Koopa: Learning Non-stationary Time Series Dynamics with Koopman Predictors

Yong Liu*, Chenyu Li*, Jianmin Wang, Mingsheng Long✉
School of Software, BNRist, Tsinghua University, China
{liuyong21,lichenyu20}@mails.tsinghua.edu.cn, {jimwang,mingsheng}@tsinghua.edu.cn

Yong Liu

Chenyu Li

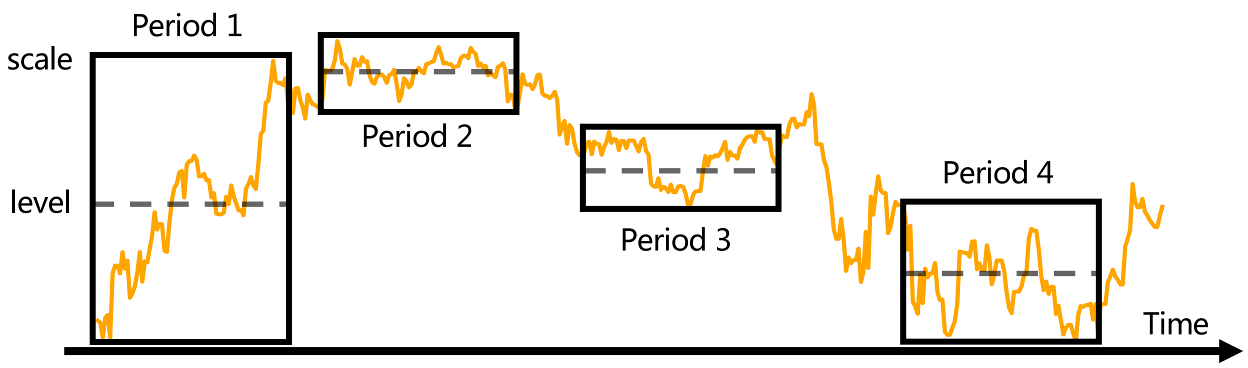
Jianmin Wang

Mingsheng Long

Non-stationary Time Series

Real-world series are always non-stationary, making the forecast extremely hard

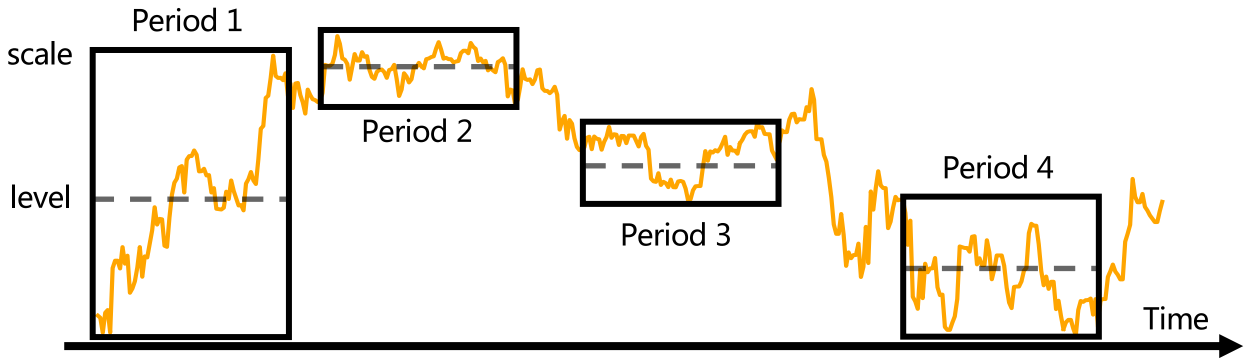
- Complicated series variations –> Challenges the model capacity
- Time-variant distribution -> Deep Models struggle to generalize well



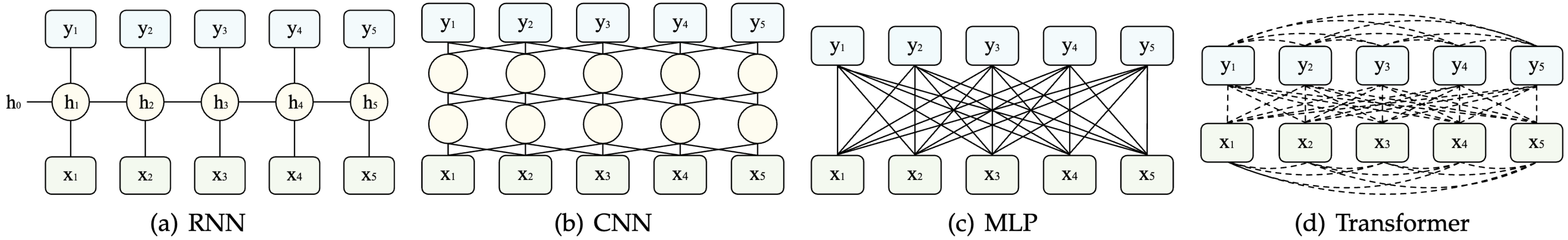
Non-stationary Time Series

Real-world series are always non-stationary, making the forecast extremely hard

- Complicated series variations –> Challenges the model capacity
- Time-variant distribution -> Deep Models struggle to generalize well



Few model tackles non-stationary series considering the inherent properties



Non-stationary Series as Dynamics

Real-world time series act like time-variant dynamics

- Complicated series variations –> Non-linear dynamical system (can be simplified as LDS)
- Time-variant distribution -> Multiple Localized Koopman operators

Koopman Theory

Dynamical System / Dynamics

$$x_{t+1} = \mathbf{F}(x_t)$$

Describes the transition of system state

$\mathbf{F}$ is hard to identify because of nonlinearity, but simple in LDS (linear operator/matrix)

