Multivariate Time Series Prediction Using Tensor Decompositions

Anastasia Sozykina

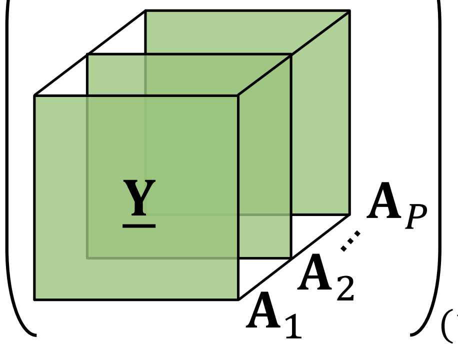
March, 2022

Skoltech

Tensor Vector Auto-Regression (VAR)

$$\mathbf{y}_t = \mathbf{A}_1 \mathbf{y}_{t-1} + \mathbf{A}_2 \mathbf{y}_{t-2} + \cdots + \mathbf{A}_P \mathbf{y}_{t-P} + \boldsymbol{\epsilon}_t$$

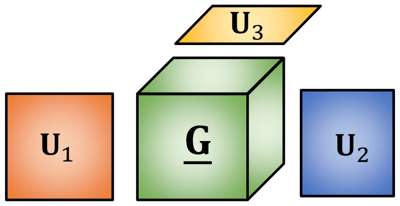
$\mathbf{y}_t = \Big($



$\Big)_{(1)} \mathbf{z}_\mathrm{t} + \boldsymbol{\epsilon}_t,$

$$\mathbf{z}_\mathrm{t} = \left(\mathbf{y}_{t-1}^{\mathrm{T}}, \cdots, \mathbf{y}_{\mathrm{t}-P}^{\mathrm{T}}\right)^{\mathrm{T}}$$

$\mathbf{y}_t = \Big($



$\Big)_{(1)} \mathbf{z}_\mathrm{t} + \boldsymbol{\epsilon}_t,$

$$\mathbf{y}_t = \mathbf{U}_1 \underline{\mathbf{G}}_{(1)} (\mathbf{U}_3 \otimes \mathbf{U}_2)^{\mathrm{T}} \mathbf{z}_t + \boldsymbol{\epsilon}_t = \mathbf{U}_1 \underline{\mathbf{G}}_{(1)} \mathrm{vec}(\mathbf{U}_2^T \mathbf{Z}_t \mathbf{U}_3) + \boldsymbol{\epsilon}_t,$$

where $\mathbf{Z}_t = (\mathbf{y}_{t-1}, \cdots, \mathbf{y}_{t-P})$

Tensor Vector Auto-Regression (VAR)

Multilinear low-rank least squares estimator

$$\underline{\mathbf{A}}_{MLR} \equiv \left[\!\left[\underline{\widehat{\mathbf{G}}}, \widehat{\mathbf{U}}_{\mathbf{1}}, \widehat{\mathbf{U}}_{\mathbf{2}}, \widehat{\mathbf{U}}_{\mathbf{3}} \right]\!\right] = \arg\min L(\underline{\mathbf{G}}, \mathbf{U}_{\mathbf{1}}, \mathbf{U}_{\mathbf{2}}, \mathbf{U}_{\mathbf{3}})$$

$$L(\underline{\mathbf{G}}, \mathbf{U}_{\mathbf{1}}, \mathbf{U}_{\mathbf{2}}, \mathbf{U}_{\mathbf{3}}) = \frac{1}{T} \sum_{t=1}^{T} \left\| \boldsymbol{y}_t - \left(\underline{\mathbf{G}} \times_1 \mathbf{U}_{\mathbf{1}} \times_{\mathbf{2}} \mathbf{U}_{\mathbf{2}} \times_{\mathbf{3}} \mathbf{U}_{\mathbf{3}} \right)_{(1)} \boldsymbol{x}_t \right\|_2^2$$

Skoltech

Tensor Vector Auto-Regression (VAR)

