Машинное обучение в прикладных и научных задачах, решаемых в лаборатории информационных технологий им. М.Г. Мещерякова

Александр Ужинский
auzhinskiy@jinr.ru

Industry 4.0

The term “Industry 4.0” is used to signify the beginning of the fourth industrial revolution – the previous three being mechanical production, mass production, and then the digital revolution. It could be argued that Industry 4.0 is simply an amalgamation of the three previous eras in manufacturing, but Industry 4.0 is poised to be much more impactful than that.

Industrial Internet of Things (IIoT), Automation, Artificial Intelligence, Big Data & Analytics, The Cloud, Cybersecurity, Simulations, Robotics, Smart manufacture, Mobile devices, Smart manufacture, etc.

- экологический мониторинг с использованием данных дистанционного зондирования земли

- использование современных средств автоматизации и машинного обучения для нужд сельского хозяйства

- прочее

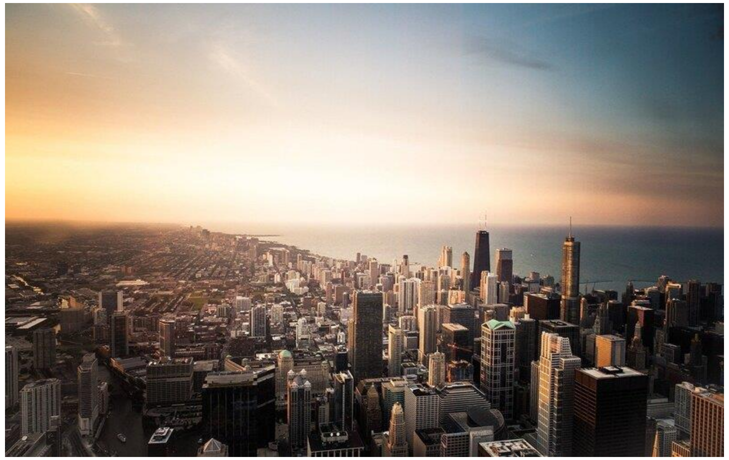
Environmental Concerns

The environmental problems like global warming, acid rain, air pollution, urban sprawl, waste disposal, ozone layer depletion, water pollution, climate change and many more affect every human, animal, and nation on this planet.

Over the last few decades, the exploitation of our planet and the degradation of our environment has gone up at an alarming rate. As our actions have been not in favor of protecting this planet, we have seen natural disasters striking us more often in the form of flash floods, earthquakes, blizzards, tsunamis, and cyclones.





Air pollution

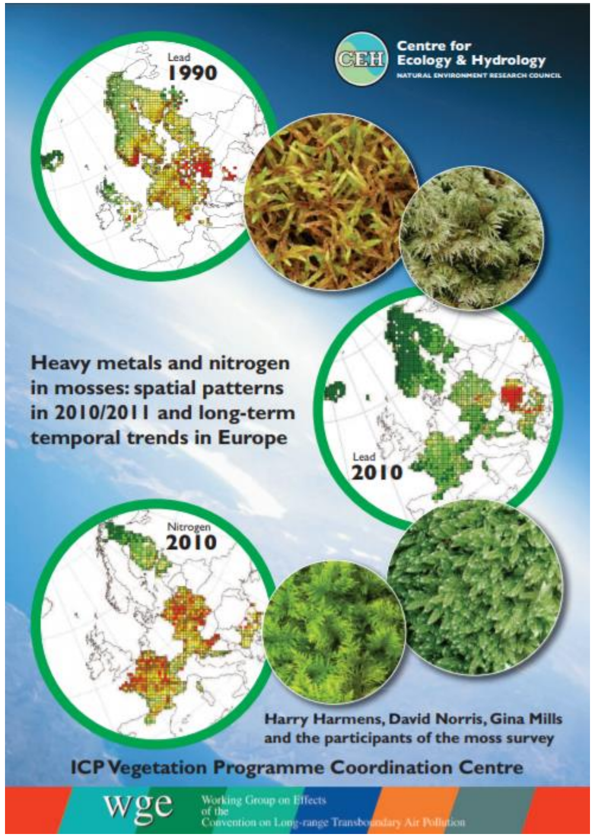
Air pollution has a significant **negative impact** on the various components of ecosystems, **human health**, and ultimately, causes significant **economic damage**.

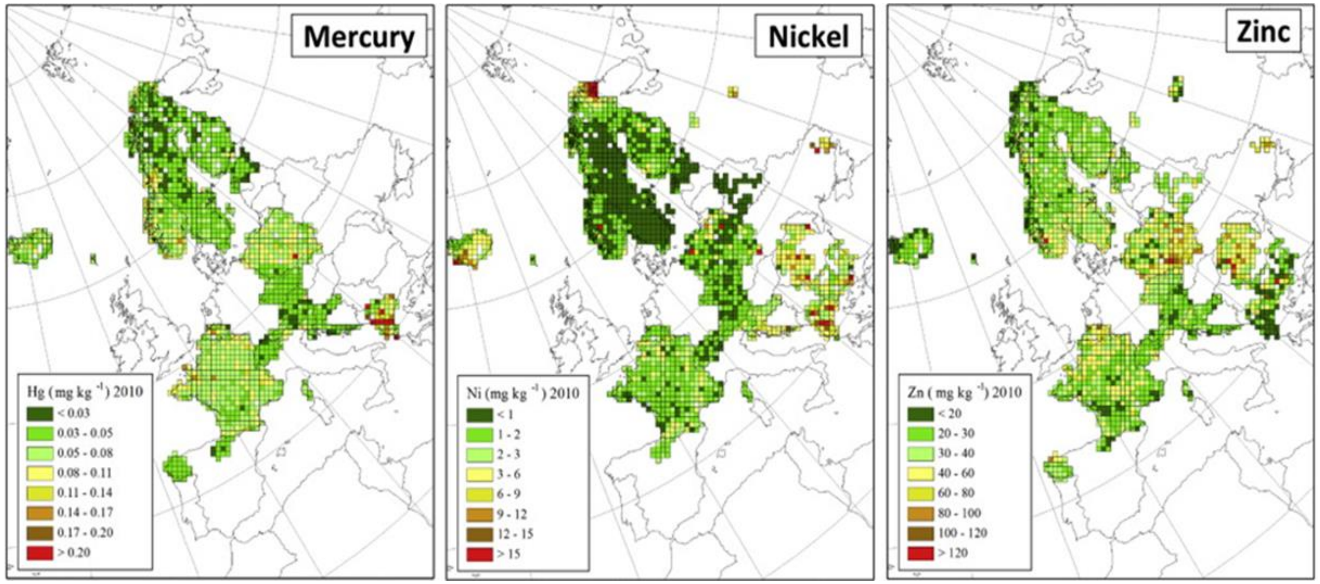
More than nine out of 10 of the world's population – 92% – lives in places where **air pollution exceeds safe limits**, according to research from the World Health Organization (WHO).

There are regional and international **environment control programs**. They use different techniques and tools but as a result, they all want to understand **what is the current situation** and how it will evolve.

ICP Vegetation

The aim of the **UNECE International Cooperative Program (ICP)** Vegetation in the framework of the United Nations Convention on Long-Range Transboundary Air Pollution is to **identify the main polluted areas of Europe**, produce regional maps and further develop the understanding of the long-range transboundary pollution. Atmospheric deposition study of heavy metals, nitrogen, persistent organic compounds (POPs) and radionuclides is based on the analysis of naturally growing mosses through moss surveys carried out **every 5 years**. The program is realized in **43 countries of Europe and Asia**. Mosses are collected at thousands of sites

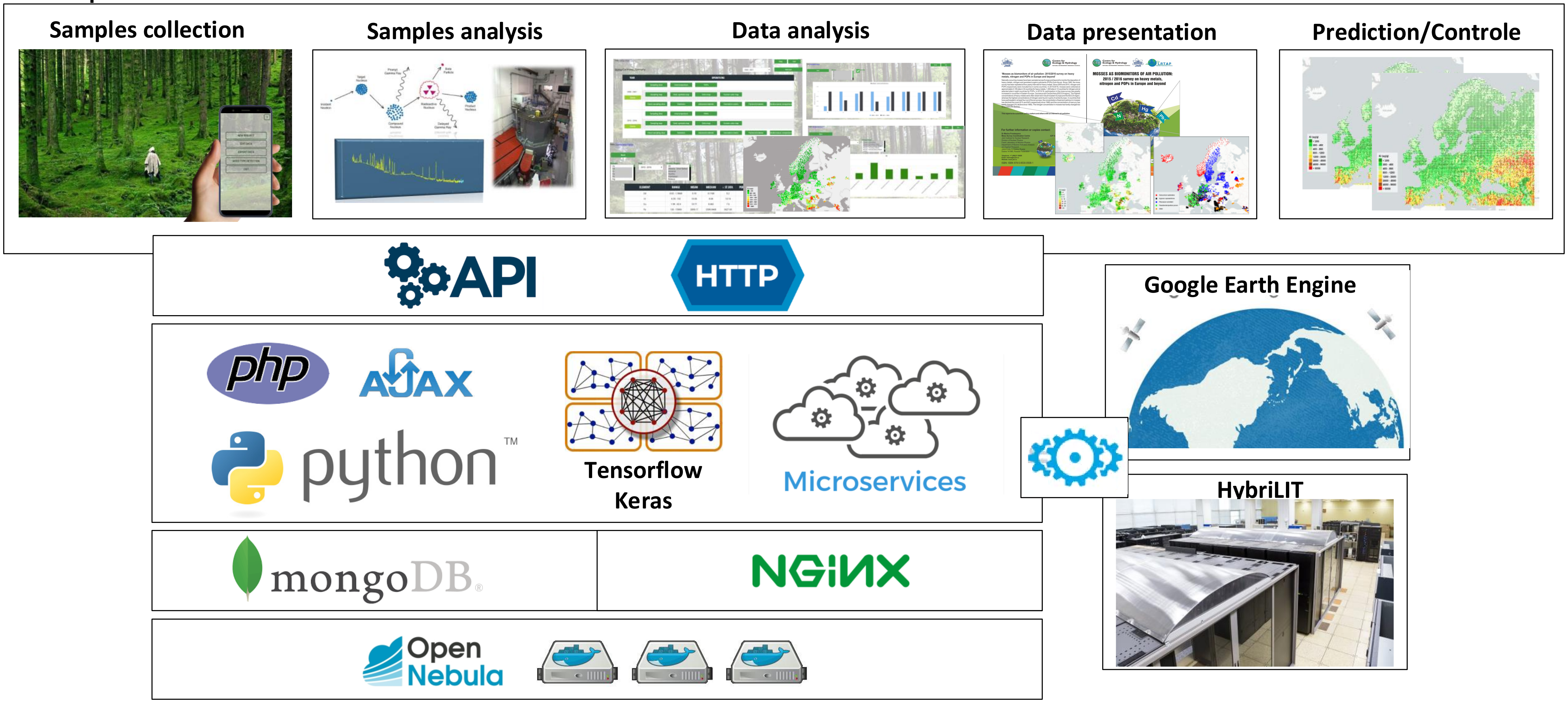




Examples of distribution maps in Atlas 2010

Since 2014 the JINR Frank Laboratory of Neutron Physics sector of neutron activation analysis is the **coordinator of the ICP** Vegetation program