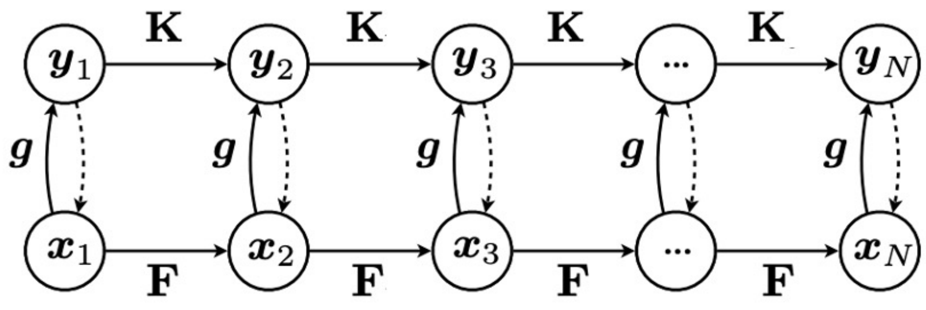
Koopman Operator

$$\mathcal{K} \circ g(x_t) = g\big(\mathbf{F}(x_t)\big) = g(x_{t+1})$$

Projected states by measurement function g

and governed by a linear operator $\mathcal{K}$

Koopman theory bridges nonlinear dynamics and high-dimensional LDS



$\mathbf{F} : \boldsymbol{x}_k \mapsto \boldsymbol{x}_{k+1}$

$\boldsymbol{g} : \boldsymbol{x}_k \mapsto \boldsymbol{y}_k$

$\mathbf{K} : \boldsymbol{y}_k \mapsto \boldsymbol{y}_{k+1}$

Cope with complicated series variations by linear layers

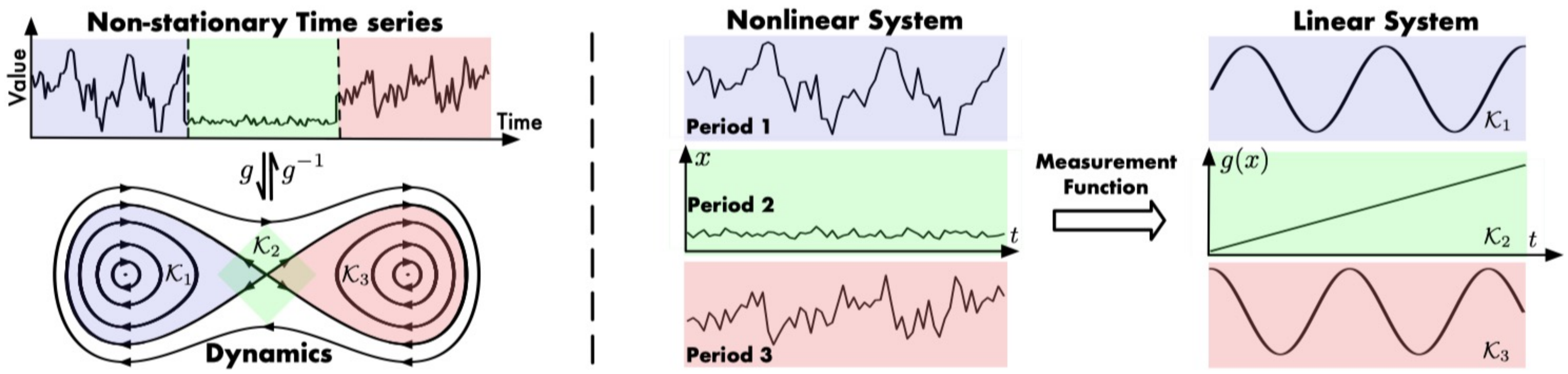
Cope with Non-stationarity

A complicated dynamical system can be time-variant

- After projected to LDS, the transitions can still differ in regions $\mathcal{K}$ -> $\mathcal{K}_t$
- Modern Koopman theory utilizes multiple localized operators to describe regions

🌟 Time series as dynamics: **transitions differ in periods**

discriminately portrayed by multiple linear operators



Disentanglement

Wold's Theorem

$$X_t = \eta_t + \sum_{j=0}^{\infty} b_j \varepsilon_{t-j}$$

holds for every weak-stationary time series X_t in the input window

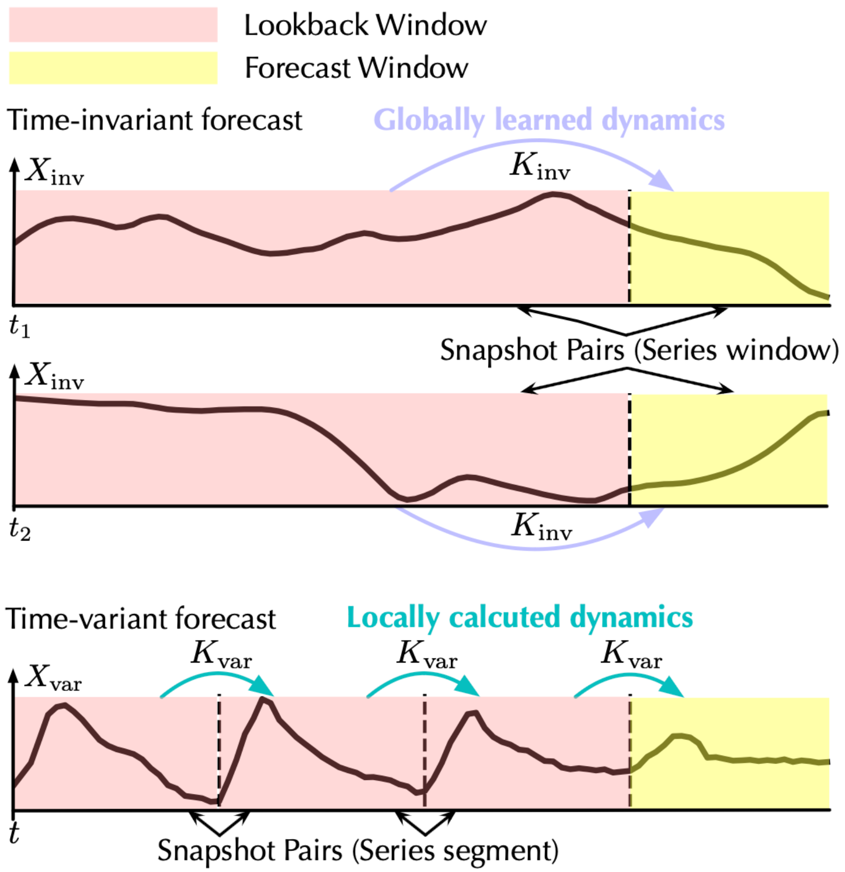
η_t - Deterministic component with long-term invariance (e.g. sin(t))