Multilinear low-rank least squares estimation

$F_{t-1} = \sigma(\boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_{\boldsymbol{t-1}}, \cdots)$) be the σ-field generated by $\{\boldsymbol{\epsilon}_s, s < t\}$, and
$\mathbf{X}_t = (\mathbf{y}_{t-1}, \cdots, \mathbf{y}_{t-P})$

$$\mathbb{E}(\mathbf{y}_\mathbf{t} \,|F_{t-1}) = \left(\left(\mathbf{x}_t^T(\mathbf{U}_3 \otimes \mathbf{U}_2)\underline{\mathbf{G}}_{(1)}^T\right) \otimes \mathbf{I}_N\right) vec(\mathbf{U}_1)$$

$$= \mathbf{U}_1\underline{\mathrm{G}}_{(1)}\left((\mathbf{U}_3^T\mathbf{X}_t^T) \otimes \mathbf{I}_{r2}\right) vec(\mathbf{U}_2^T) = \mathbf{U}_1\underline{\mathbf{G}}_{(1)}(\mathbf{I}_{r3} \otimes (\mathbf{U}_2^T\mathbf{X}_t))vec(\mathbf{U}_3)$$

$$= \left(\left((\mathbf{U}_3 \otimes \mathbf{U}_2)^T\mathbf{x}_t\right)^T \otimes \mathbf{U}_1\right) vec(\underline{\mathbf{G}}_{(1)})$$

Skoltech

Tensor Vector Auto-Regression (VAR) ALS

Algorithm Alternating least squares algorithm for $\widehat{\mathcal{A}}_{\mathrm{MLR}}$

Initialize: $\mathcal{A}^{(0)}$

HOSVD: $\mathcal{A}^{(0)} \approx \mathcal{G}^{(0)} \times_1 \boldsymbol{U}_1^{(0)} \times_2 \boldsymbol{U}_2^{(0)} \times_3 \boldsymbol{U}_3^{(0)}$ with multilinear ranks (r_1, r_2, r_3)

repeat $k = 0, 1, 2, \ldots$

$$\boldsymbol{U}_1^{(k+1)} \leftarrow \arg\min_{\boldsymbol{U}_1} \sum_{t=1}^{T} \|\boldsymbol{y}_t - ((\boldsymbol{x}_t'(\boldsymbol{U}_3^{(k)} \otimes \boldsymbol{U}_2^{(k)})\mathcal{G}_{(1)}^{(k)\prime}) \otimes \boldsymbol{I}_N)\mathrm{vec}(\boldsymbol{U}_1)\|_2^2$$

$$\boldsymbol{U}_2^{(k+1)} \leftarrow \arg\min_{\boldsymbol{U}_2} \sum_{t=1}^{T} \|\boldsymbol{y}_t - \boldsymbol{U}_1^{(k+1)}\mathcal{G}_{(1)}^{(k)}((\boldsymbol{X}_t\boldsymbol{U}_3^{(k)})' \otimes \boldsymbol{I}_{r_2})\mathrm{vec}(\boldsymbol{U}_2')\|_2^2$$

$$\boldsymbol{U}_3^{(k+1)} \leftarrow \arg\min_{\boldsymbol{U}_3} \sum_{t=1}^{T} \|\boldsymbol{y}_t - \boldsymbol{U}_1^{(k+1)}\mathcal{G}_{(1)}^{(k)}(\boldsymbol{I}_{r_3} \otimes (\boldsymbol{U}_2^{(k+1)\prime}\boldsymbol{X}_t))\mathrm{vec}(\boldsymbol{U}_3)\|_2^2$$

$$\mathcal{G}^{(k+1)} \leftarrow \arg\min_{\mathcal{G}} \sum_{t=1}^{T} \|\boldsymbol{y}_t - (((\boldsymbol{U}_3^{(k+1)} \otimes \boldsymbol{U}_2^{(k+1)})'\boldsymbol{x}_t)' \otimes \boldsymbol{U}_1^{(k+1)})\mathrm{vec}(\mathcal{G}_{(1)})\|_2^2$$

$$\mathcal{A}^{(k+1)} \leftarrow \mathcal{G}^{(k+1)} \times_1 \boldsymbol{U}_1^{(k+1)} \times_2 \boldsymbol{U}_2^{(k+1)} \times_3 \boldsymbol{U}_3^{(k+1)}$$

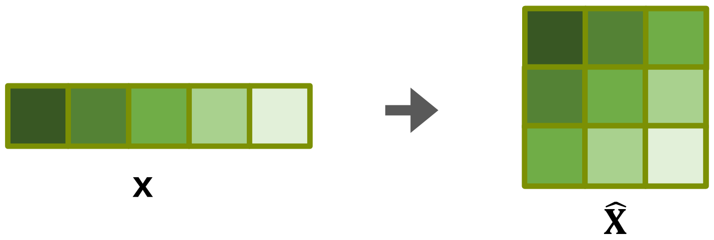
until convergence

Finalize: $\widehat{\boldsymbol{U}}_i \leftarrow$ top r_i left singular vectors of $\widehat{\mathcal{A}}_{(i)}$ with positive first elements, $1 \leq i \leq 3$