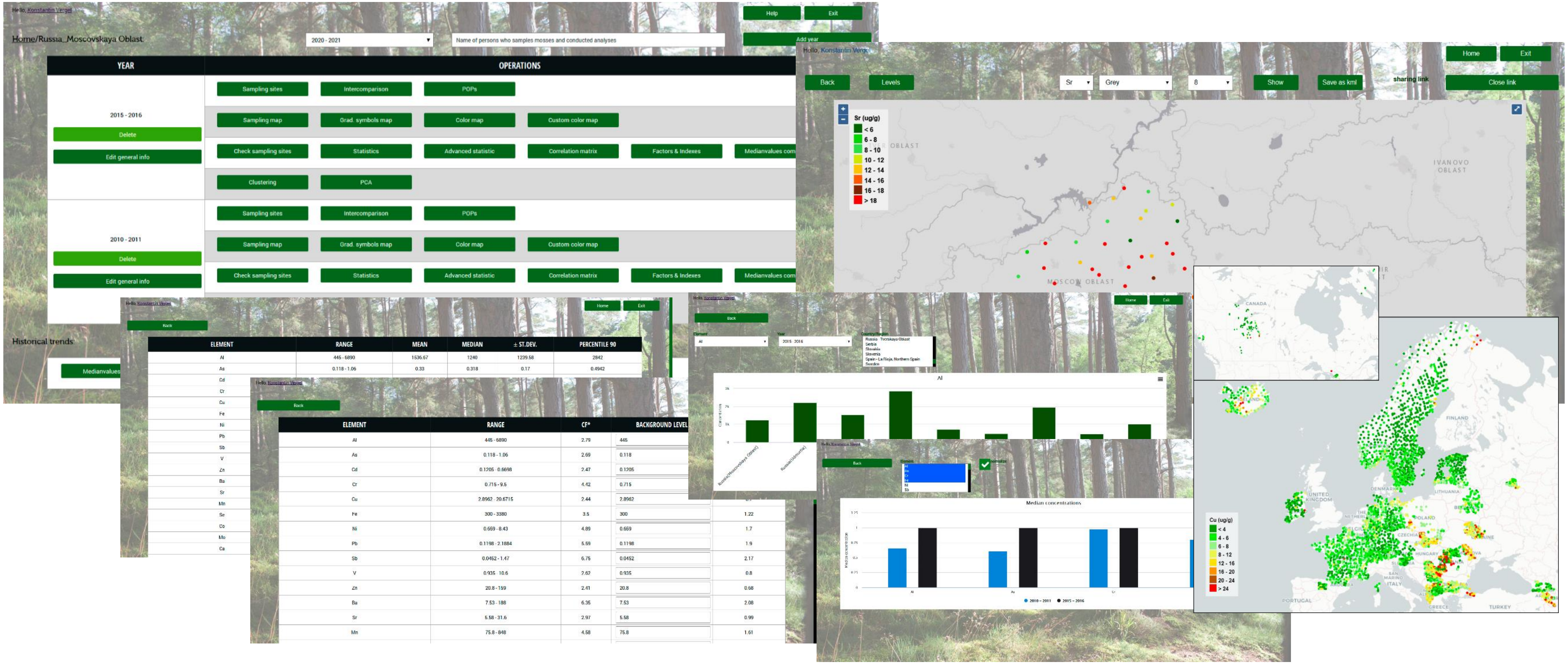
The platform



Since the launch of the first version of the platform, a mobile application has been developed to simplify the process of collecting and verifying data, deep learning models for image classification and pollution prediction based on remote sensing data, various functional blocks implemented in a microservice architecture to automate a number of operational tasks, and the analytical capabilities of the system are also expanded.

DMS

The Data Management System (DMS) of the UNECE ICP Vegetation was developed at the Laboratory of Information Technologies and consists of a set of interconnected services and tools deployed and hosted at the Joint Institute for Nuclear Research (JINR) cloud infrastructure. DMS is intended to provide its participants with a **modern unified system of collecting, analyzing and processing of biological monitoring data.**

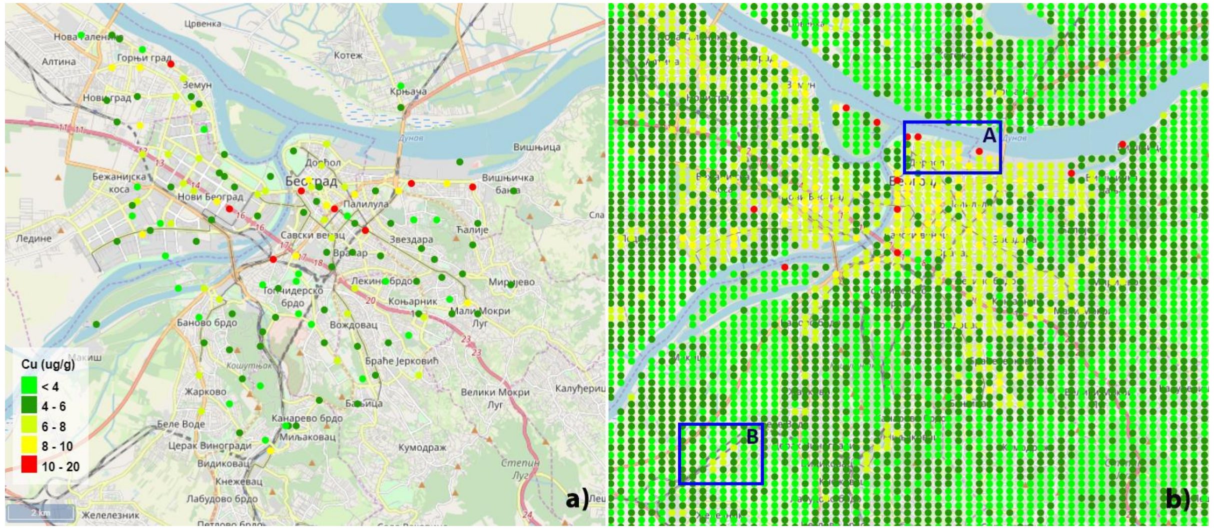


DMS. Atlas 2015-2016

The project, designed to automate typical operations with data and the creation of regional maps, absorbed more and more new approaches and technologies and now it may be classified as intelligent platforms.

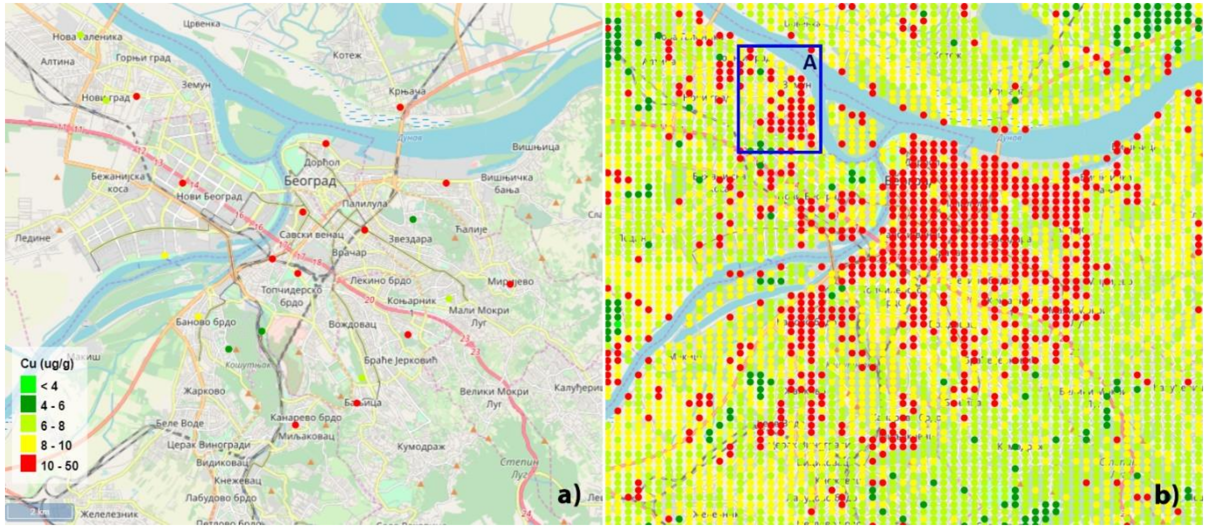
Modeling - Motivation & Benefits

Modelling of air pollution can be a good option for overcoming gaps in the data gathering.

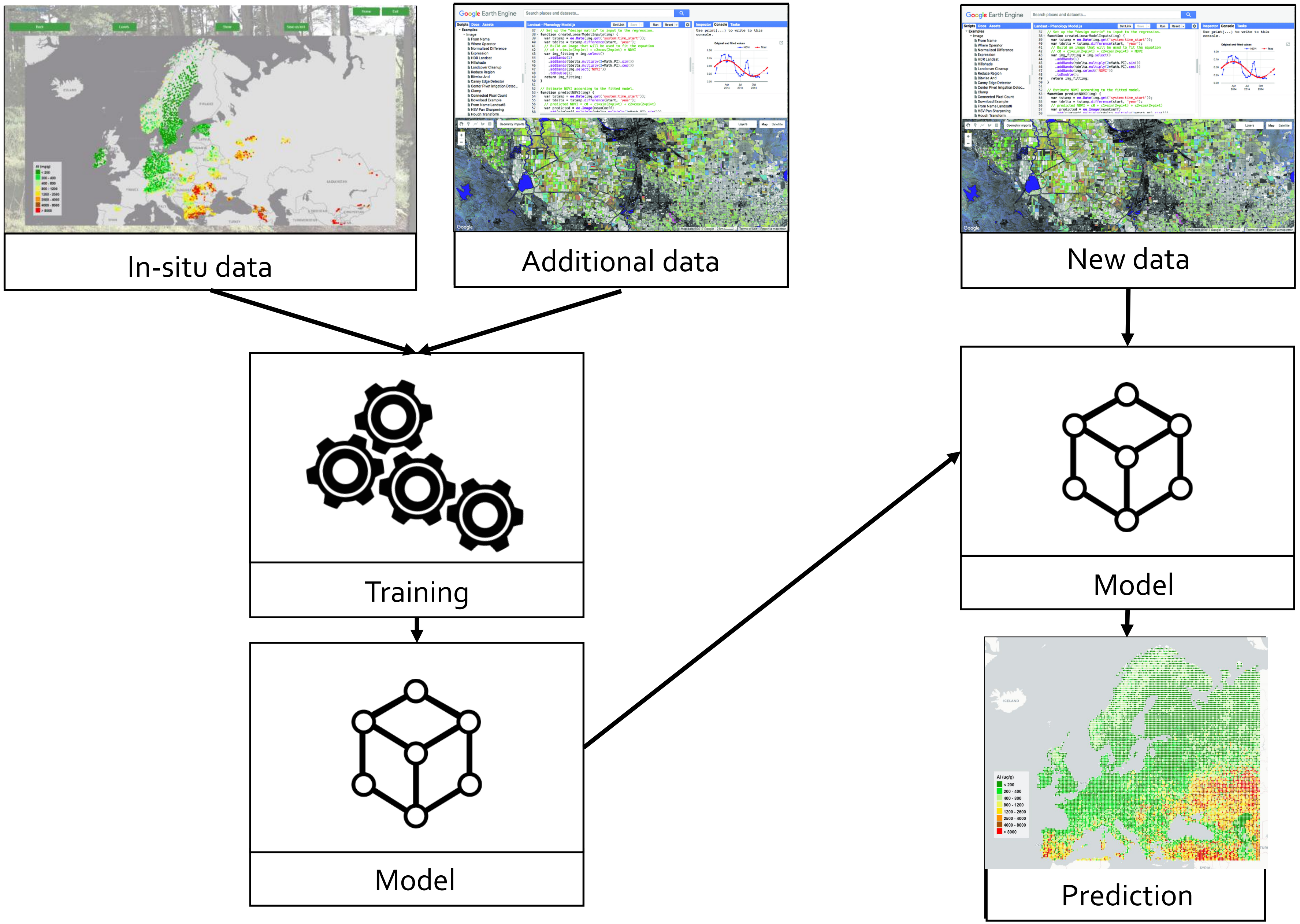


Modelling allows us to:

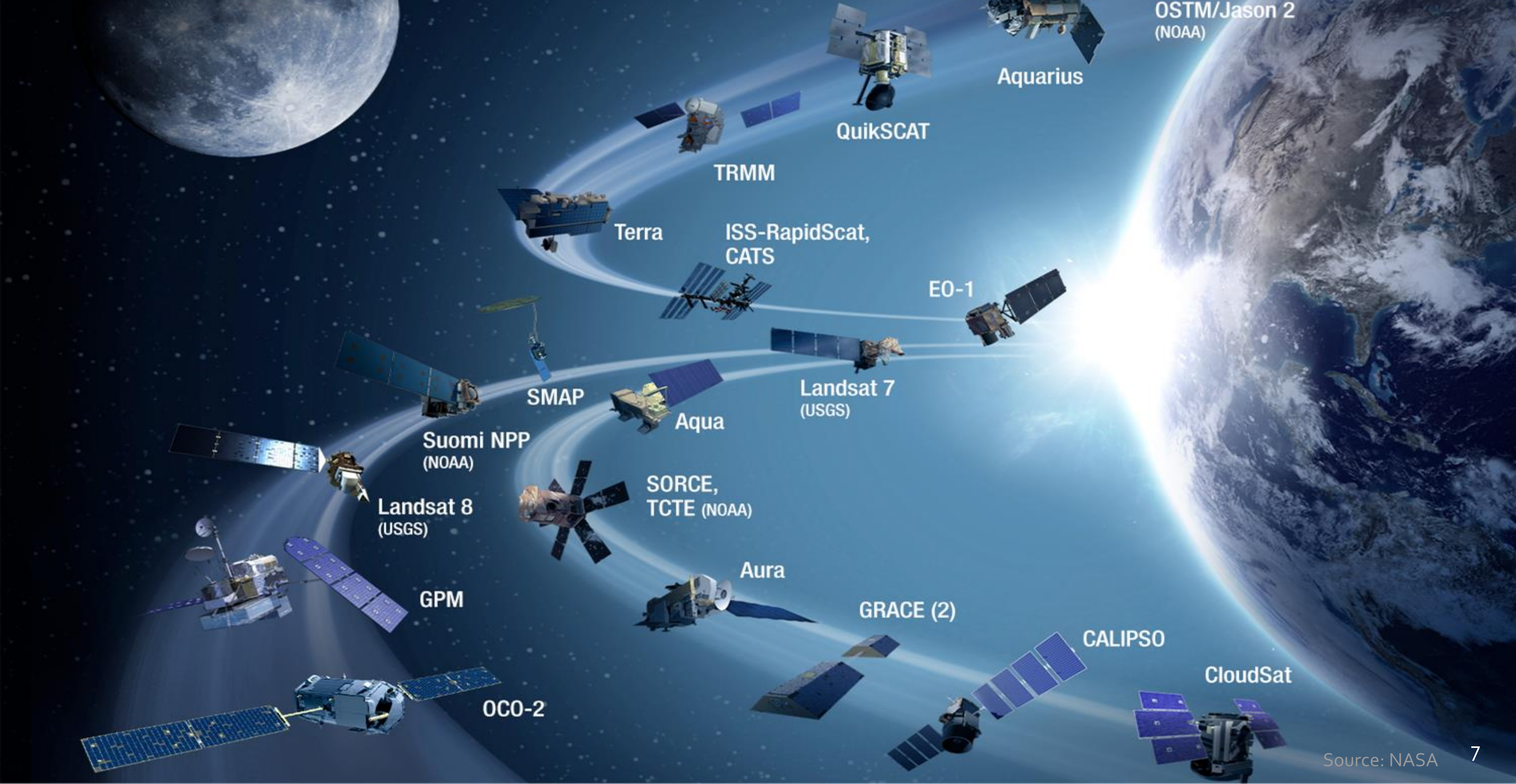
- monitor the evaluation of situation when it needed,
- get detailed information about areas of interests,
- check the situation at the cross border areas,
- partly automate the environment control process.



Machine learning (Supervised learning)



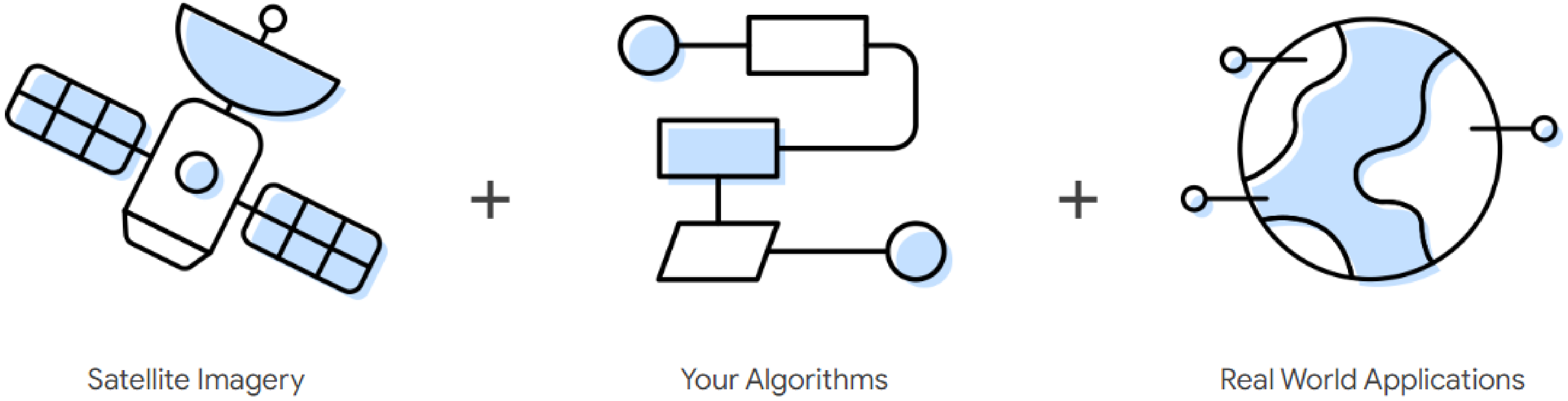
Satellite programs



Source: NASA

Google Earth Engine

Google Earth Engine combines a multi-petabyte catalog of satellite imagery and geospatial datasets with planetary-scale analysis capabilities. Scientists, researchers, and developers use Earth Engine to detect changes, map trends, and quantify differences on the Earth's surface. Earth Engine is now available for commercial use, and remains free for academic and research use.

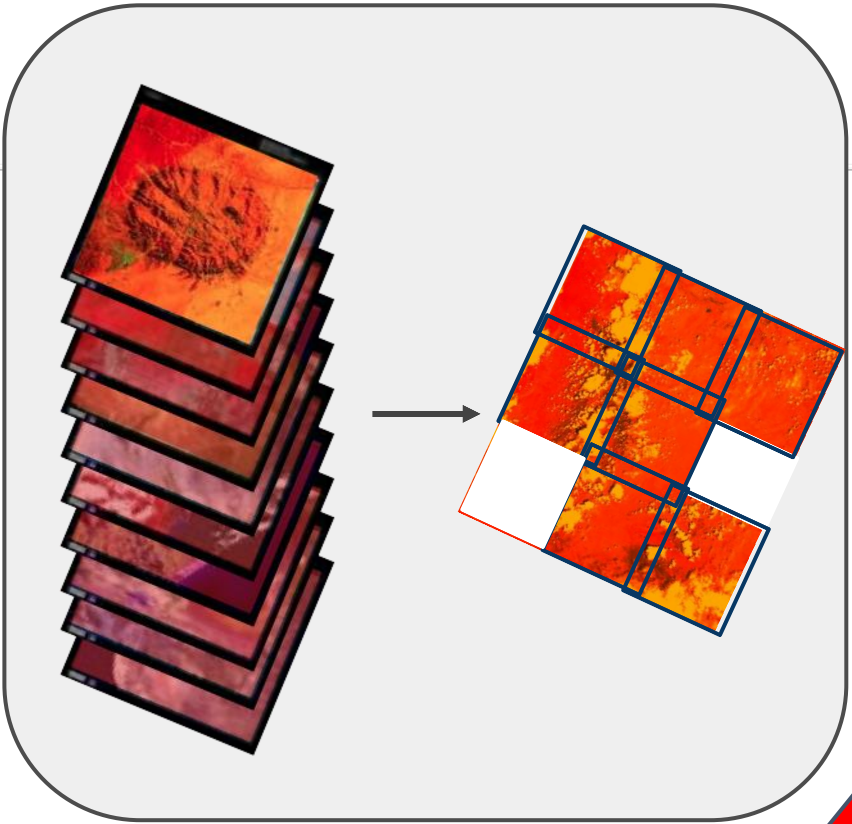


Earth Engine provides easy, web-based access to an extensive catalog of satellite imagery and other geospatial data in an analysis-ready format. The data catalog is paired with scalable compute power backed by Google data centers and flexible APIs that let you seamlessly implement your existing geospatial workflows. This enables cutting-edge, global scale analysis and visualization.

