



South Ural  
State University

National Research  
University

# Monitoring and forecasting crop yields

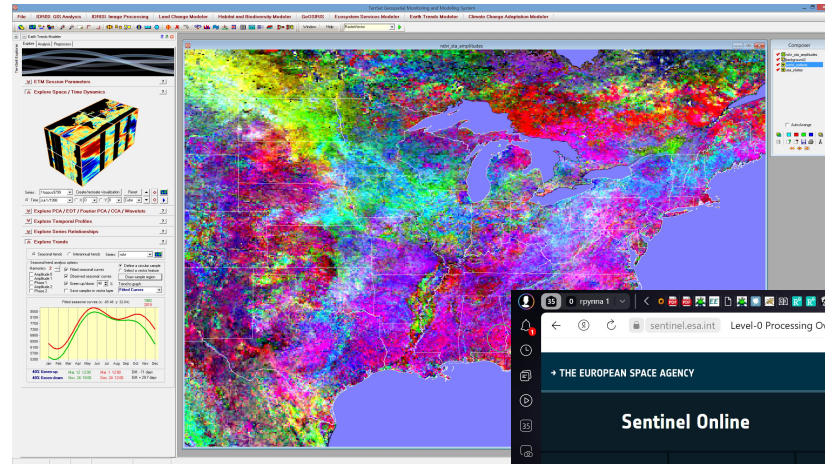
## Team:

- **Makarovskikh T.A. (presenter)**
- **Abotaleb M.S.A.**
- **Panyukov A.V.**

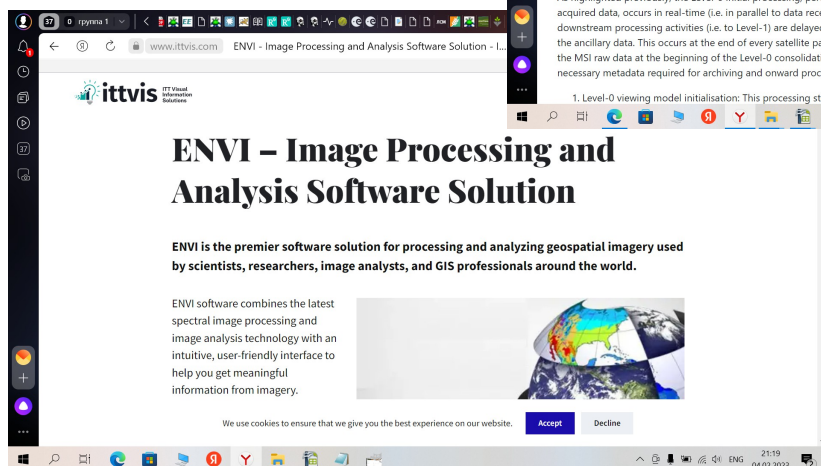
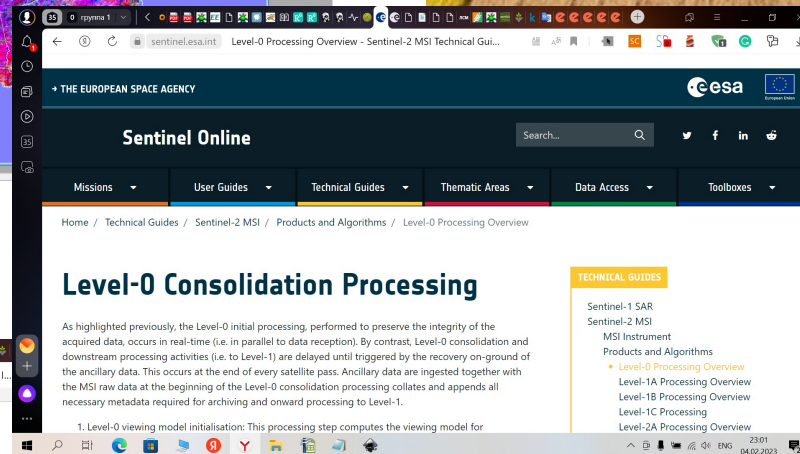
St.Peterburg, 30.03.2023