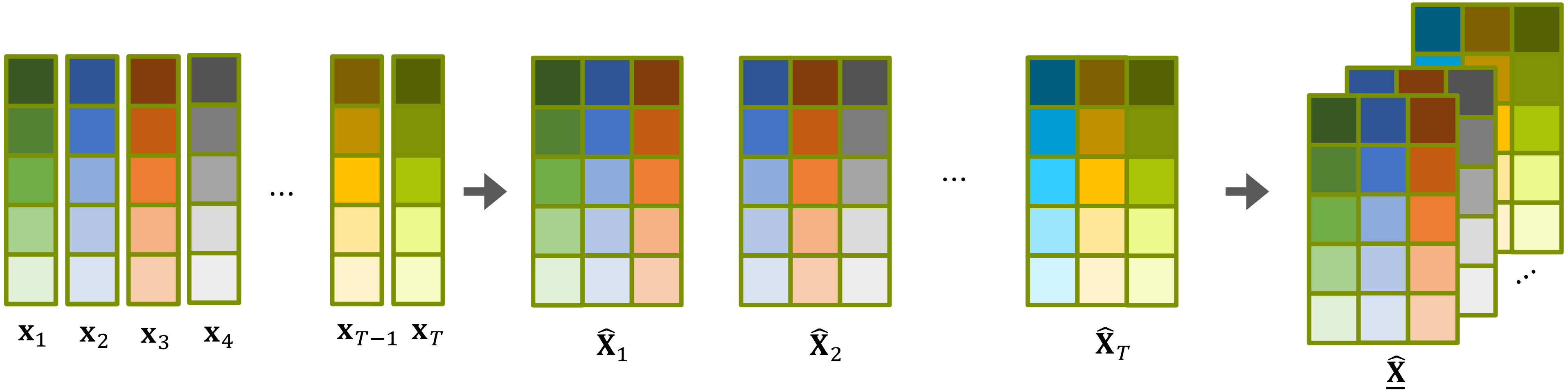
$$\widehat{\mathcal{G}} \leftarrow [\![\widehat{\mathcal{A}}; \widehat{\boldsymbol{U}}_1', \widehat{\boldsymbol{U}}_2', \widehat{\boldsymbol{U}}_3']\!]$$

Skoltech

Block Hankel Tensor-ARIMA



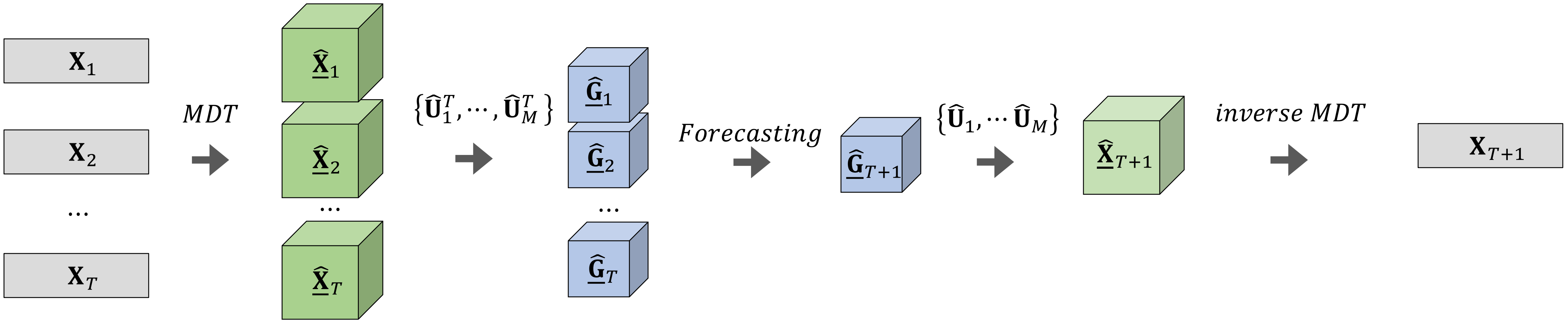
Hankelization of vector (Tensorization)