Reduce

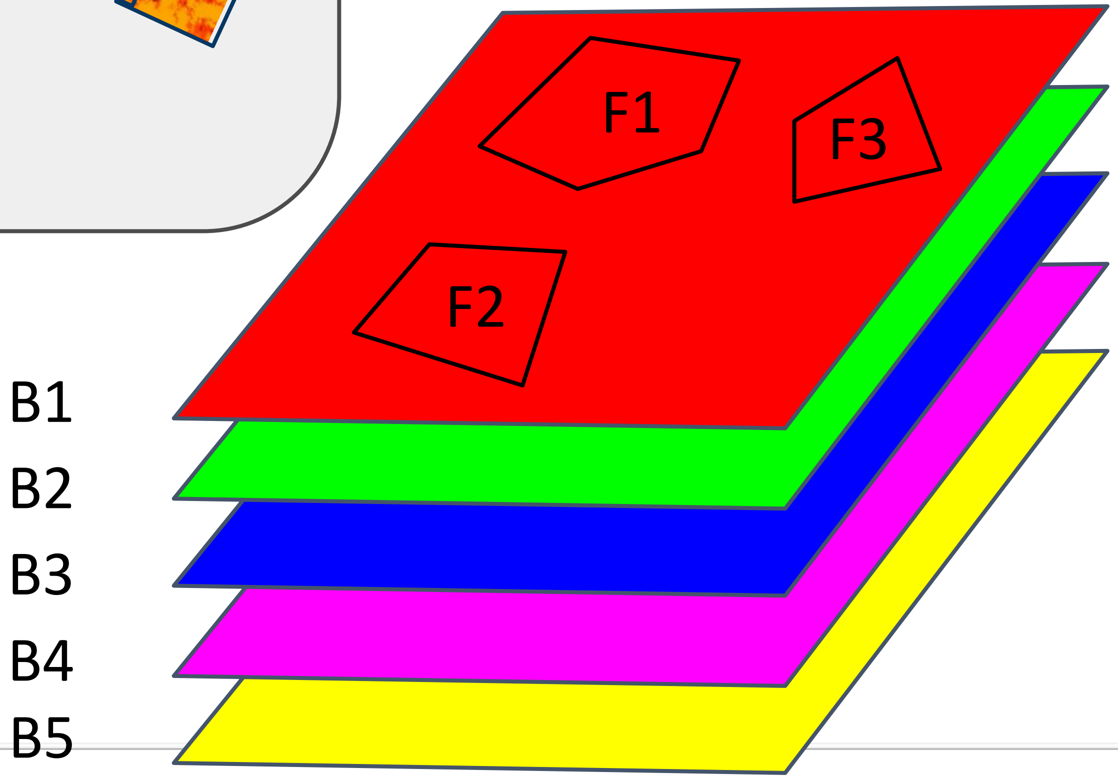
Aggregate everything in a collection

"Reduction"



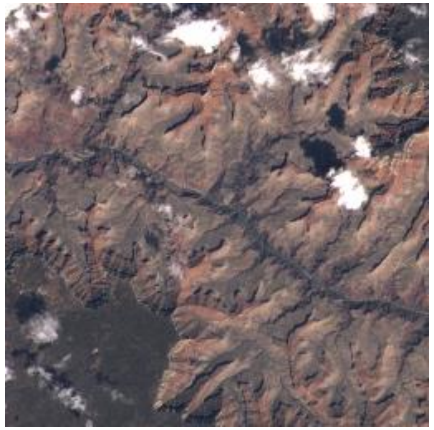
Examples

- Summed area over all features
- Median-pixel composite
- Train a classifier



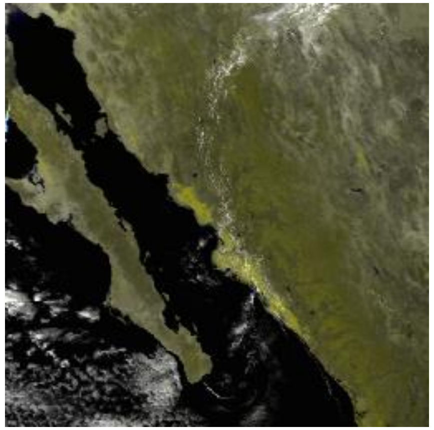
Google Earth Engine

There are more than **100 satellite programs** and modeled datasets. Google Earth Engine has **JavaScript online editor** to create and verify code and **python API** to communicate with user's applications.

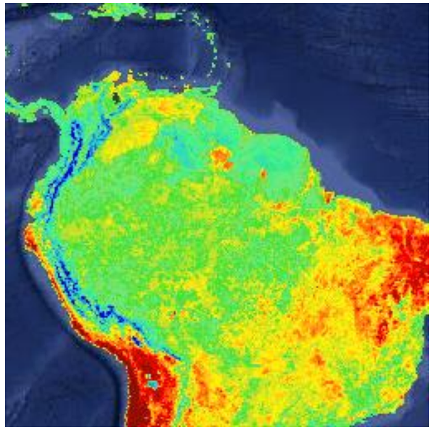


Landsat (15-30m Resolution)

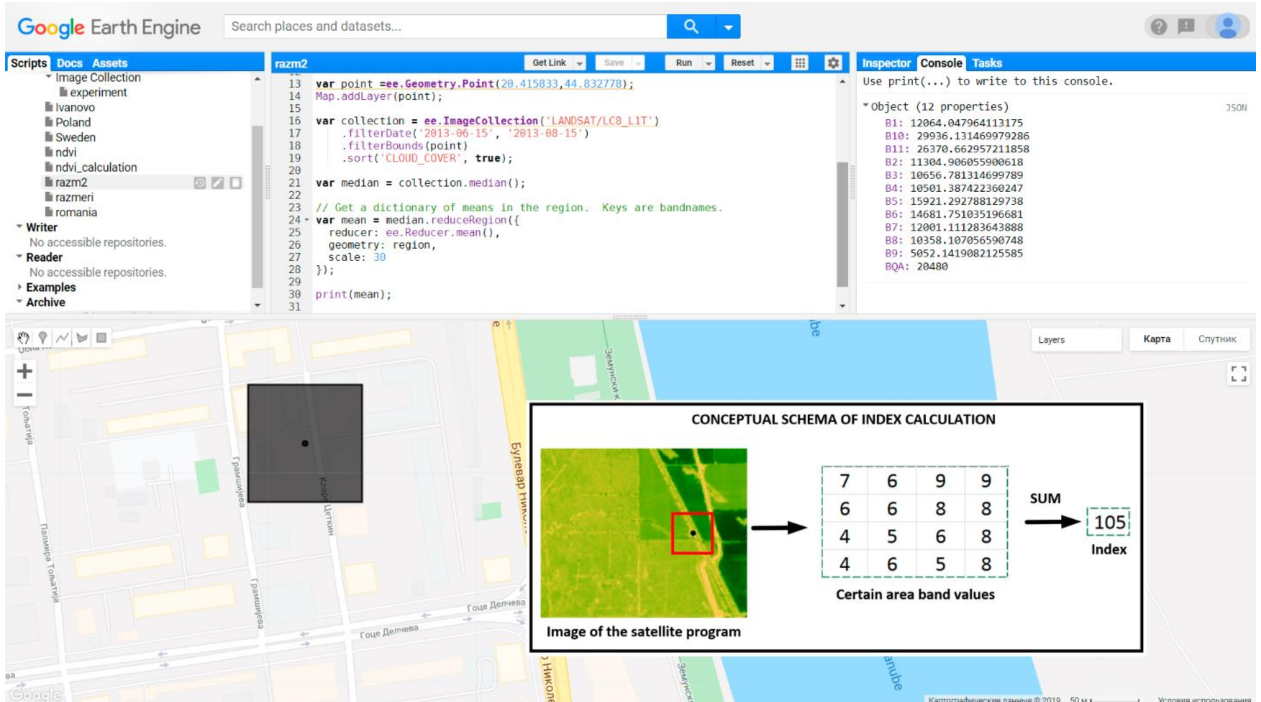
Modis (250-500m Resolution)



Sentinel (250-500m Resolution)

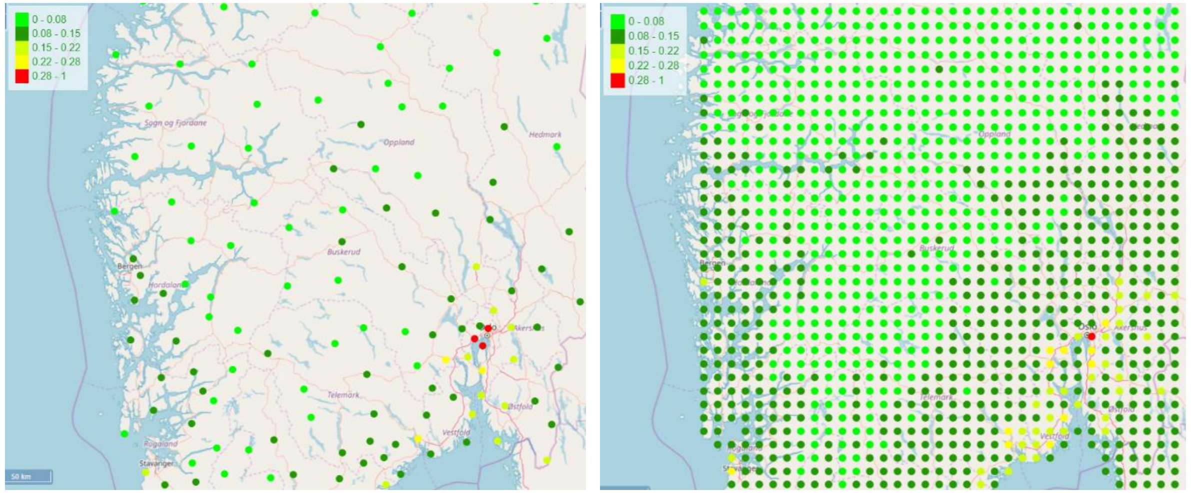


The MOD11A2 V6 average 8-day land surface temperature (LST) in a 1200 x 1200 kilometer grid.



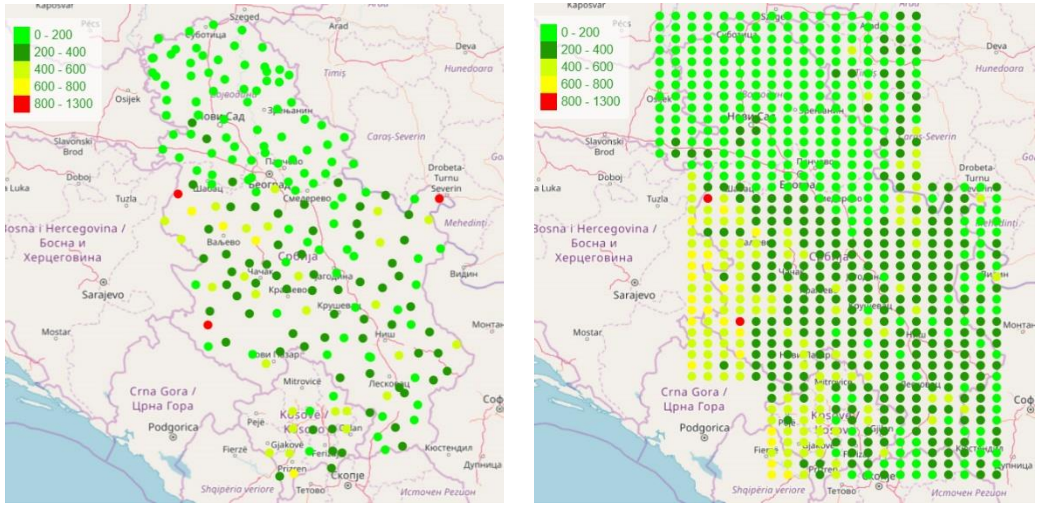
Specify program and time-period to get a collection of images, for example, program – “MODIS/006/MOD09A1” from 2013-06-15 to 2013-08-15 (the period relevant for in situ biomonitoring). Then, define the analyzed area, for example, a square kilometer, with center at the coordinates where sampling was performed. During the satellite data collection, under the bands (channels) of the median image, we execute some mathematical functions (max, min, median, etc.) and get the numerical values.