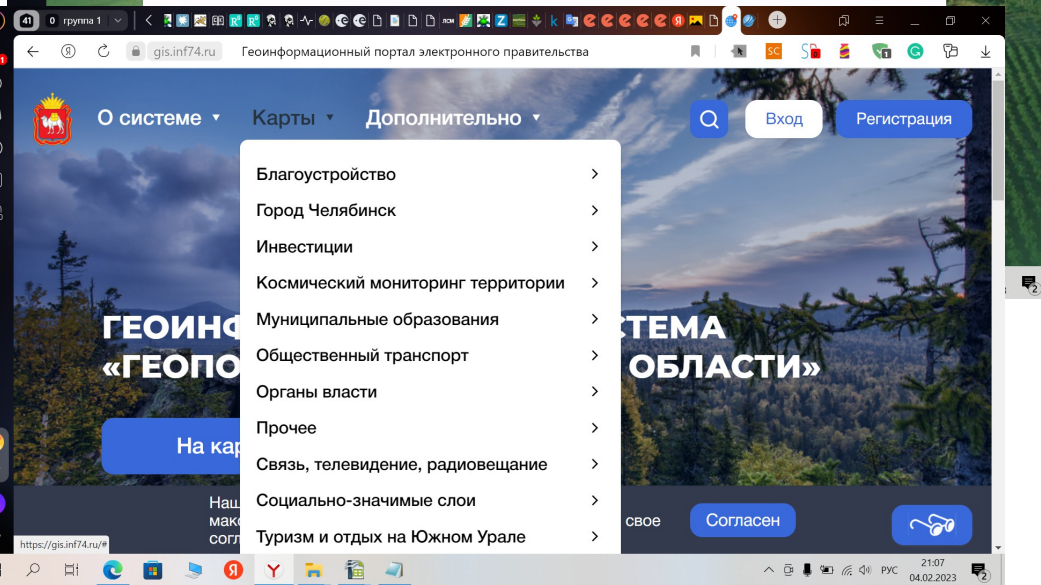
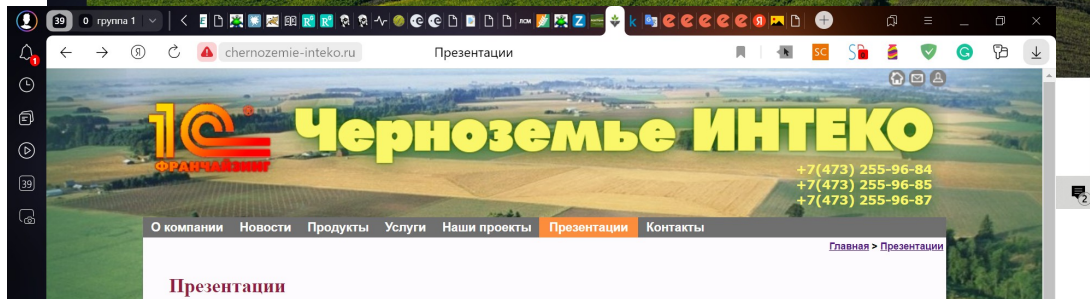
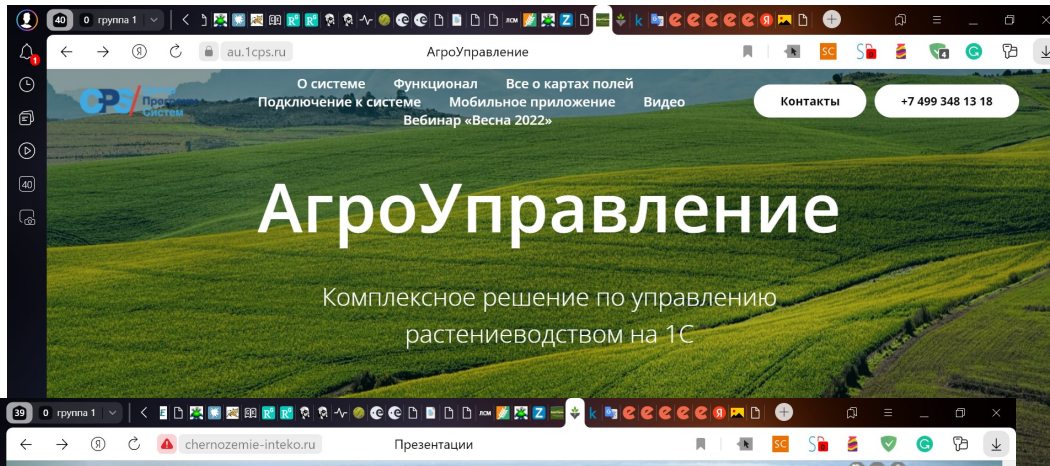
# Precision farming today



OneSoil



# Precision farming in Russia



Yakushev, V., Yakushev, V.: Mathematical models and methods of realizing information technology procedures in precision agriculture 34(4), 280–283

Tokarev, K.E., Lebed, N.I., Kuzmin, V.A., Chernyavsky, A.N.: Theory and technologies of irrigation control for crops based on information technologies decision support and mathematical modelling 60(4), 1–16.

Terekhin, E.A.: Analysis of the seasonal dynamics of NDVI index and the reective properties of corn in the Belgorod Region. Modern problems of remote sensing of the Earth from space 11(4), 244253 (2014).

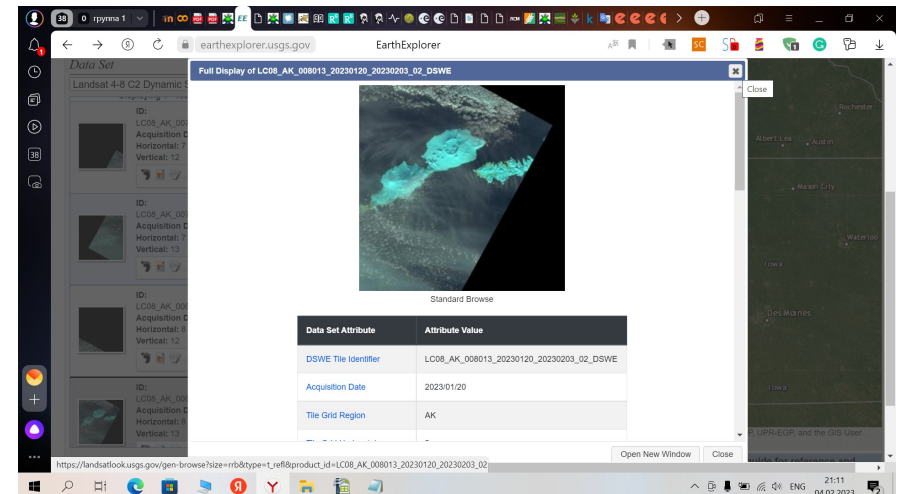
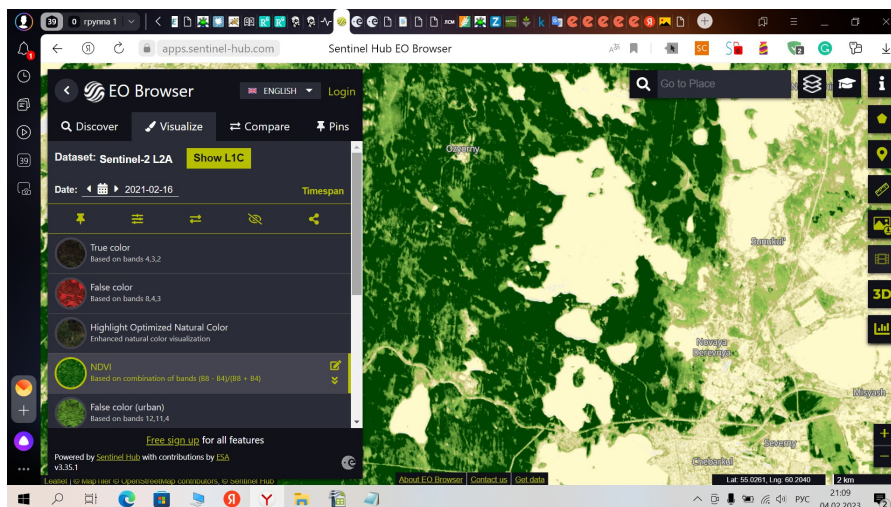
Bukhovets, A. G., Semin, E.A., Kostenko, E.I., Yablonovskaya, S.I. Modelling of the dynamics of the NDVI vegetation index of winter wheat under the conditions of the CFD. Bulletin of the Voronezh State Agrarian University. 2. 186199 (2018).