Multi-way delay embedding (extension of Hankelization)

Skoltech

Block Hankel Tensor-ARIMA



$$\Delta^d \hat{\underline{\mathbf{G}}}_t = \Delta^d \hat{\underline{\mathbf{X}}}_t \times_1 \hat{\mathbf{U}}_1^T \times_2 \hat{\mathbf{U}}_2^T \cdots \times_M \hat{\mathbf{U}}_M^T$$

$$\Delta^d \hat{\underline{\mathbf{G}}}_{T+1} = \sum_{i=1}^{p} \alpha_i \Delta^d \hat{\underline{\mathbf{G}}}_{T-i} - \sum_{i=1}^{q} \beta_i \hat{\epsilon}_{T-i}$$

Skoltech

Results of the Experiment. Yahoo.finance

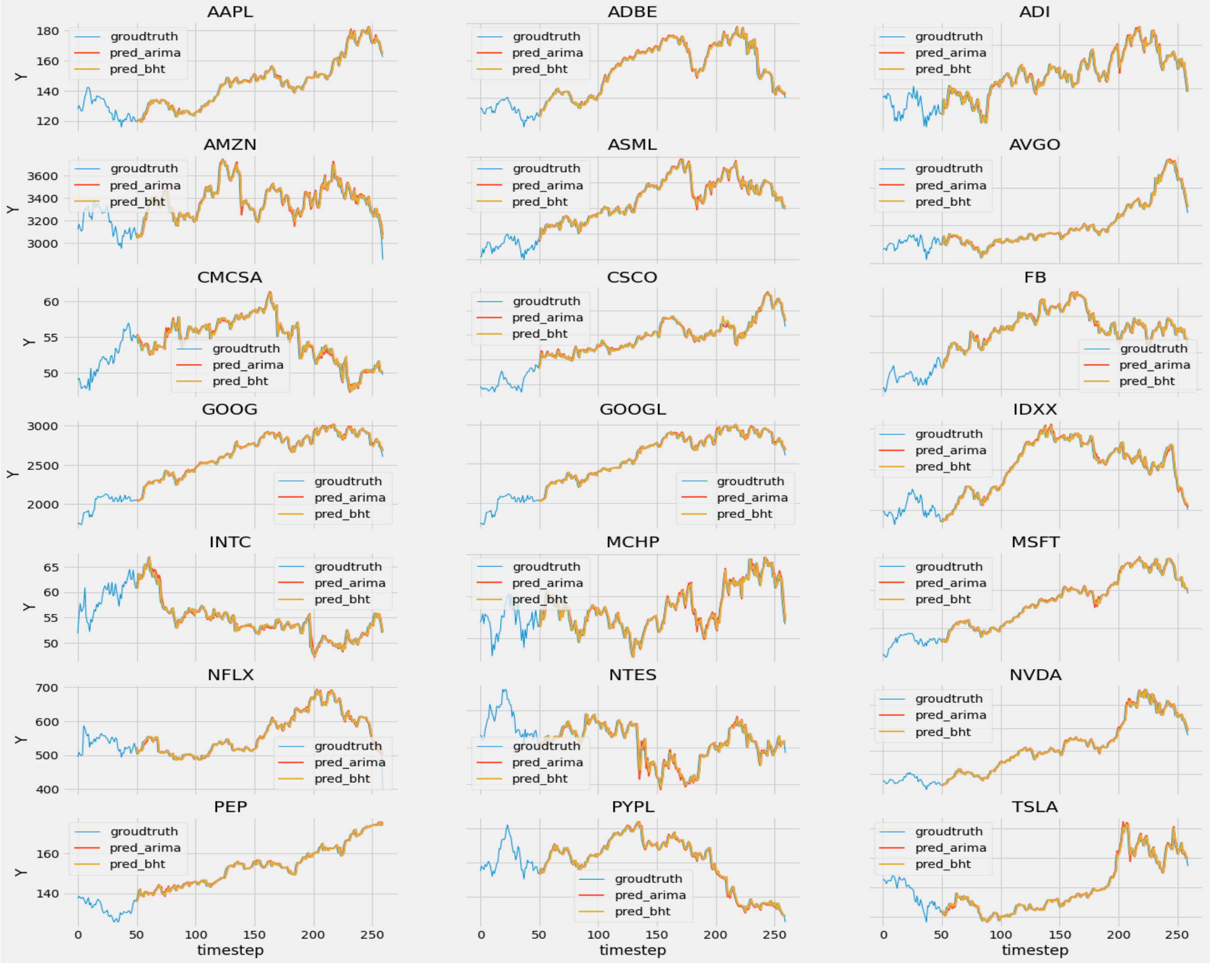


Fig. 1 BHT-ARIMA(1, 1, 0) with MDT-rank [21, 4], Tucker rank [4, 4] and window size - 50