Results on the regional level

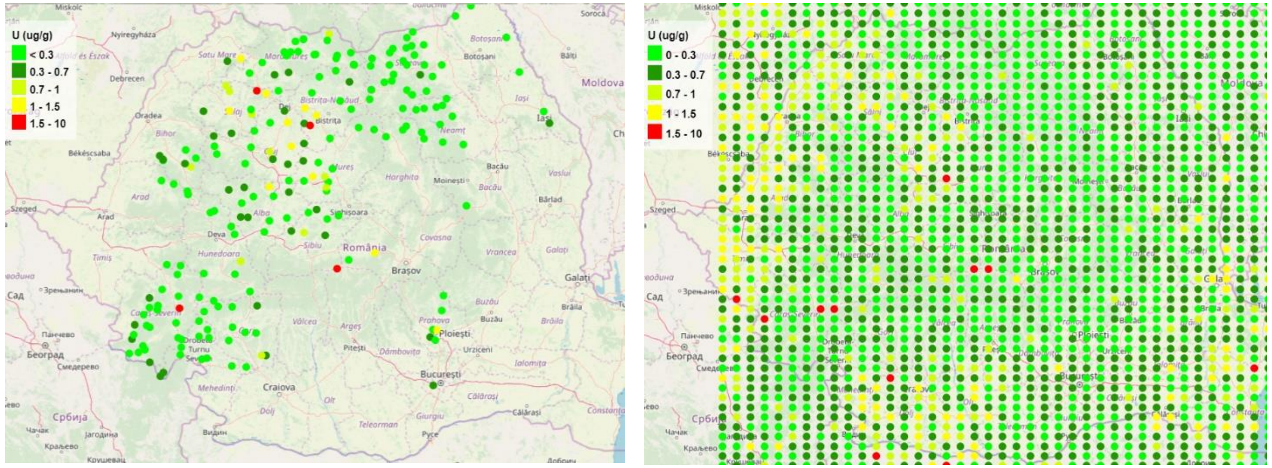


Sb at Norway. Left – real life, right - prediction

Candidates for modeling:
Al, As, Cr, Cu Fe, Mn, Ni, Pb, V, Sb, U …



Mn at Serbia. Left – real life, right - prediction



U at Romania. Left – real life, right - prediction

Local level

The goal of this study was to facilitate the highly resolved mapping of the presence of potentially toxic elements in the air of an urban area, which is typically characterised by high and variable pollution. + to check whether model can keep appropriate accuracy during long time period.

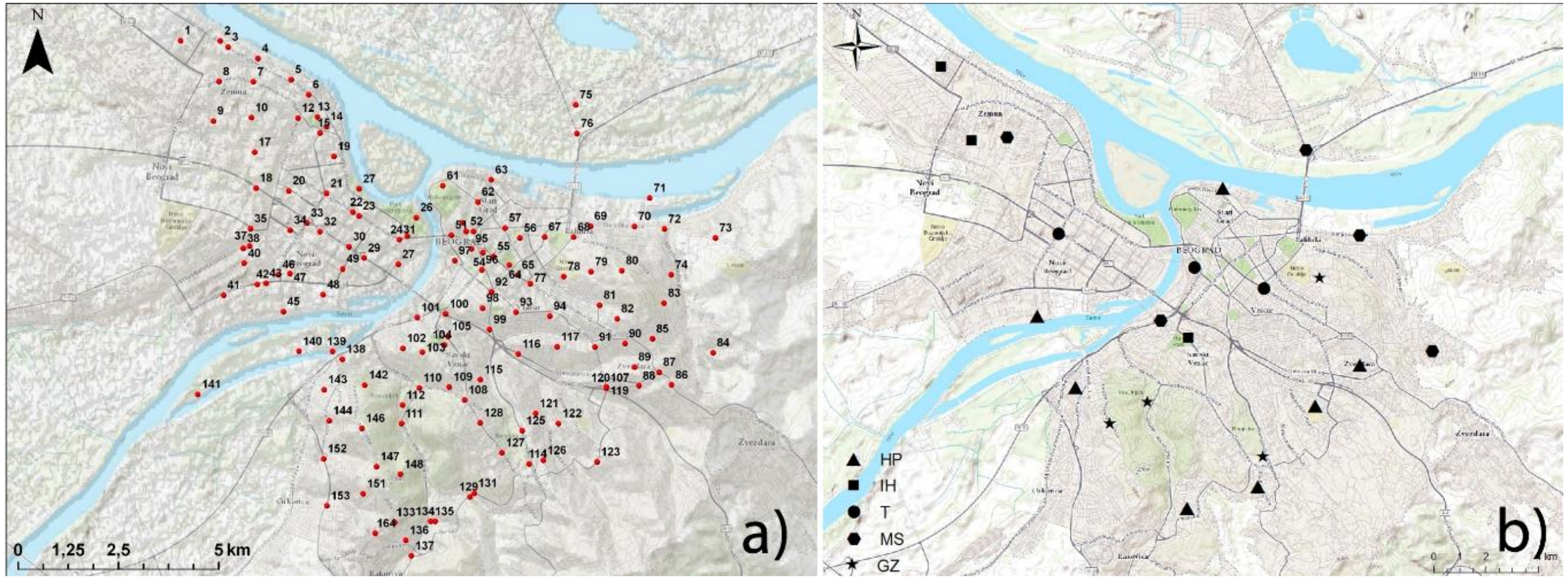


Figure 1. Moss bag biomonitoring across the Belgrade urban area; maps of the sampling sites during two seasons: (a) summer (urban, suburban and green zones) and (b) winter (U–urban sites, GZ–green zones)

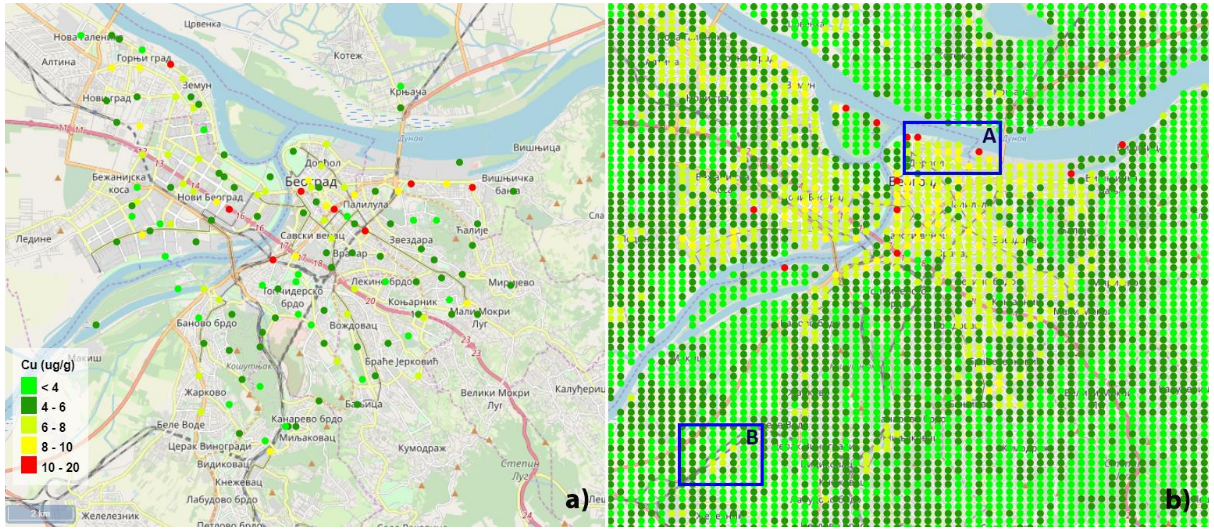


Figure 3. Concentration of Cu in the summer of 2013 (Belgrade): a) real measurements, and b) prediction values; area A represents central part of Old Belgrade with permanently high traffic flow; area B represents a large railway terminal

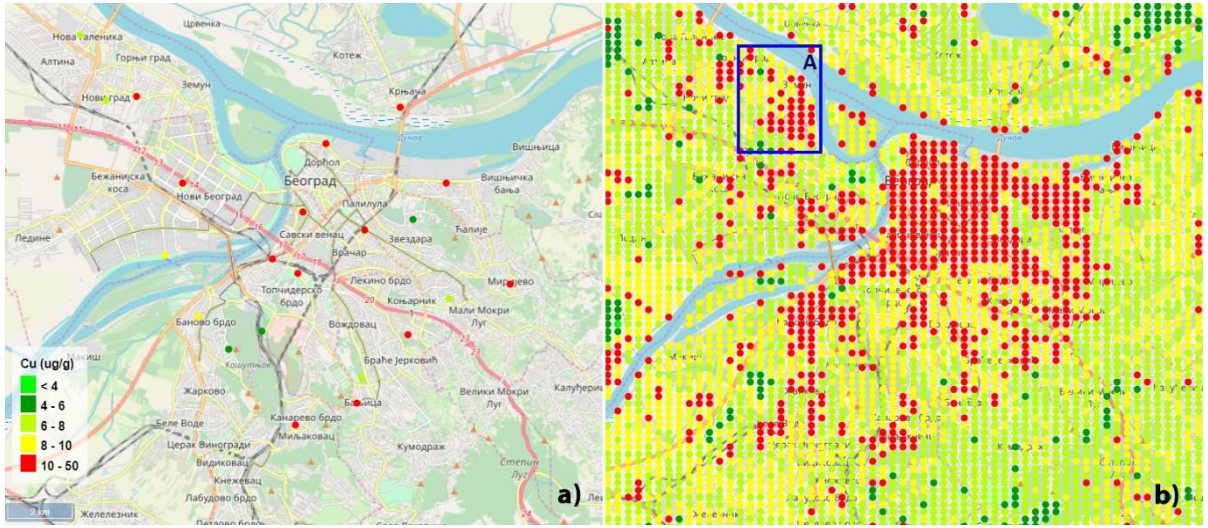


Figure 4. Concentration of Cu in the winter season 2013/2014 (Belgrade): a) real measurements, and b) prediction values; area A represents an old city core highly polluted in winter season