Greben, A.S., Krasovskaya, I.G. Analysis of the main methods for forecasting yields using space monitoring data, in relation to grain crops in the steppe zone of Ukraine. Radio electronic and computer systems. 2 (54). 170180 (2012).

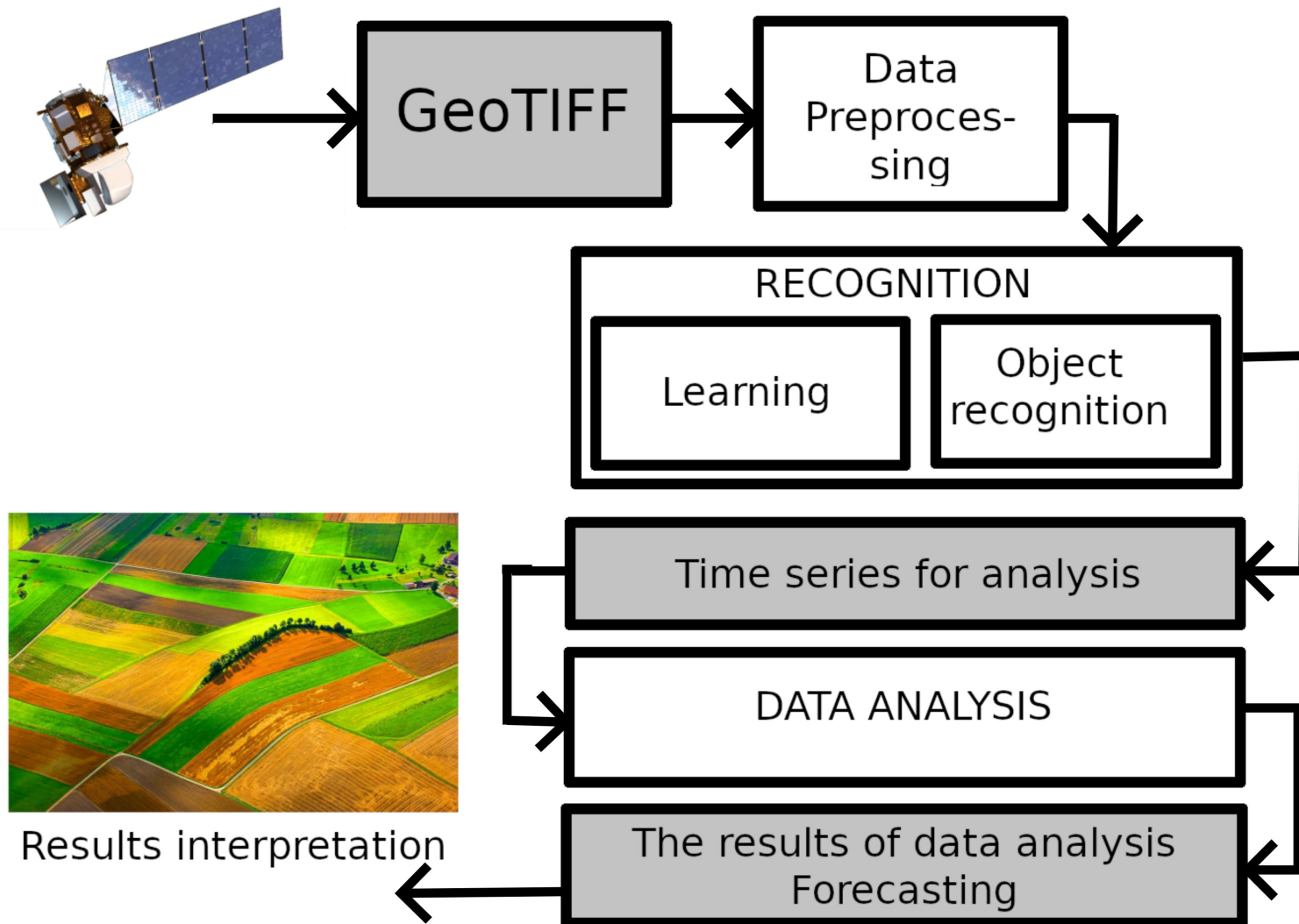
Spivak, L.F., Vitkovskaya, I.S., Batyrbayeva, M.Zh., Kauazov, A.M. Analysis of the results of forecasting the yield of spring wheat based on time series of statistical data and integral indices of vegetation. Modern problems of remote sensing of the Earth from space. 12 (2). 173182 (2015)

