrix} = \begin{bmatrix} \mu & 0 & 0 \\ 0 & \lambda & -\lambda \\ 0 & 0 & 2\mu \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_1^2 \end{bmatrix}$$

It can also be spanned by three eigenfunctions $[\phi_1, \phi_2, \phi_3] = [x_1, x_1^2, x_2 - \frac{\lambda}{\lambda - 2\mu}x_1^2]$

$$\frac{d}{dt}\begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \end{bmatrix} = \begin{bmatrix} \mu & 0 & 0 \\ 0 & 2\mu & 0 \\ 0 & 0 & \lambda \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \end{bmatrix}$$

Any multiplication of $[\phi_1, \phi_2, \phi_3]$ is still an eigenfunction.

We want to show our model with only trainable global operator can correctly identify the Koopman invariant subspace from the synthetic data generated based on this system. We generate 36 time series with $\mu = -0.1$ and $\lambda = -1$ and the initial values are uniformly sampled from $[-1, 1]^2$. We tried four different basis dictionaries, including $D_1 = \{x_1, x_2, x_1^2\}$, $D_2 = \{x_1, x_2, x_1^2, x_2^2\}$, $D_3 = \{x_1, x_2, x_1^2, x_2^2, x_1^3, x_2^3\}$, $D_4 = \{x_1, x_2, x_1^2, x_2^2, x_1^3, x_2^3, x_1^4, x_2^4\}$ and we trained and tested the models on different initial conditions. The left figure in Fig. 5 shows eigenvalues learnt by KNF and some true eigenvalues. Since the learned Koopman matrix may not be unique given the flexibility in the coefficients of measurement functions, and the compositions of eigenfunctions are also eigenfunctions, there are many possible eigenvalues. But we can still see that most of the learned eigenvalues match the true eigenvalues. The right three figures in Fig. 5 shows that KNF can make accurate predictions with different dictionaries.

A.3 ADDITIONAL EXPERIMENTAL DETAILS

Table 6 shows the hyperparameter tuning ranges, including the learning rate, the hidden dimension and number of layers of deep neural network modules in both baselines and our model, number of predictions made at each autoregressive step, the length of input observations, the forecasting window size during training, and whether to use ReVIN. For different modules in a model, we tune hyperparameters, such as the number of layers/hidden dimensions separately. For instance, in FedFormer, the encoder and decoder may have different numbers of layers. As for the nonlinearity, we use ReLU for all layers.

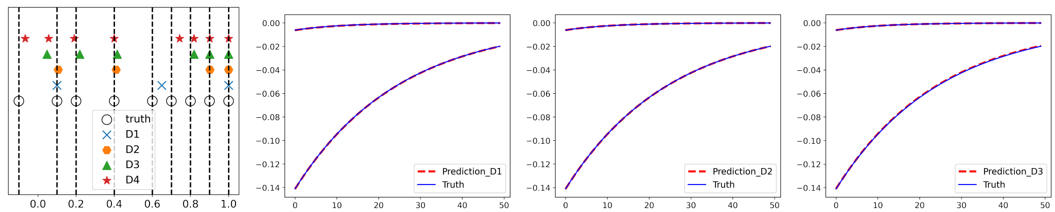

Figure 5: Left: Learnt eigenfunctions by KNF with different dictionaries. Right: Predictions from KNF with on the simple nonlinear system.

Learning rate	Hidden dim	#Layers	#forecasting window size during training	#Predictions made at each autoregressive step	Input length	Whether to use ReVIN
1e-1∼1e-5	64∼1024	3∼7	1∼15	1∼10	5∼50	True/False

Table 6: Hyperparamter Tuning Ranges.

A.4 Additional Results

sMAPE	Monthly(18)	Weekly(13)	Daily(14)	Hourly(48)	Yearly(6)	Quarterly(8)
`Montero et.al`	12.639	7.625	3.097	11.506	13.528	9.733
`Smyl`	12.126	7.817	3.170	**9.328**	13.176	9.679
`Nbeats-I+G`	12.024	-	-	-	**12.924**	**9.212**
`KNF`	**11.930**	**7.254**	**2.990**	11.294	13.800	10.008

Table 7: sMAPE for `KNF` and baselines on six M4 datasets for different sampling frequencies. The numbers in parentheses are the number of prediction steps. `KNF` achieves the state-of-the-art prediction performance at Weekly, Daily and Monthly frequencies.