ε_t - White noise as the stationary input of linear filter $\{b_j\}$ (varies in periods / windows)

Inspirations

- Time series can be decomposed into time-invariant and time-variant parts
- Two components should be portrayed in different ways

Koopa



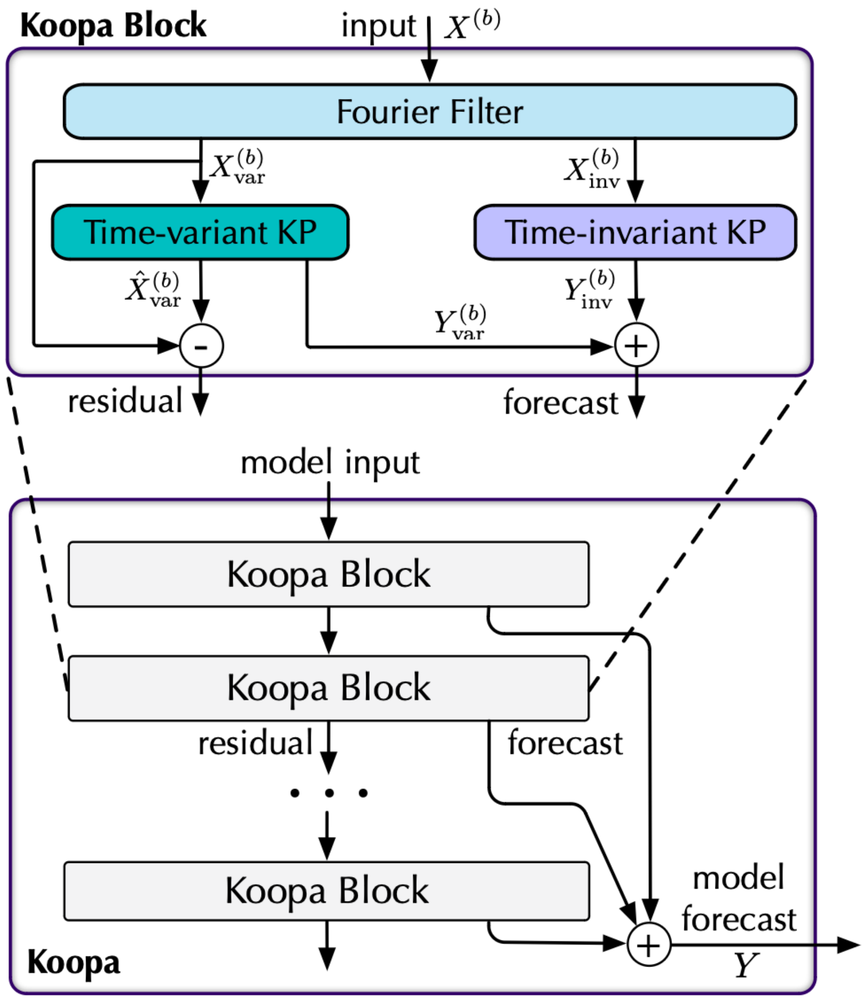
Disentanglement

- Disentangle time-invariant and time-variant dynamics
- Utilizes frequency statistics based on Fourier analysis

Two Koopman Predictors (KP)

- Globally learned parameter as operator
- Locally calculated operator within a short period

Deep Residual Structure



Koopa Block

$$X_{\text{var}}^{(b)},\ X_{\text{inv}}^{(b)} = \text{FourierFilter}(X^{(b)})$$

$$Y_{\text{inv}}^{(b)} = \text{TimeInvKP}(X_{\text{inv}}^{(b)})$$

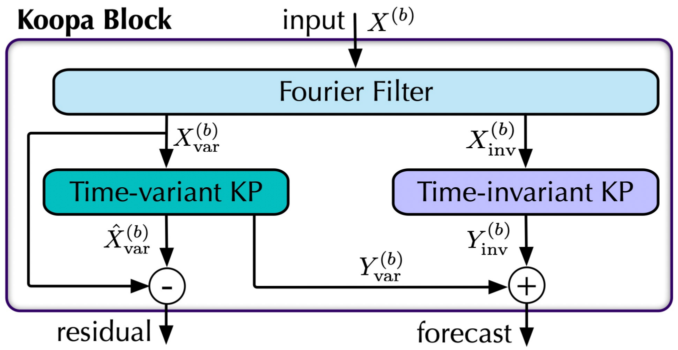
$$\hat{X}_{\text{var}}^{(b)}, Y_{\text{var}}^{(b)} = \text{TimeVarKP}(X_{\text{var}}^{(b)})$$