# Datasets

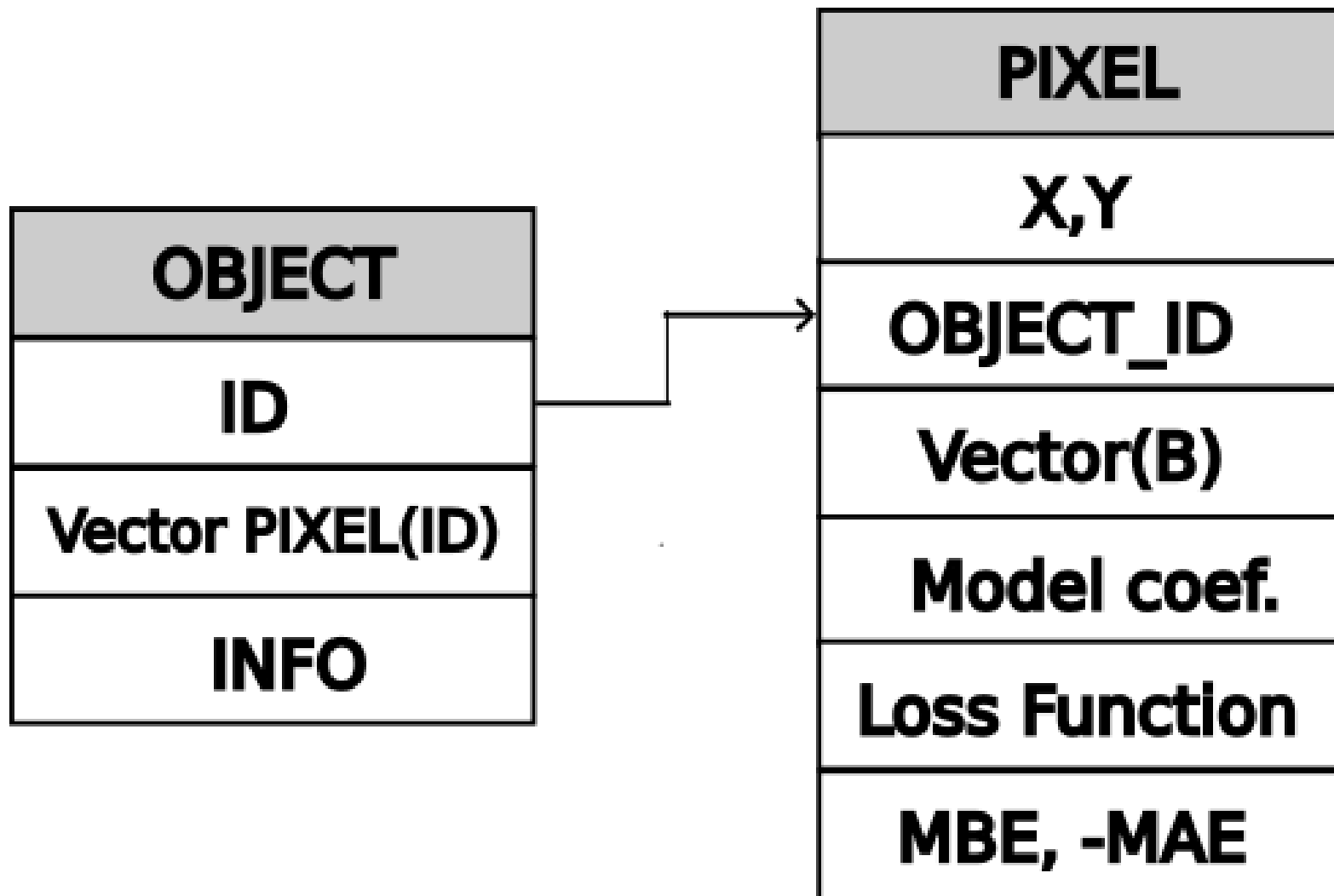
- <https://apps.sentinel-hub.com>
- <https://earthexplorer.usgs.gov/>
- <https://www.kaggle.com>



# The system for intelligent photomonitoring



# The database of objects



# Data preprocessing algorithm for one classified object

- **Input:**

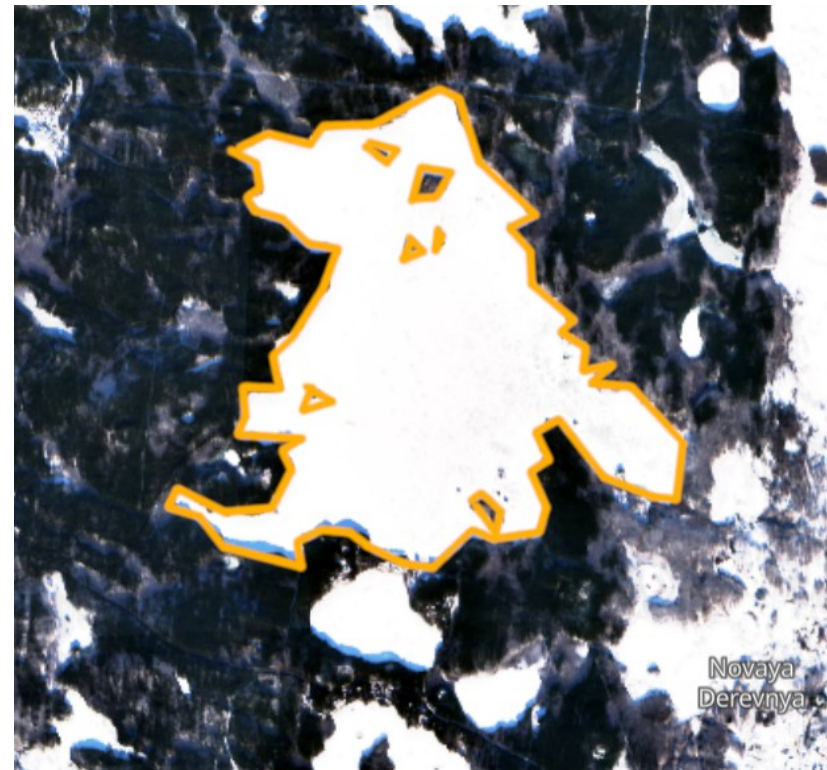
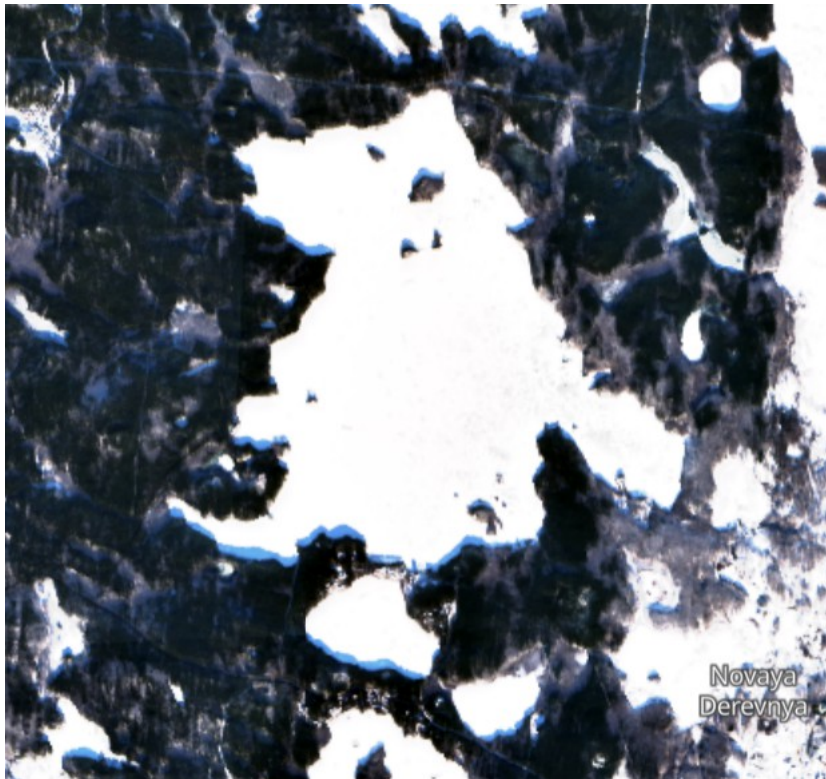
- **M** sets of files in \*.tiff format (**K** bands for each set) for time period
- Each file has **N**= $N_1 \times N_2$  tiles.