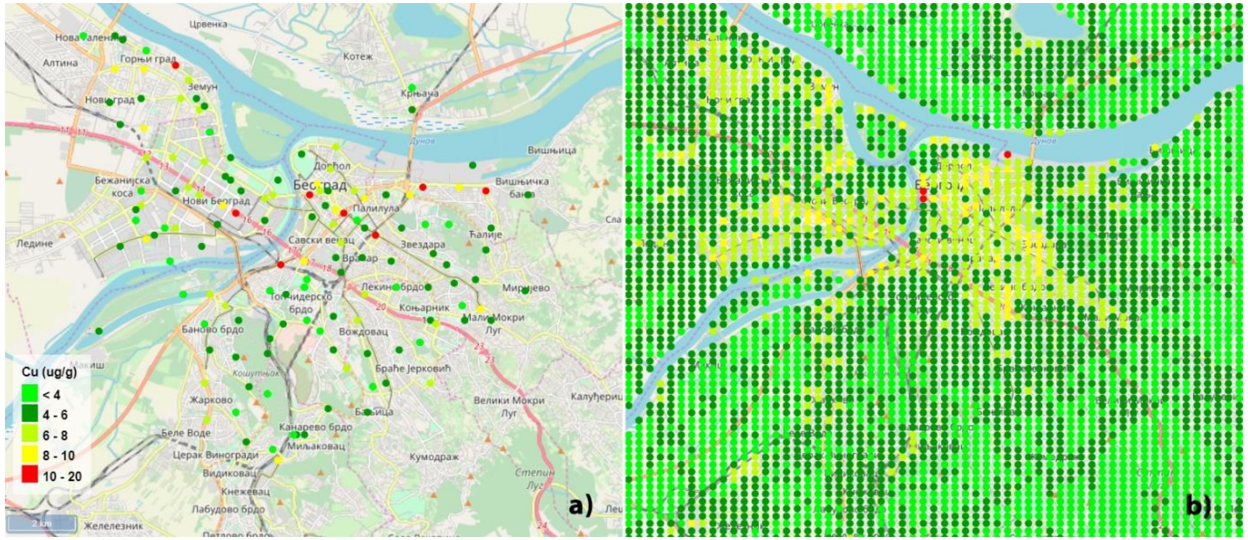
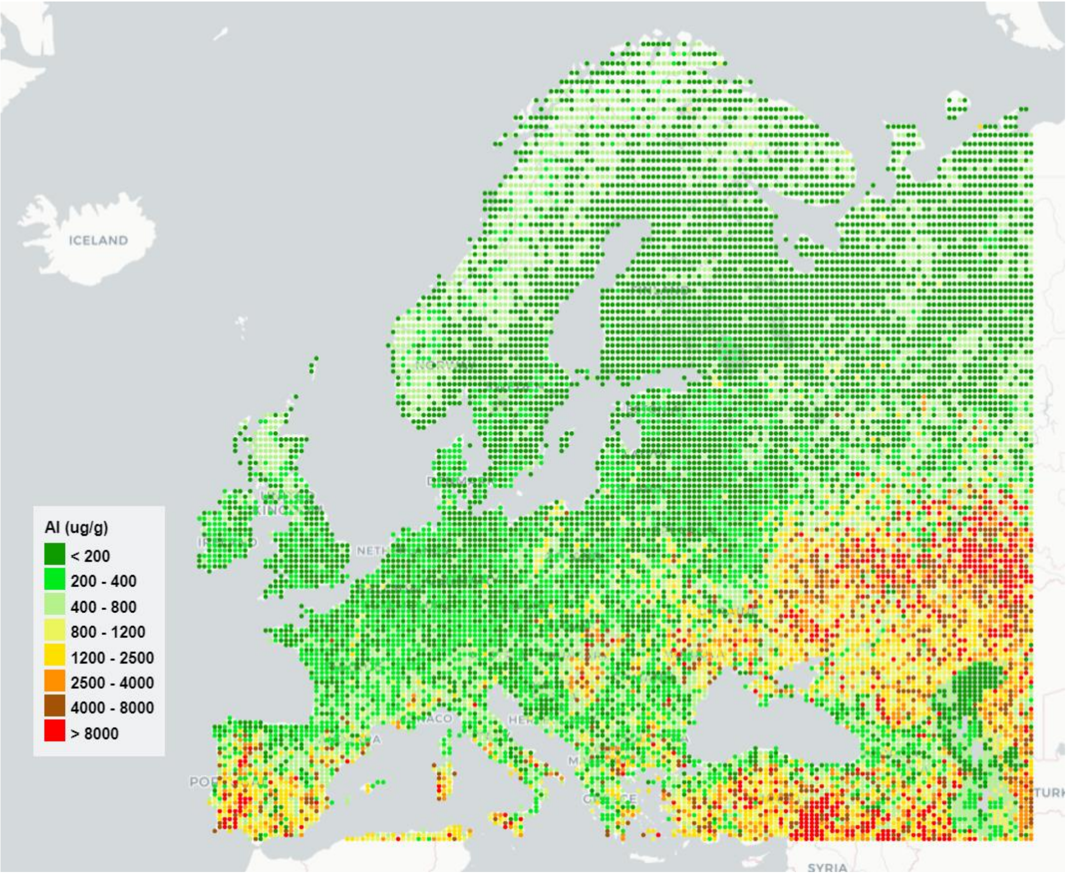
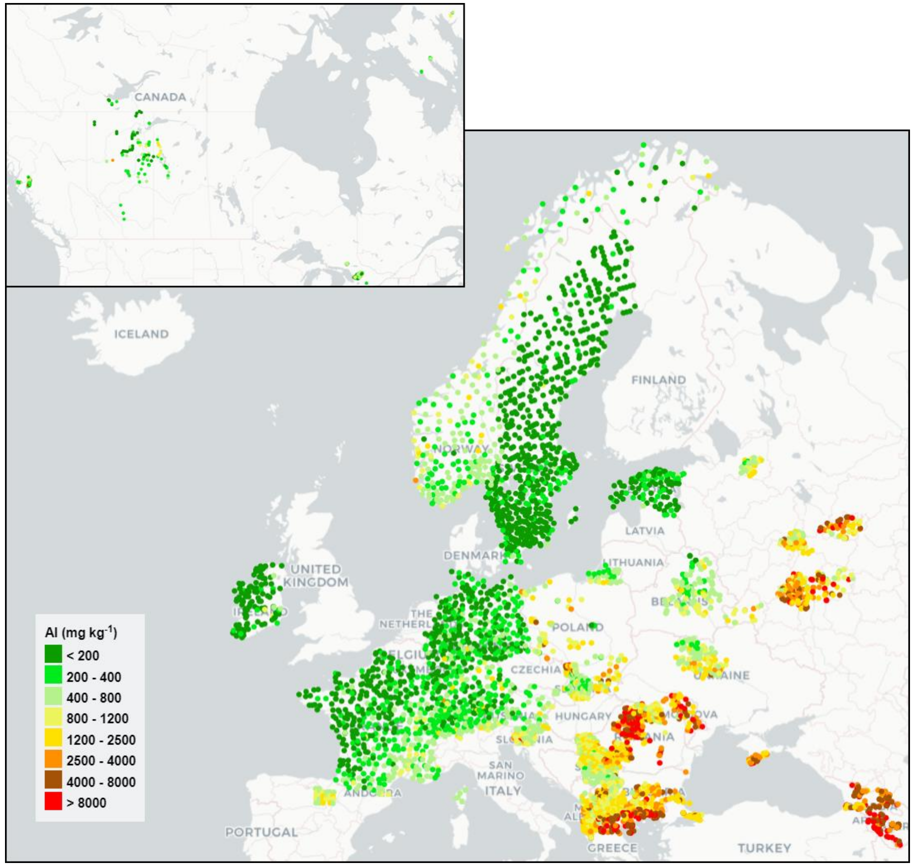


Figure 5. Concentration of Cu in Belgrade: a) biomonitoring measurements in the summer of 2013, and b) prediction for 2018

Prediction of air pollution by potentially toxic elements over urban area by combining satellite imagery, Moss Biomonitoring Data and Machine Learning
A. Uzhinskiy,
M. Anitich Urorevice, M. Frontasyeva, Ciencia e Tecnica Vitivinicola Journal, ISSN:2416-3953, 12, 35, 2020

Draft maps (model trained on Full data)





- More than 4900 sampling sites
- Information about tens elements
- Satellite images of tens programs hundreds of bends

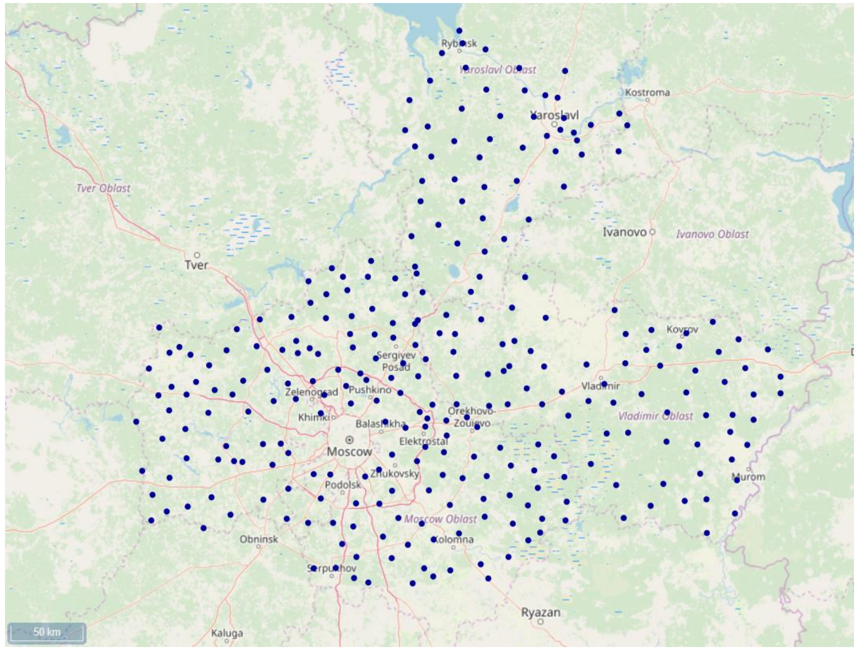
- Imbalanced data (Cr)

0.5 1461
1.5 575
2.5 322
3.5 203
5 221
8 186
14 95
17 144

Approaches:

- Regression and classification (priority) tasks
- Data balancing methods
- Statistical models (learning trees, boosting, etc)
- Deep Neural Networks models
- etc.

Machine learning and neural networks



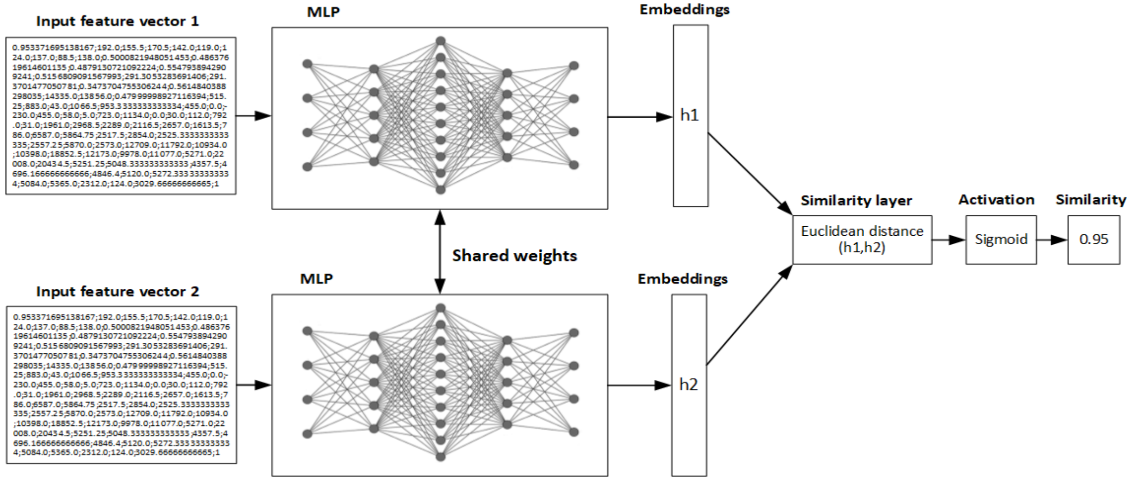
We use the information on 73, 53, and 156 samples from the Vladimir, Yaroslavl, and Moscow regions gathered in 2018 - 2019.