Portray the residual of previously fitted dynamics

- Enhance model capability
- Benefit training with stable operators

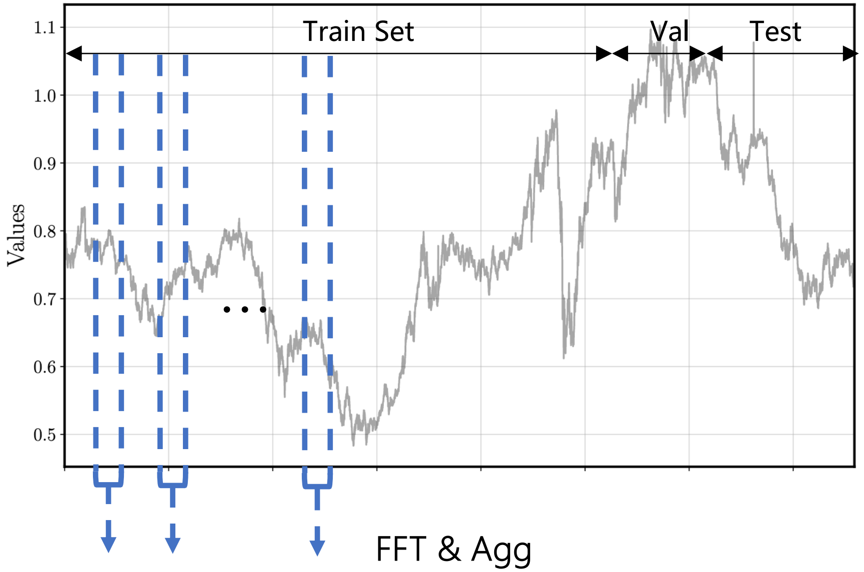
$$X^{(b+1)} = X_{\text{var}}^{(b)} - \hat{X}_{\text{var}}^{(b)},\ \ Y = \sum \left(Y_{\text{var}}^{(b)} + Y_{\text{inv}}^{(b)}\right)$$

Fourier Filter



Disentangle by time-agnostic properties

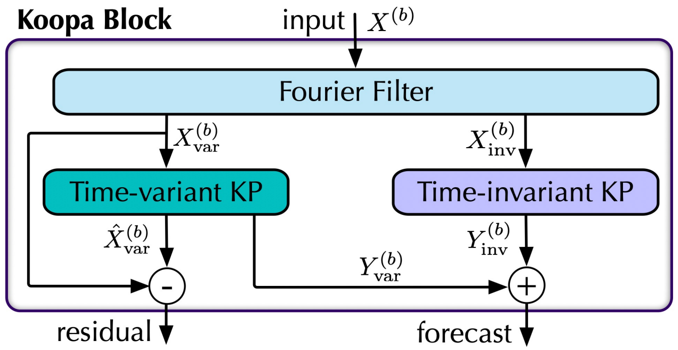
- Statistics: Level, Scale, Moments….
- Period, Frequency, Spectrums



1. Compute FFT of each window of the training set
2. Calculate the averaged amplitude of each spectrum $\mathcal{S}$
3. Take the top percent of $\boldsymbol{\alpha}$ as the subset $\mathcal{G}_\alpha \subset \mathcal{S}$
   - Contains dominant spectrums shared among all windows

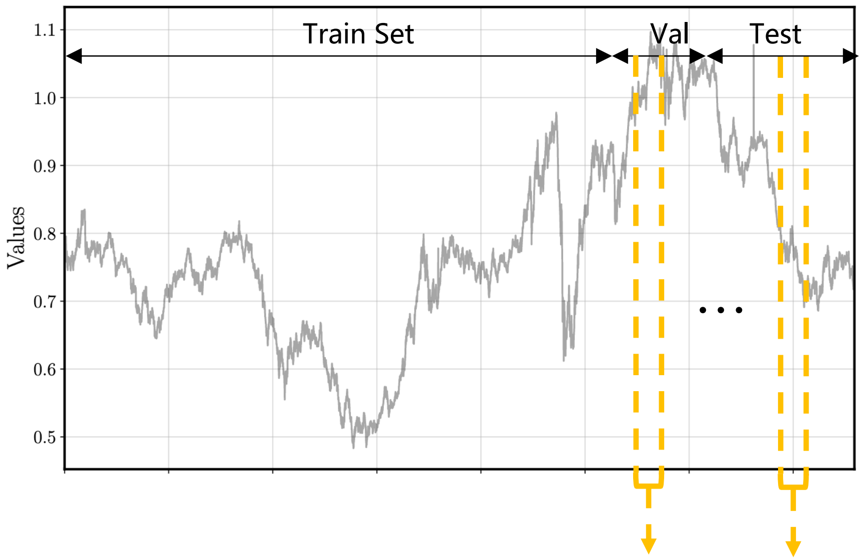
Window Set —FFT & Agg→ Amplitude distribution of Spectrums —Truncate→ Spectrums shared among periods