	RMSE	Execute time
BHT-ARIMA	**18.24**	**192.15s**
ARIMA	20.26	442.71s

Table 1. Methods performance

Skoltech

Results of the Experiment. Air Quality Data Set

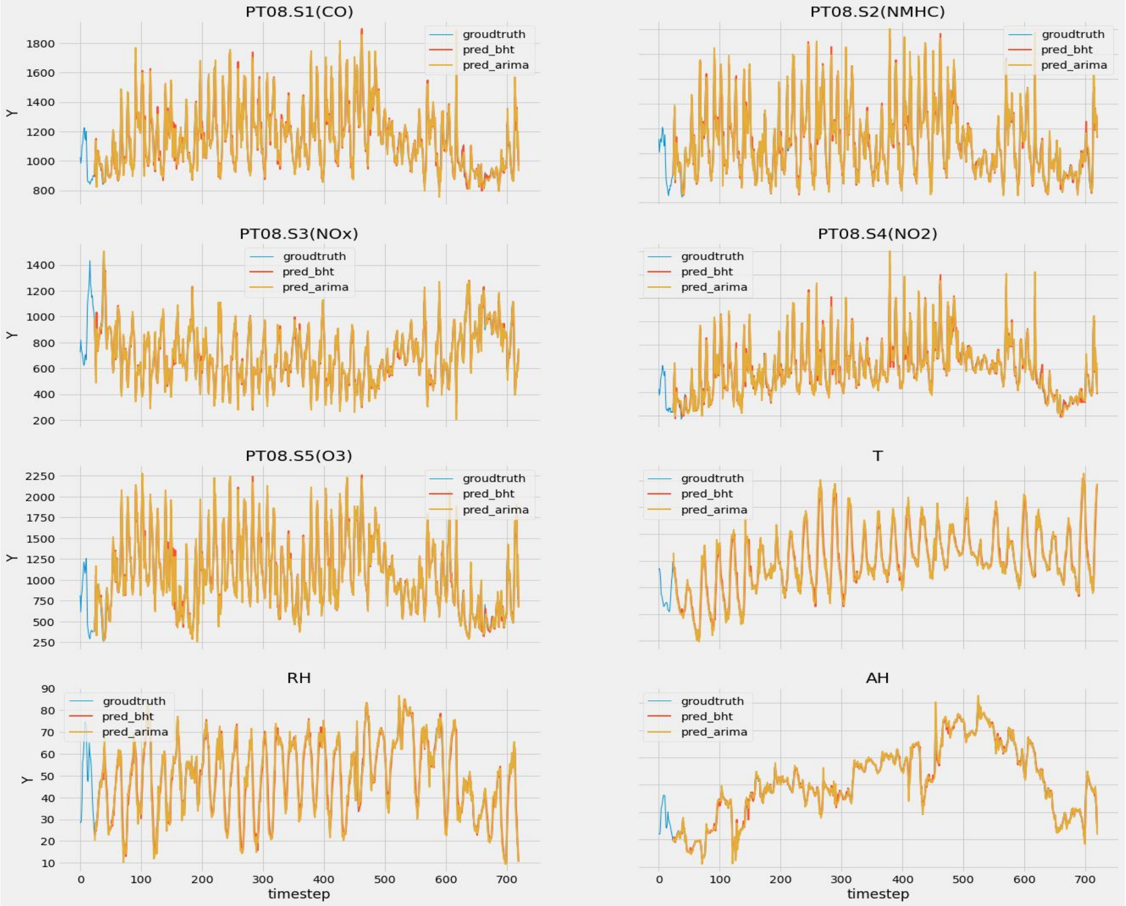


Fig. 2 BHT-ARIMA(2, 1,1) with MDT-rank [8, 4], Tucker rank [4, 4] and window size - 24