The indices are gathered based on data from 13 programs for 281 sampling sites, and their linkage with the concentration of 18 heavy metals is verified. Altogether 9 HMs, i.e., Al, Fe, Sb, Na, Sc, Sm, Tb, Th, and U, look very prospective for modeling.

We examine three approaches: Gradient Boosting, Multilayer perceptron, and Siamese network.

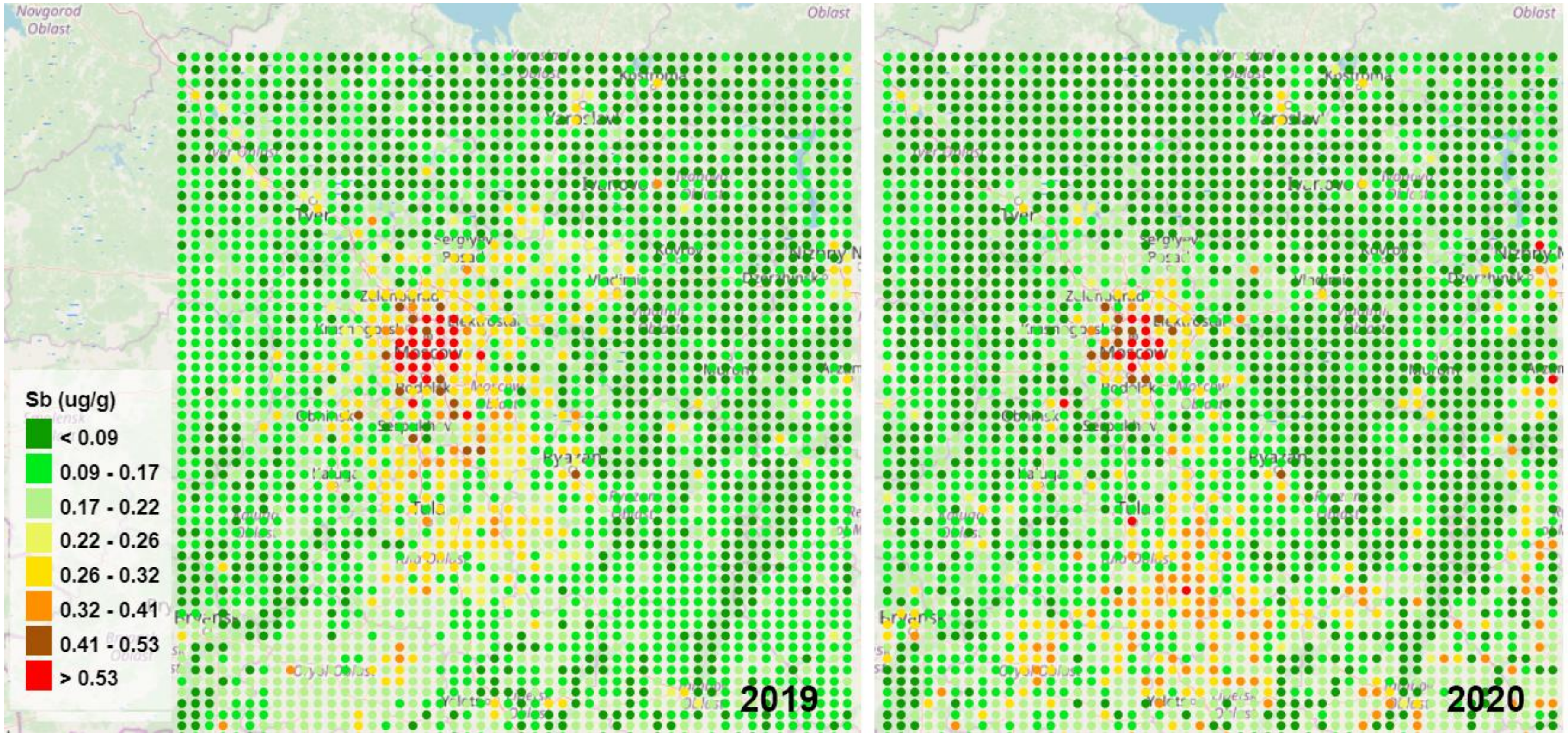
	Al		Fe		Sb	
	Acc si	Acc ai	Acc si	Acc ai	Acc si	Acc ai
GB	0.91	0.92	0.92	0.93	0.94	0.94
MLP	0.89	0.91	0.92	0.92	0.89	0.92
SNN	0.92	0.93	0.93	0.93	0.93	0.94

Table 2. Mean accuracy of the models. GB is gradient boosting. MLP is the multilayer perceptron. SNN is the Siamese neural network. Acc Si is the accuracy on the selected indices. Acc Al is the accuracy on all indices.



Siamese network architecture

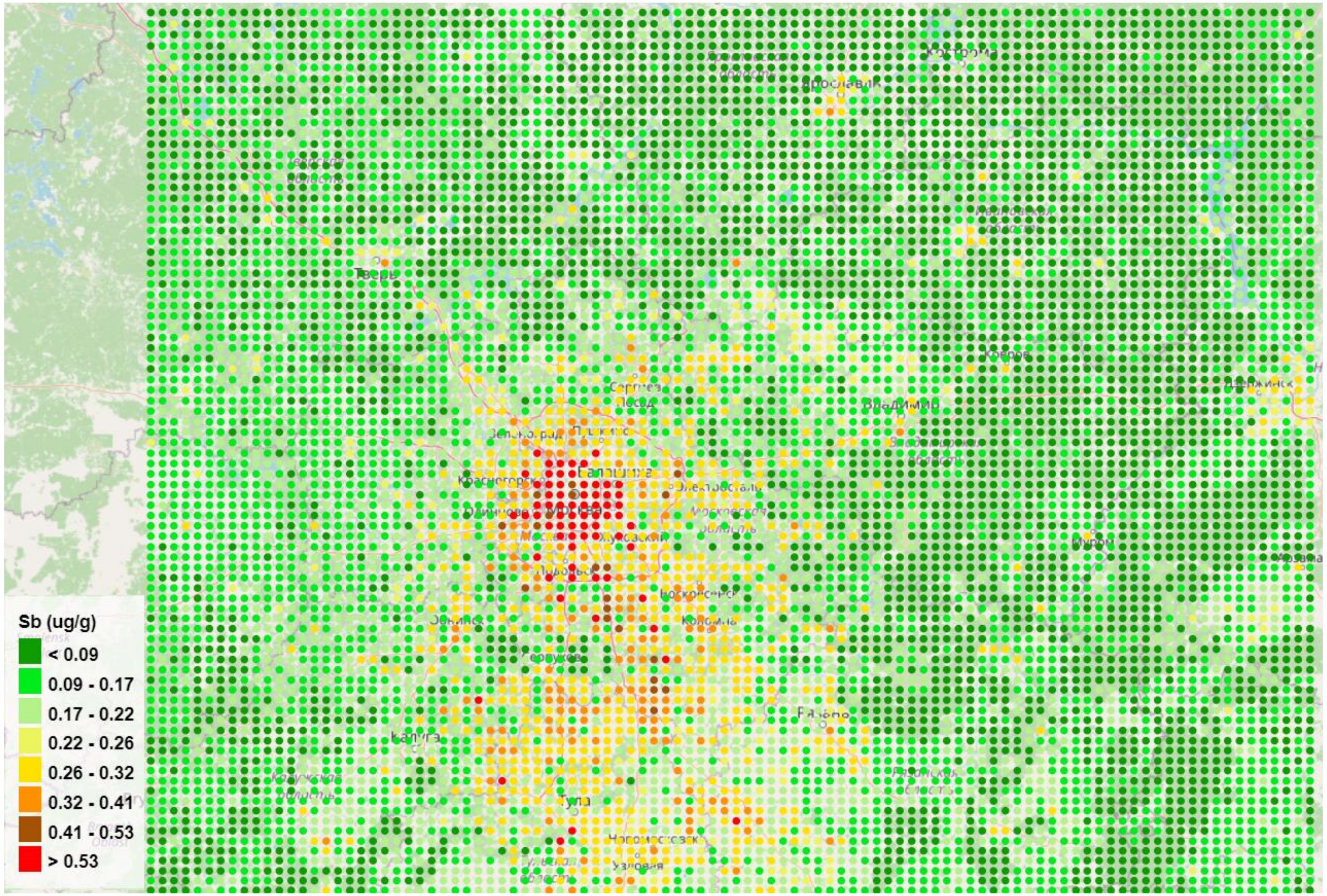
Results (2019 – 2020)



Sb contamination prediction for 2019 (left) and 2020 (right)

The lockdown in Russia that lasted for approximately 1.5 months imposed different limitations. Most of the limitations restricted the movement activities of the population. According to the official statistics, industrial production in Russia decreased by 2.9% from the past, by the end of 2020.

Results (High spatial resolution)