Fourier Filter



Disentangle by time-agnostic properties

- Statistics: Level, Scale, Moments….
- Period, Frequency, Spectrums



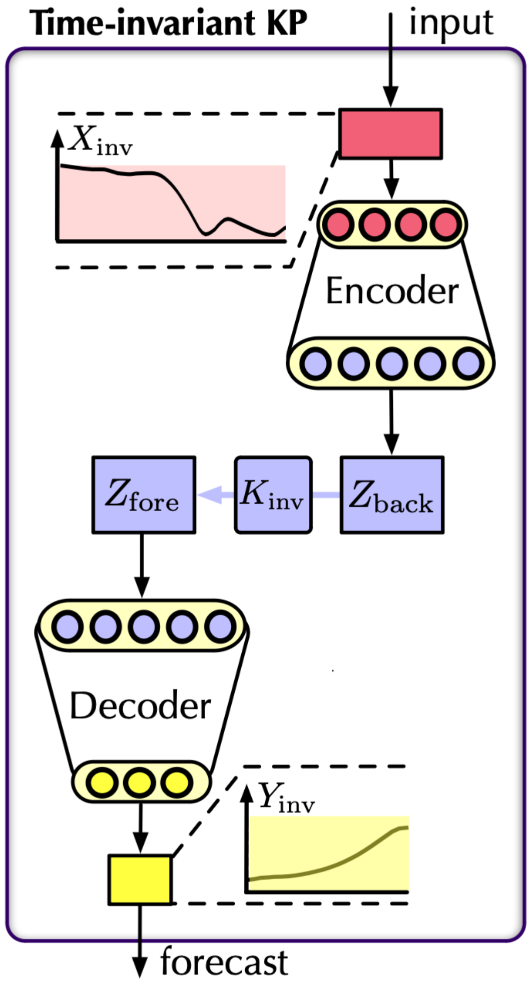
4. Disentangle by $\mathcal{G}_\alpha$ and its complementary $\bar{\mathcal{G}}_\alpha$

$$X_{\text{inv}} = \mathcal{F}^{-1}\big(\text{Filter}\left(\mathcal{G}_\alpha,\ \mathcal{F}(X)\right)\big)$$

$$X_{\text{var}} = \mathcal{F}^{-1}\big(\text{Filter}\left(\bar{\mathcal{G}}_\alpha,\ \mathcal{F}(X)\right)\big) = X - X_{\text{inv}}$$

Input Series $\xrightarrow{\text{Filter}}$ Time-invariant & Time-variant componets

Time-invariant Koopman Predictor



Globally shared dynamics from the lookback to forecast window

$$\mathbf{F} : X_{\text{inv}} \mapsto Y_{\text{inv}} \qquad \mathbb{R}^{T\times C} \mapsto \mathbb{R}^{H\times C}$$

- Utilize Encoder/Decoder to learn the measurement function

$$Z_{\text{back}} = \text{Encoder}(X_{\text{inv}}) \qquad \mathbb{R}^{T\times C} \mapsto \mathbb{R}^{D}$$

$$Y_{\text{inv}} = \text{Decoder}(Z_{\text{fore}}) \qquad \mathbb{R}^{D} \mapsto \mathbb{R}^{H\times C}$$

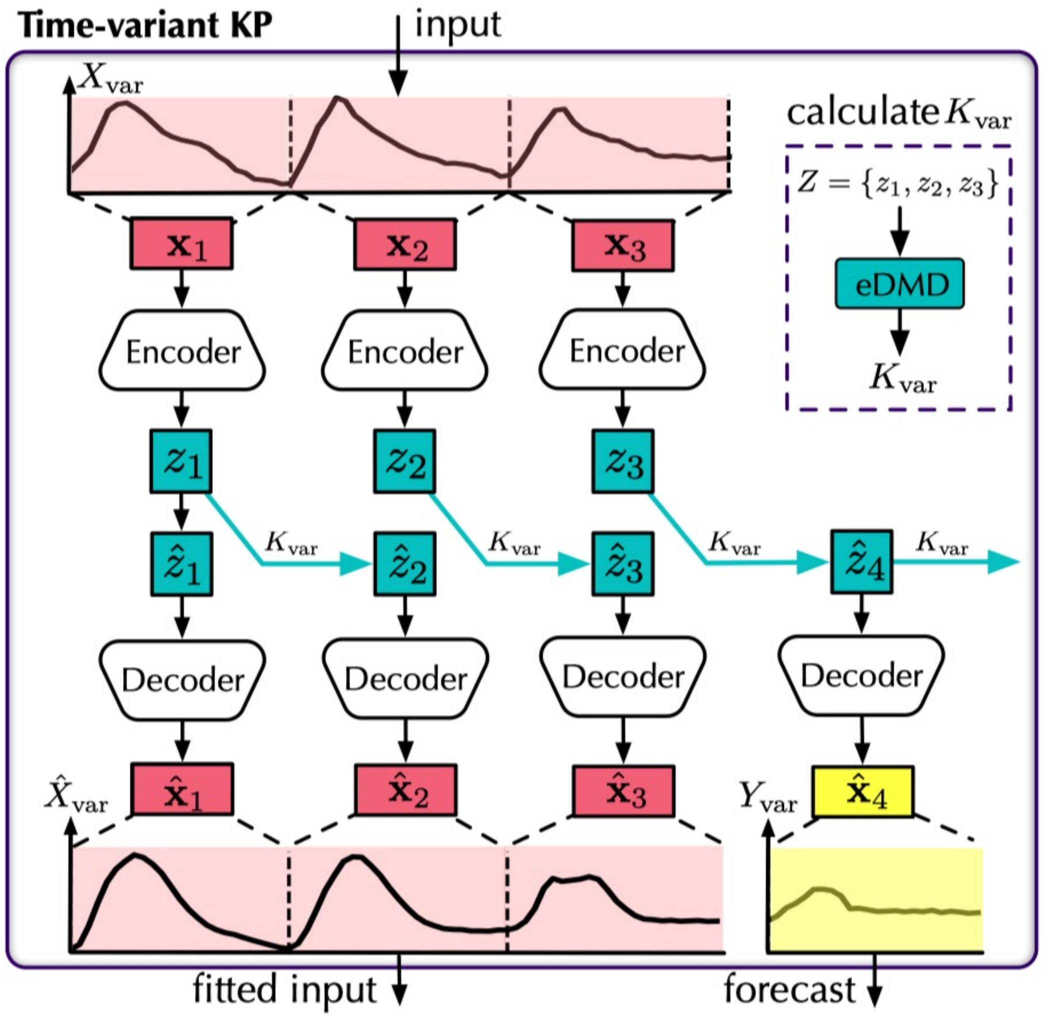
- Operator as a learnable parameter

$$Z_{\text{fore}} = K_{\text{inv}} Z_{\text{back}} \qquad \mathbb{R}^{D} \mapsto \mathbb{R}^{D}$$