	RMSE	Execute time
BHT-ARIMA	**109.4**	**419.33s**
ARIMA	116.4	553.15s

Table 2. Methods performance

1. PT08.S1 (tin oxide) hourly averaged sensor response
2. PT08.S2 (titania) hourly averaged sensor response
3. PT08.S3 (tungsten oxide) hourly averaged sensor response
4. PT08.S4 (tungsten oxide) hourly averaged sensor response
5. PT08.S5 (indium oxide) hourly averaged sensor response
6. T Temperature
7. RH Relative Humidity (%)
8. AH Absolute Humidity

Skoltech

Results of the Experiment

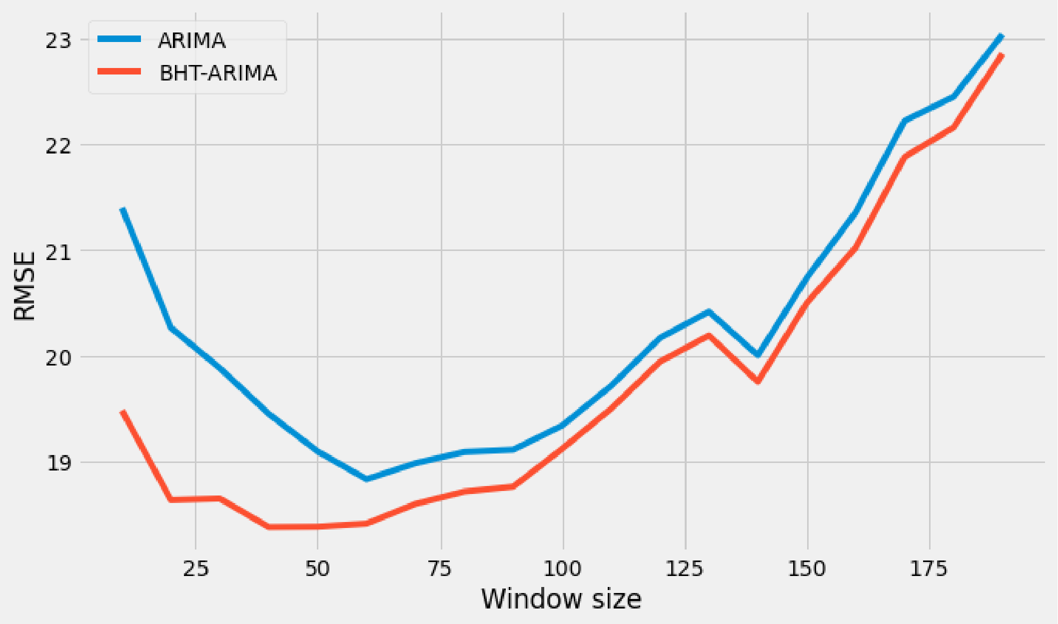


Fig. 3. Dependence of RMSE on the window size

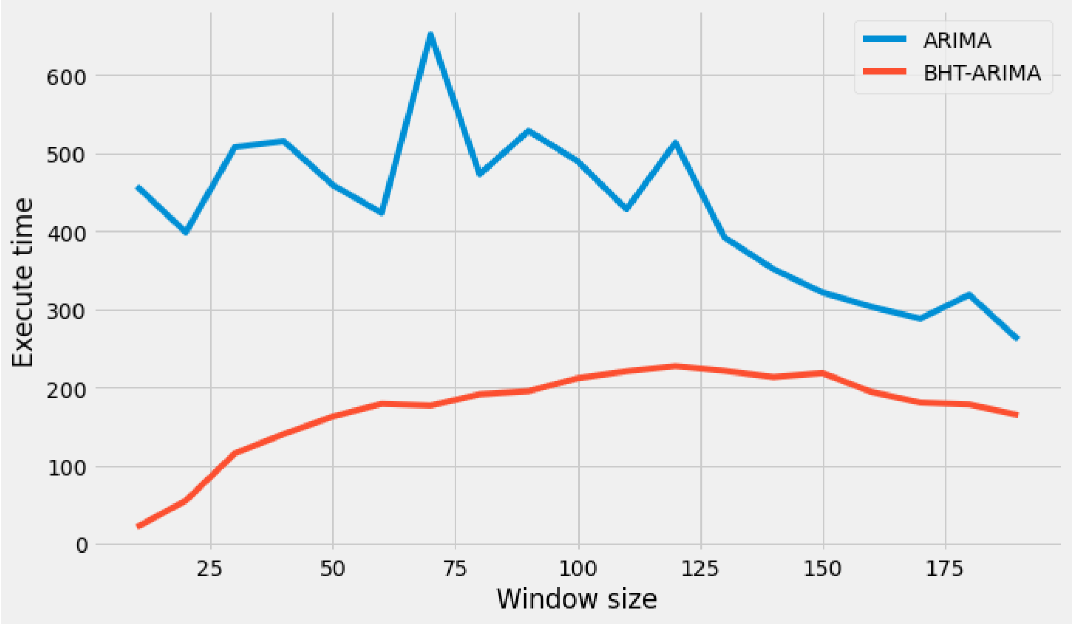


Fig. 4. Dependence of execute time on the window size

Skoltech

Results of the Experiment

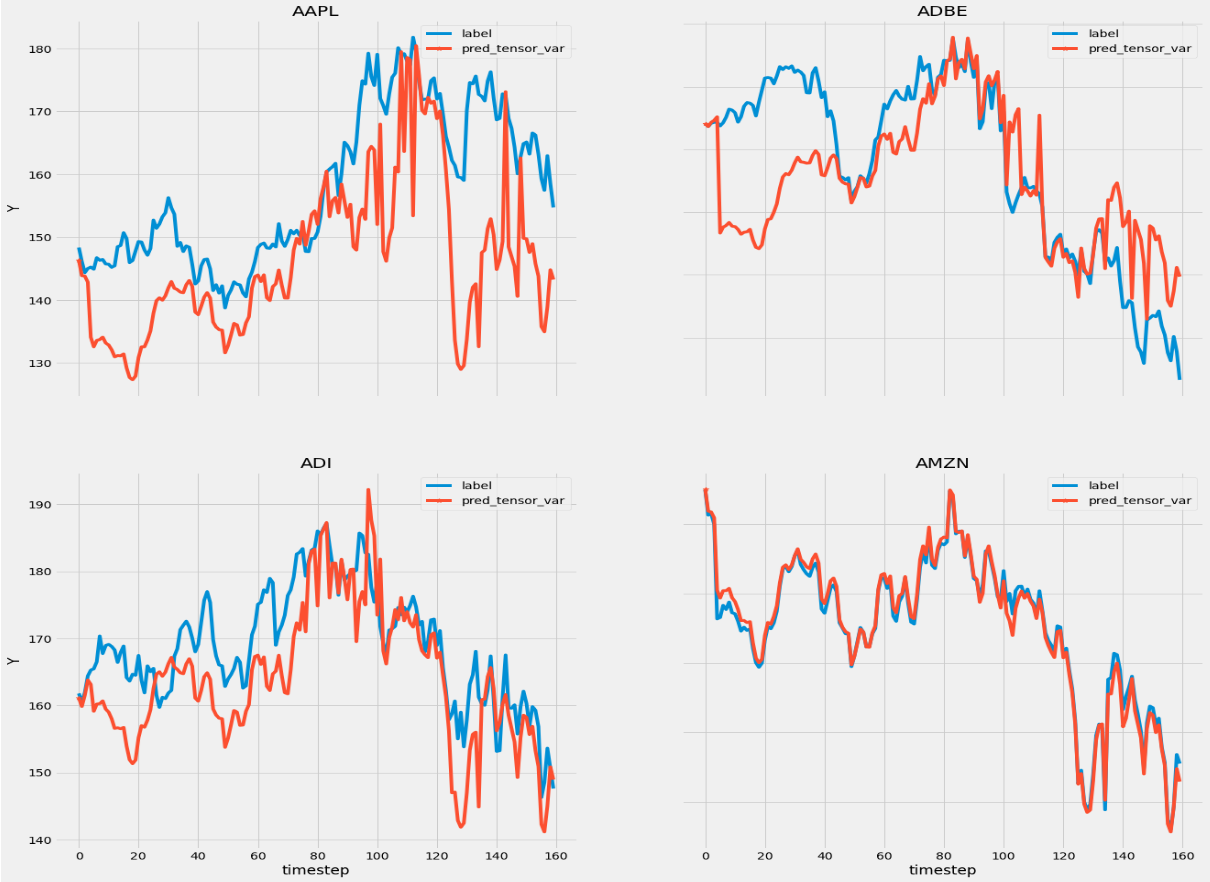


Fig. 5 Tensor-VAR(1) with train size - 100

	RMSE
Tensor-VAR	29.86

Table 3. Method performance

Skoltech

Thank you for attention!

Skoltech