- **Output:**

- $M \times N \times K$  matrix, each row corresponding to a tile, columns containing values for each range).
  - **M** — the number of observation points in time
  - **N** — the number of tiles
  - **K** — the number of bands

# Step 1. Load the data

- Read GeoTIFF classified data and digital number (DN) values as separate arrays



## Step 2. Data restructuring

- Save the resulting arrays as a list of tiles.

This is the data organization used by most machine learning and analytic algorithms.

<b>P<sub>1</sub></b>	<b>P<sub>2</sub></b>	<b>P<sub>3</sub></b>		<b>P<sub>4</sub></b>
	<b>P<sub>5</sub></b>	<b>P<sub>6</sub></b>		<b>P<sub>7</sub></b>
<b>P<sub>8</sub></b>	<b>P<sub>9</sub></b>	<b>P<sub>10</sub></b>		<b>P<sub>11</sub></b>
	<b>P<sub>12</sub></b>	<b>P<sub>13</sub></b>	<b>P<sub>14</sub></b>	<b>P<sub>15</sub></b>

# Step 3. Transformation

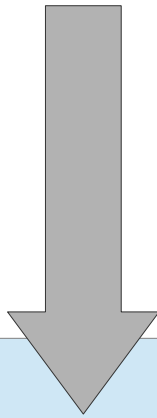
- Transformation of datasets so that the model perceives each row as a separate tile (separate training object).

$P_1$	$B_{11}$	$B_{12}$	...	$B_{1k}$
$P_2$	$B_{21}$	$B_{22}$	...	$B_{2k}$
...	...	...	...	...
$P_n$	$B_{n1}$	$B_{n2}$	...	$B_{nk}$

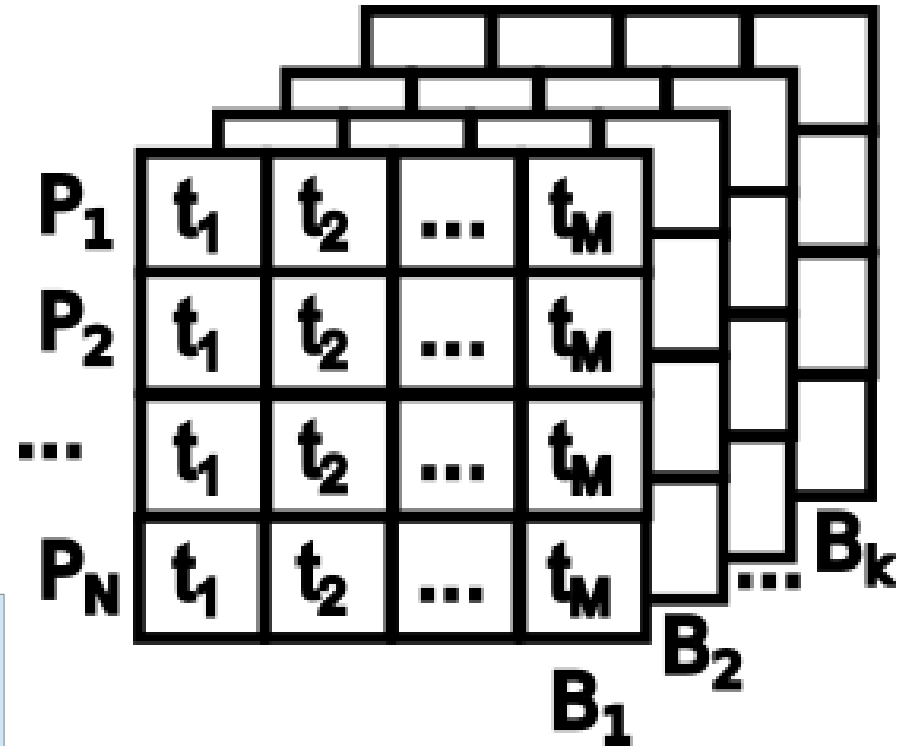
Band	mkm	Resolution (m)
1 - Blue	0.45-0.52	30
2 - Green	0.52-0.60	30
3 - Red	0.63-0.69	30
4 - Near infrared	0.76-0.90	30
5 - Middle infrared	1.55-1.75	30
6 - Far infrared	2.08-2.35	120
7 - Thermic	10.4-12.4	30