- No reconstruction branch like previous KAE

Time-variant Koopman Predictor

Localized dynamics within the lookback window and advance forward to forecast



$$\mathbf{F}:\mathbf{x}_t\mapsto\mathbf{x}_{t+1}\qquad \mathbb{R}^{S\times C}\mapsto\mathbb{R}^{S\times C}$$

- Segment the lookback time series

$$\mathbf{x}_j=[x_{(j-1)S+1},\ldots,x_{jS}]^\top$$

- Utilize Encoder/Decoder

$$z_j=\text{Encoder}(\mathbf{x}_j),\ \hat{\mathbf{x}}_j=\text{Decoder}(\hat{z}_j)$$

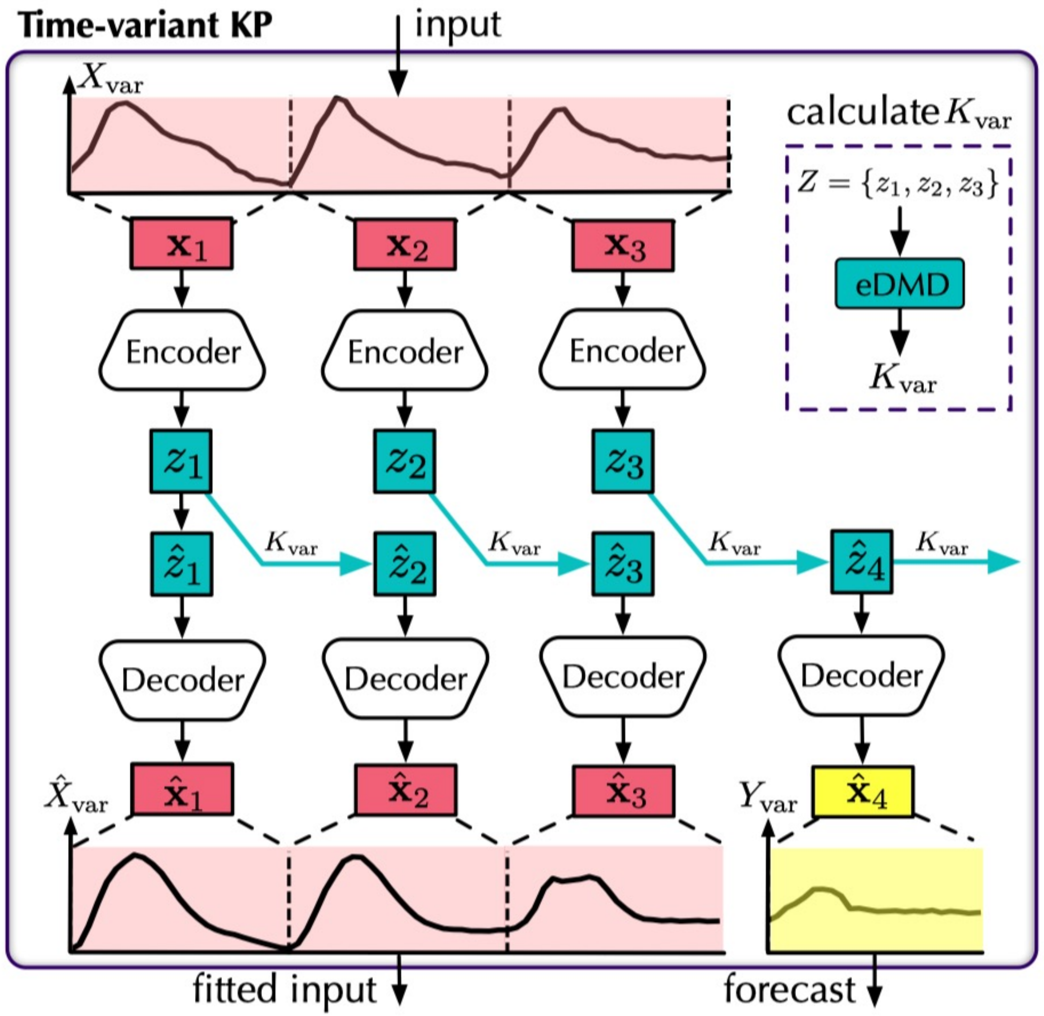
- Calculate operator by eDMD

$$Z_{\text{back}}=[z_1,z_2,\ldots,z_{\frac{T}{S}-1}]\quad Z_{\text{fore}}=[z_2,z_3,\ldots,z_{\frac{T}{S}}]$$

$$\Rightarrow\quad K_{\text{var}}=Z_{\text{fore}}Z_{\text{back}}^{\dagger}$$

Time-variant Koopman Predictor

Localized dynamics within the lookback window and advance forward to forecast



$$\mathbf{F} : \mathbf{x}_t \mapsto \mathbf{x}_{t+1} \quad \mathbb{R}^{S \times C} \mapsto \mathbb{R}^{S \times C}$$

- Reconstruct and forward advance dynamics

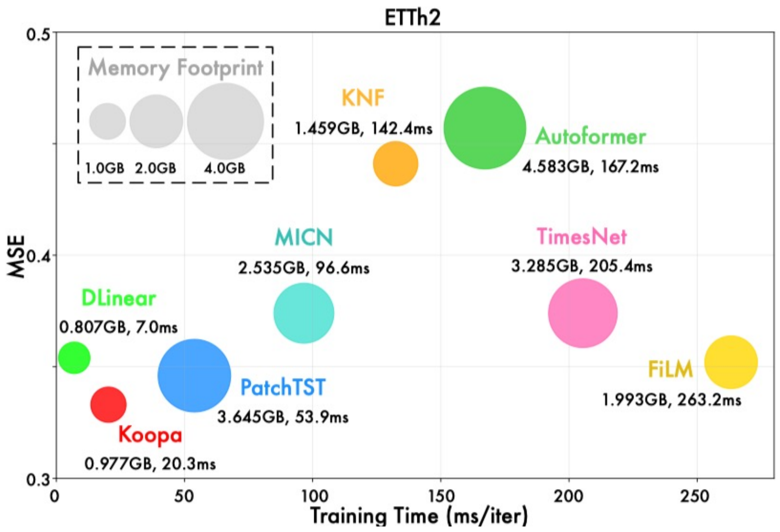
$$[\hat{z}_1, \hat{z}_2, \ldots, \hat{z}_{\frac{T}{S}}] = [z_1, K_{\text{var}} Z_{\text{back}}]$$

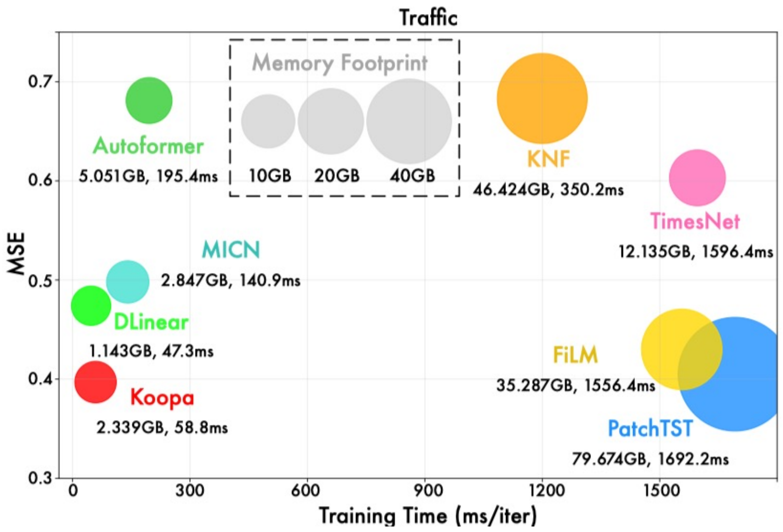
$$\hat{z}_{\frac{T}{S}+t} = (K_{\text{var}})^t z_{\frac{T}{S}}$$

- Rearrange segments as fitted input and forecast

$$\hat{X}_{\text{var}} = [\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_{\frac{T}{S}}]^\top, \; Y_{\text{var}} = [\hat{\mathbf{x}}_{\frac{T}{S}+1}, \ldots, \hat{\mathbf{x}}_{\frac{T}{S}+\frac{H}{S}}]^\top$$

Time Series Forecasting





Balanced Performance/Efficiency

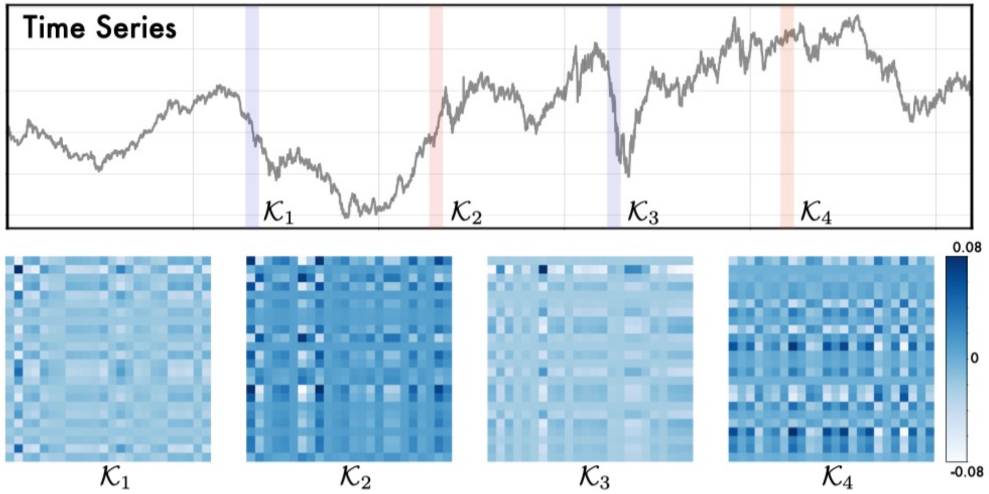
- Comparable to SOTA PatchTST
- Averaged Saving
  - 77.3% training cost
  - 76.0% memory footprint

	Models	**Koopa**	N-HiTS	N-BEATS	PatchTST	TimesNet	DLinear	MICN	KNF	FiLM	Autoformer
Weighted Average	sMAPE	**11.863**	11.960	11.910	13.022	11.930	12.418	13.023	12.126	12.489	14.057
	MASE	**1.595**	1.606	1.613	1.814	1.597	1.656	1.836	1.641	1.690	1.954
	OWA	**0.858**	0.861	0.862	0.954	0.867	0.891	0.960	0.874	0.902	1.029