# Forming time series

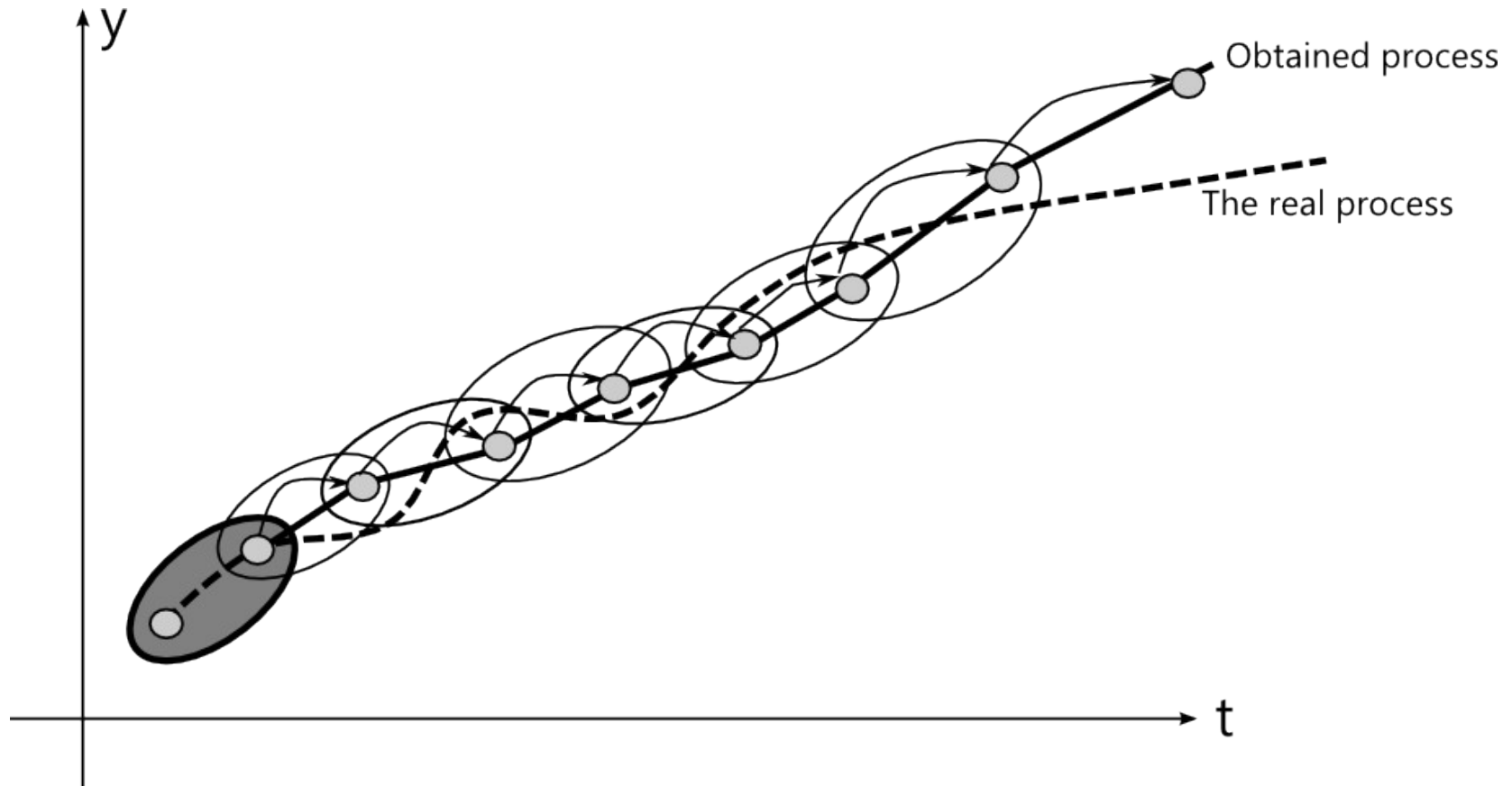
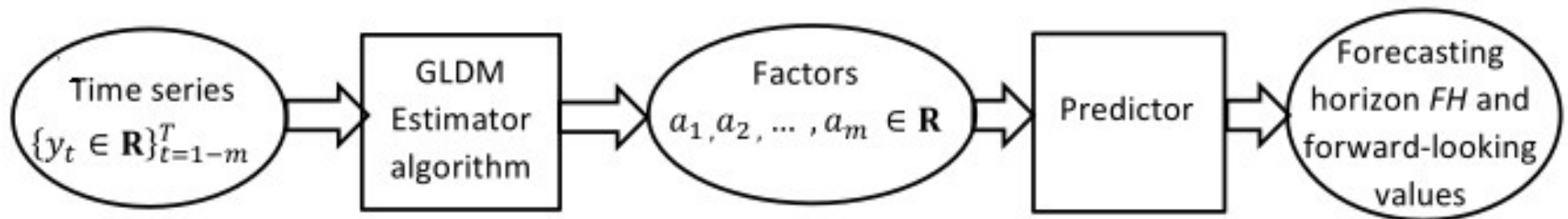
- Run steps 1-3 for all figures of interest



**The set of time series  
for analysing  
and modelling  
for each band**



# General Least Deviation Method



# GLDM Estimator

$$y_t = \sum_{j=1}^{n(m)} a_j g_j(\{y_{t-k}\}_{k=1}^m) + \varepsilon_t, \quad t = 1, 2, \dots, T$$

- $y_i$  — the values of time series
- $g_j$  — defined functions (5 for the model of 2<sup>nd</sup> order, and 9 for the model of 3<sup>rd</sup> order)
- $\varepsilon_t$  — unknown error

## Aim:

- Define coefficients of model  $a_j$

# Functions $g(^*)$

$$g_{(k)}(\{y_{t-k}\}_{k=1}^m) = y_{t-k},$$

$$g_{(kl)}(\{y_{t-k}\}_{k=1}^m) = y_{t-k} \cdot y_{t-l},$$

$$k = 1, 2, \dots, m; \quad l = k, k + 1, \dots, m.$$

$$n(m) = 2m + C_m^2 = m(m + 3)/2$$

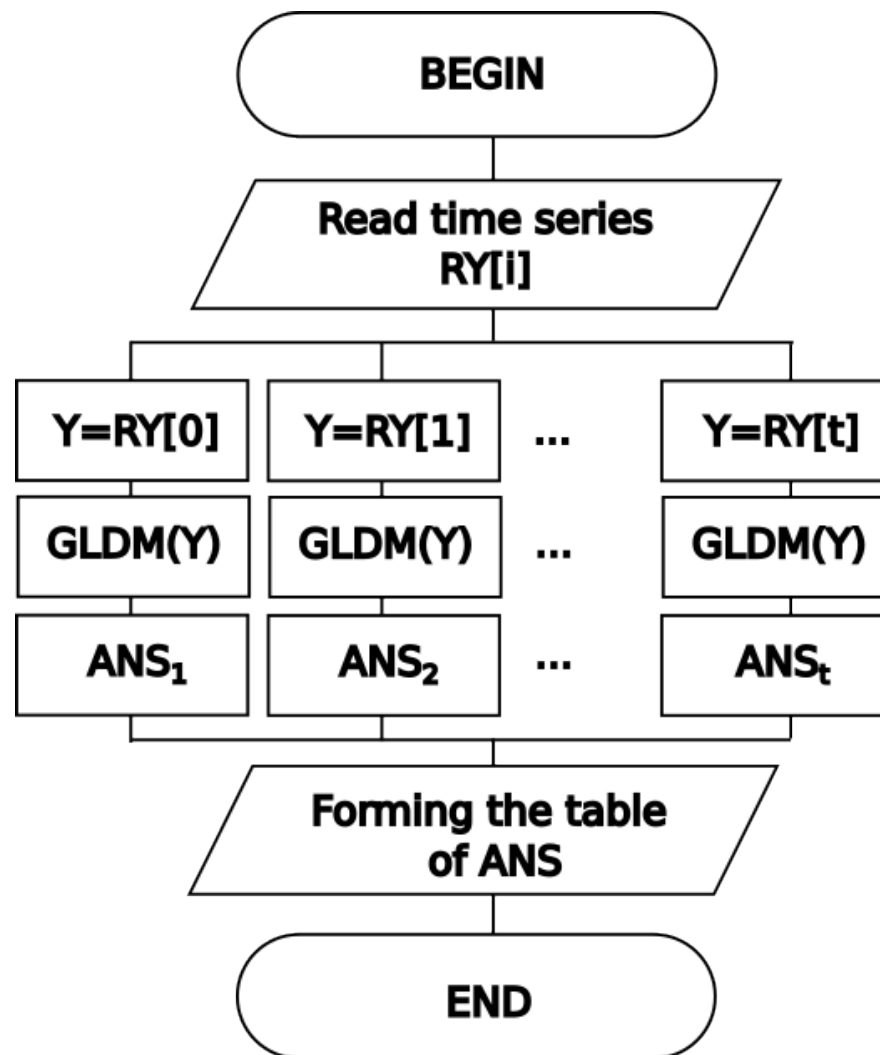
# Optimization task

$$\sum_{t=1}^T \arctan \left| \sum_{j=1}^{n(m)} a_j g_j(\{y_{t-k}\}_{k=1}^m) - y_t \right| \rightarrow \min_{\{a_j\}_{j=1}^{n(m)} \subset \mathbb{R}}$$