Achieve consistent **state-of-the-art** on univariate forecasting

- Comparable to Fast DLinear
- Averaged Reducing
  - 12.3% MSE

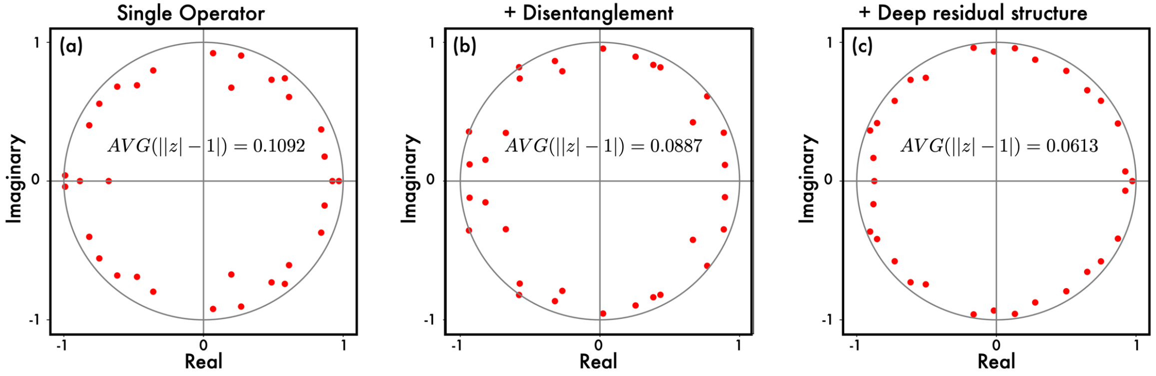
Operator Analysis



Operator Visualization

- Localized operators exhibit changing series variations in different periods
- Trending v.s. Heatmap values

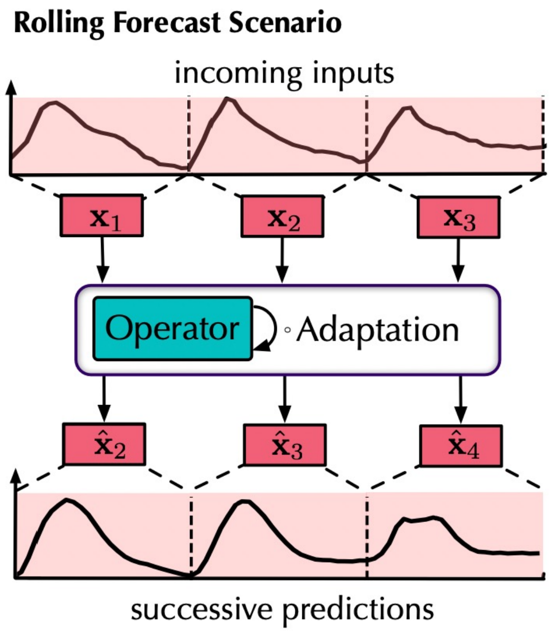
Eigenvalues Analysis



Disentanglement and deep residual structure

- More stable operators
- Improved the training stability

Scale Up Forecasting Horizon



Most deep forecasters have fixed prediction length once trained

Koopa conducts rolling forecast while adapting to varying dynamics

Dataset ADF Test Statistic	Exchange (-1.889)		ETTh2 (-4.135)		ILI (-5.406)		ECL (-8.483)		Traffic (-15.046)		Weather (-26.661)	
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Koopa	0.214	0.348	0.437	0.429	2.836	1.065	0.199	0.298	0.709	0.437	0.237	0.276
Koopa OA	**0.172**	**0.319**	**0.372**	**0.404**	**2.427**	**0.907**	**0.182**	**0.271**	**0.699**	**0.426**	**0.225**	**0.264**
Promotion (MSE)	19.6%		14.9%		14.1%		8.5%		1.4%		5.1%	

- Operator Adaptation improves the accuracy without training
- Significant promotion made in non-stationary time series

Speed Up

Algorithm 2 Accelerated Koopa Operator Adaptation.

Require: Observed embedding $Z = [z_1, \ldots, z_F]$ and successively incoming ground truth embedding $[z_{F+1}, \ldots, z_{F+L}]$ with each embedding $z_i \in \mathbb{R}^D$.

$Z_{\text{back}} = [z_1, \ldots, z_{F-1}], Z_{\text{fore}} = [z_2, \ldots, z_F]$ ▷ $Z_{\text{back}}, Z_{\text{fore}} \in \mathbb{R}^{D \times (F-1)}$

$K_{\text{var}} = Z_{\text{fore}} Z_{\text{back}}^{\dagger}, X = Z_{\text{back}} Z_{\text{back}}^{\dagger}$ ▷ $K_{\text{var}}, X \in \mathbb{R}^{D \times D}$

$\hat{z}_{F+1} = K_{\text{var}} n$ ▷ $\hat{z}_{F+1} \in \mathbb{R}^D$

for l **in** $\{1, \ldots, L\}$**:** ▷ z_{F+l} comes successively

  $m = z_{F+l-1}, n = z_{F+l}$ ▷ $m, n \in \mathbb{R}^D$

  $r = m - Xm$ ▷ $r \in \mathbb{R}^D$

  $b = r/||r||^2$ ▷ $b \in \mathbb{R}^D$

  $K_{\text{var}} \leftarrow K_{\text{var}} + (n - K_{\text{var}} m) b^{\top}$ ▷ $K_{\text{var}} \in \mathbb{R}^{D \times D}$

  $X \leftarrow X + r b^{\top}$ ▷ $X \in \mathbb{R}^{D \times D}$

  $\hat{z}_{F+l+1} = K_{\text{var}} n$ ▷ $\hat{z}_{F+l+1} \in \mathbb{R}^D$

End for

Return $[\hat{z}_{F+1}, \ldots, \hat{z}_{F+L+1}]^{\top}$ ▷ Return predicted embedding

Derived by linearity

Reduced Complexity

$\mathcal{O}(H_{\text{te}} D^3)$

$\mathcal{O}((H_{\text{te}} + D) D^2)$

Empowering Koopman forecasters on the long-term rolling forecast scenario for the first time.

Thank You!

liuyong21@mails.tsinghua.edu.cn

Code is available at https://github.com/thuml/Koopa