# Comparing GLDM and machine learning models

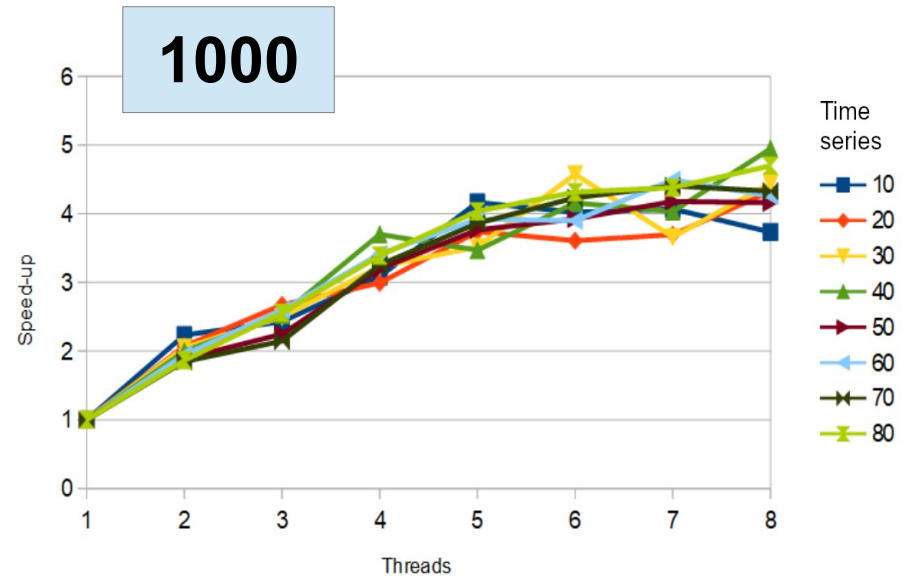
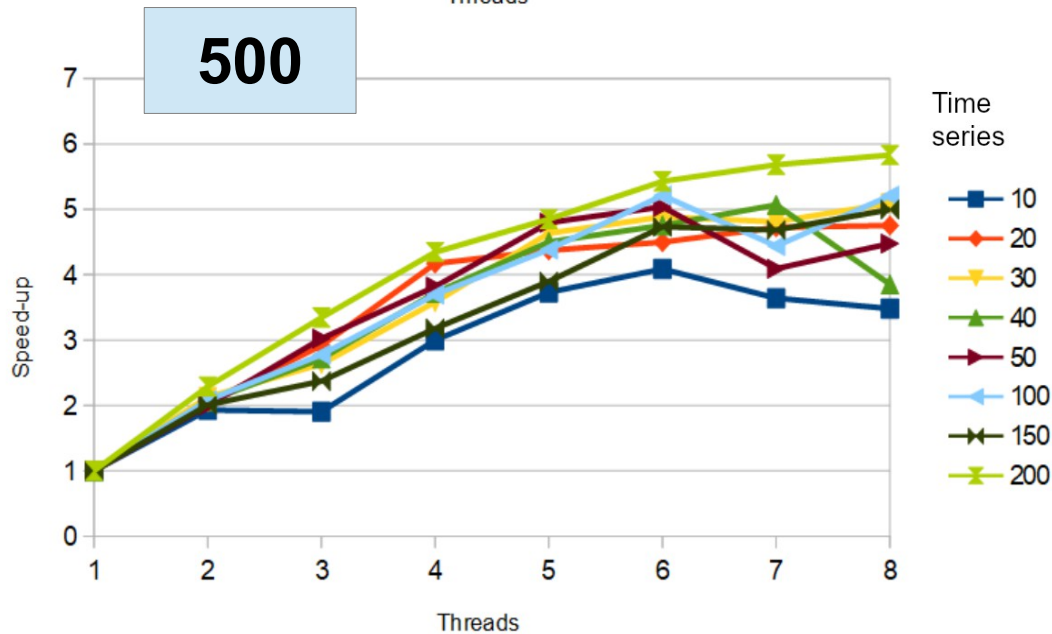
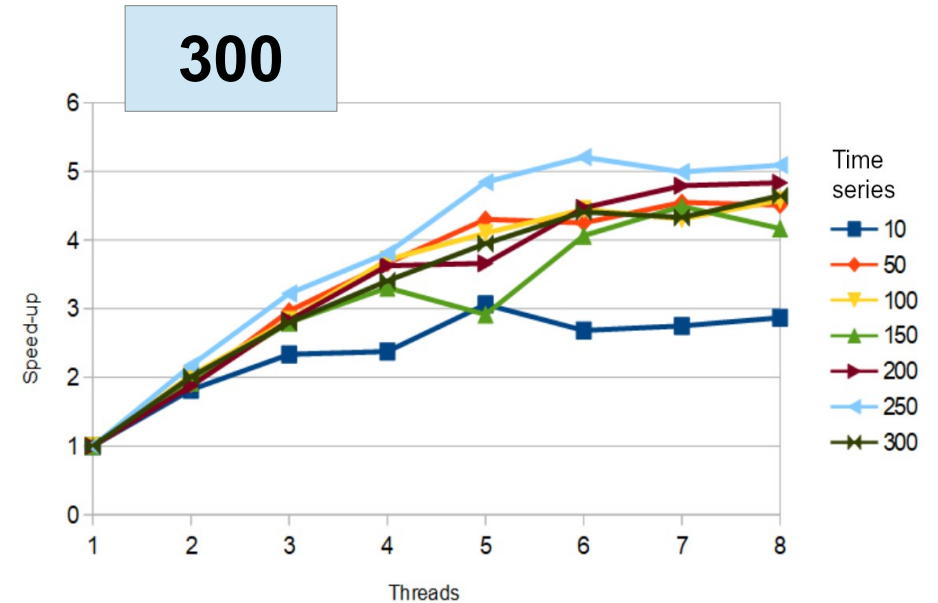
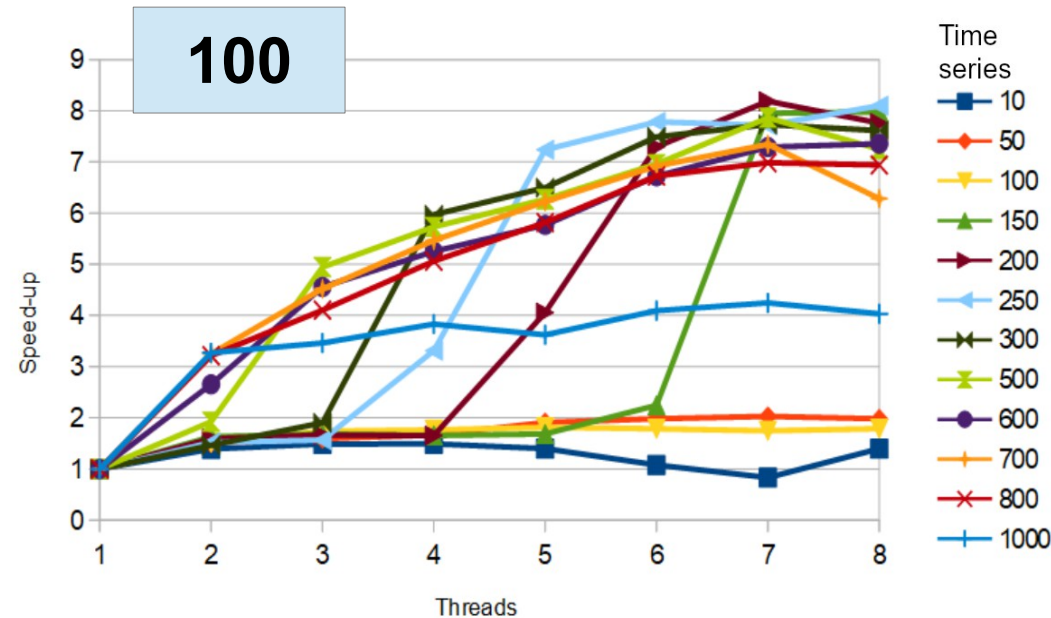
Model	MSE	RMSE	MAE	$R^2$	RRMSE	Correlation	MBE	Loss Func.
LSTM	8886.12	94.27	26.86	0.98	0.25	0.99	17.62	15571.76
LSTMs	21761.75	147.52	52.03	0.96	0.38	0.99	46.3	40929.42
BDLSTM	8703.41	93.29	28.56	0.98	0.24	0.99	0.67	590.96
GRU	15637.84	125.05	30.9	0.97	0.33	0.99	16.94	14978.54
GLDM	34.04	5.83	0.65	1	0.01	1	-0.65	165950.5

# Parallel GLDM



# Speed-up for PC

11th Gen Intel(R) Core(TM) i5-1135G7, 2.40GHz, 2.42 GHz, 16 Gb of RAM,  
8 threads



# Applications of GLDM

- Forecasting using a dynamic model of the normalized difference vegetation index (NDVI)
  - Crop yields
  - Waterlogging

# ToDo List

- To held the experiments and obtain own datasets for the fields in Chelyabinsk region
- To classify the objects using own datasets
- To obtain models for NDVI forecasting (crops forecasting)
- To solve the task of devining UAV optimal route
- Debugging parallel programs with use of supercomputer
- The task of restoring the missing data

# ToDo List

- To held the experiments and obtain datasets for the fields of Chelyabinsk region
- To develop the data clusterization algorithm to decrease the number of time series with the same properties
- Using hybrid OpenMP+MPI approach
- Optimization of RAM memory use
- Debugging parallel programs with use of supercomputer
- The task of restoring the missing data

**Thank you for attention!**