High spatial resolution of the SNN model prediction of Sb contamination

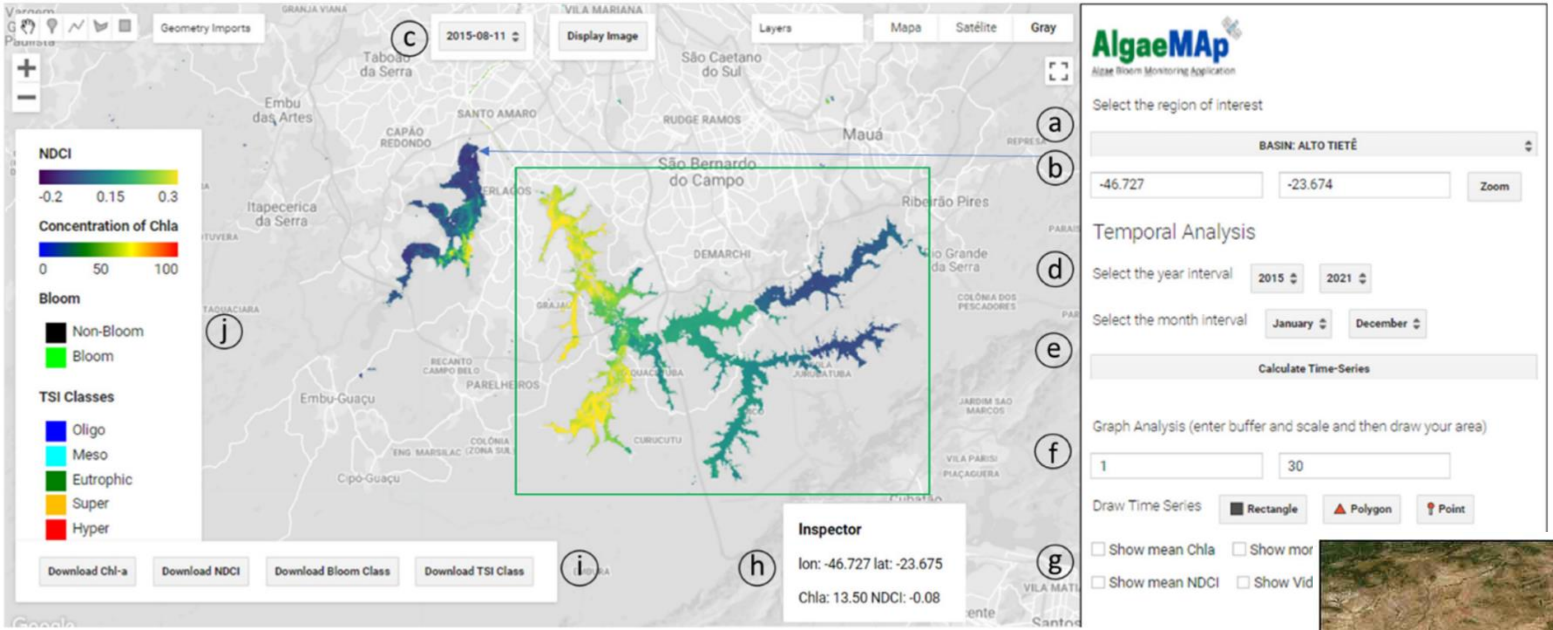
Uzhinskiy A.; Vergel K.; Central Russia heavy metal contamination model based on satellite imagery and machine learning, Computer Optics 2023; 47(1): 137-151. DOI: 10.18287/2412-6179-CO-1149.

Moscow is a thickly populated city, and the population is increasing at a fast pace. Published information reveals, there are about 12.5 million habitants in Moscow. Therefore the Sb contamination level there is bound to be very high.

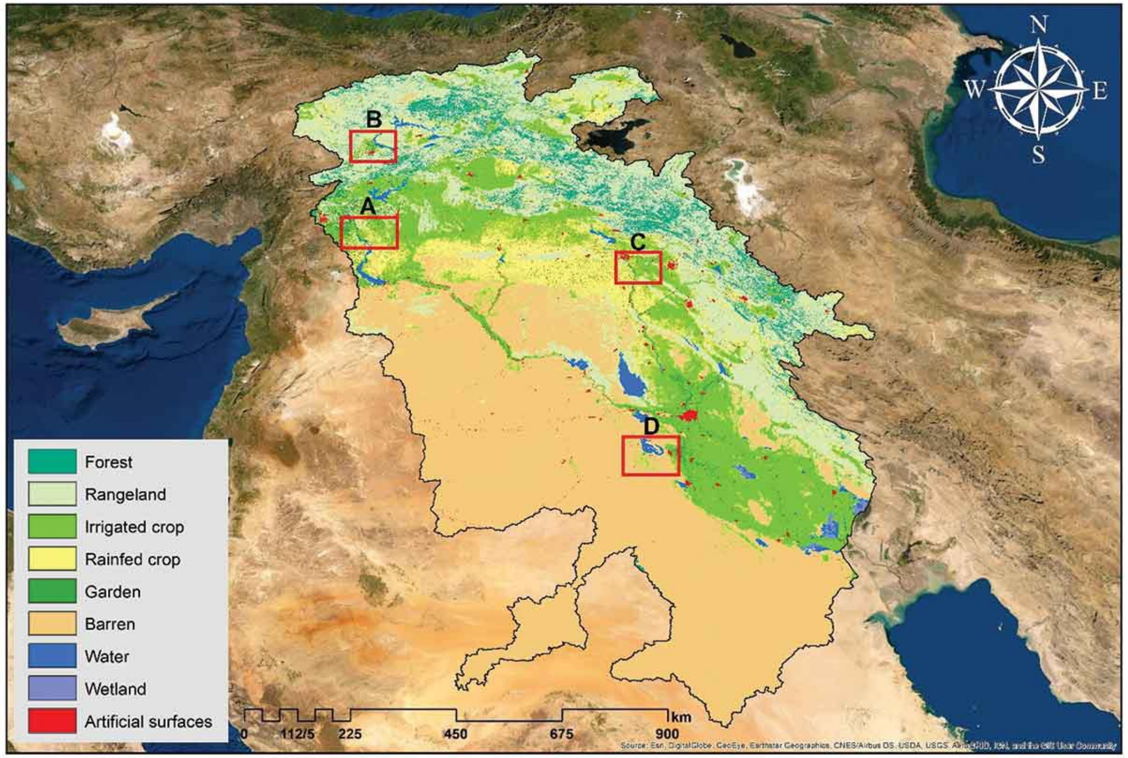
The map also reveals clusters of hot spots in large cities, such as Tula, Kaluga, Vladimir, Tver, Nizhny Novgorod, Yaroslavl, etc. It is also seen that from Sergiyev Posad to the north direction, the contamination level is rather low, except Yaroslavl, where the working oil refinery is located.

The Tula region stands out on the map. There is a multitude of industrial enterprises located in the region, i.e., chemical, metallurgical, and machine-building, besides several large thermal power plants. Huge transport nodes and federal freeways are seen, rather clearly, on the map.

Alert System for Algal Bloom



Soil Mapping and Classification



- туториал на выездном семинаре РЭУ «Научно-исследовательская деятельность ОИЯИ», 5-7 декабря 2023
- проведение демонстраций на осенней it-школе по информационным технологиям ОИЯИ, 16 – 20 October 2023
- туториал на XXVI Летней научной школе молодых учёных и специалистов «Липня-2022», 15 по 17 июля

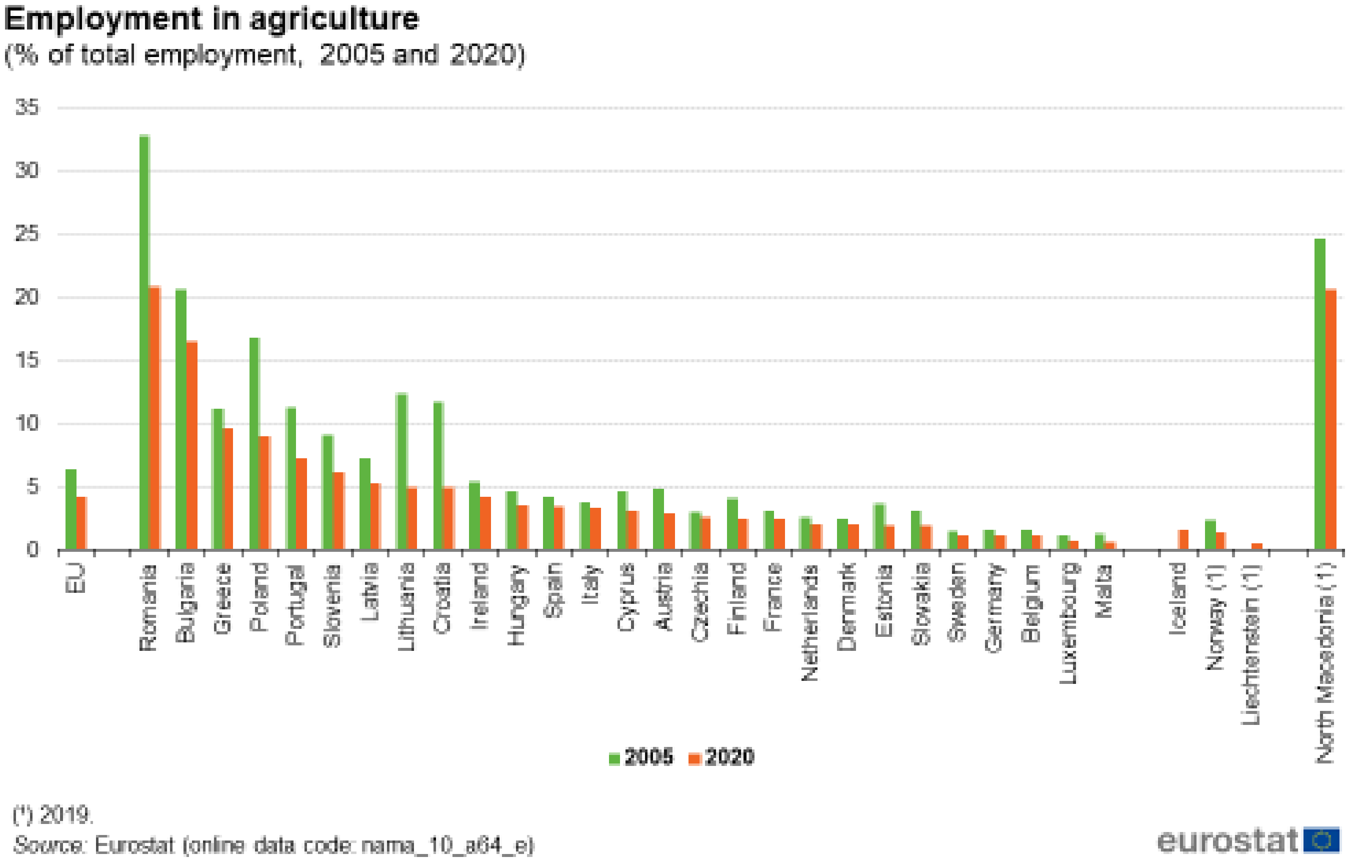
Планы: прогнозирование загрязнения воздуха с использованием Google Earth Engine с акцентом на PM 2.5 и летучие соединения (CO, SO2, NO, …) по данным станций контроля чистоты воздуха

Использование современных средств автоматизации и машинного обучения для нужд сельского хозяйства

Background


WORLD'S POPULATION
9.1 Billion
by 2050
70%
beyond
today's level
FOOD DEMAND SURGING


HELP WANTED


Employment in agriculture
(% of total employment, 2005 and 2020)
35
30
25
20
15
10
5
0
EU
Romania
Bulgaria
Greece
Poland
Portugal
Slovenia
Latvia
Lithuania
Croatia
Ireland
Hungary
Spain
Italy
Cyprus
Austria
Czechia
Finland
France
Netherlands
Denmark
Estonia
Slovakia
Sweden
Germany
Belgium
Luxembourg
Malta
Iceland
Norway (¹)
Liechtenstein (¹)
North Macedonia (¹)
2005
2020
(¹) 2019.
Source: Eurostat (online data code: nama_10_a64_e)
eurostat

Advanced technologies in agriculture

- IoT, sensors,
- remote sensing,
- big-data analysis,
- robots,
- drones,
- digitalization,
- artificial intelligence,
- etc.

There are also many interesting projects in chemistry-, biology-, genetic- and other areas

Animal husbandry is very interesting area with great impact of advanced technologies, but it is out of scope of the report!

Artificial intelligence in agriculture

- Soil management,
- problems detection,
- crop health monitoring,
- yield prediction,
- price forecasting,
- yield mapping,
- optimization of pesticides and fertilizers usage,
- etc.

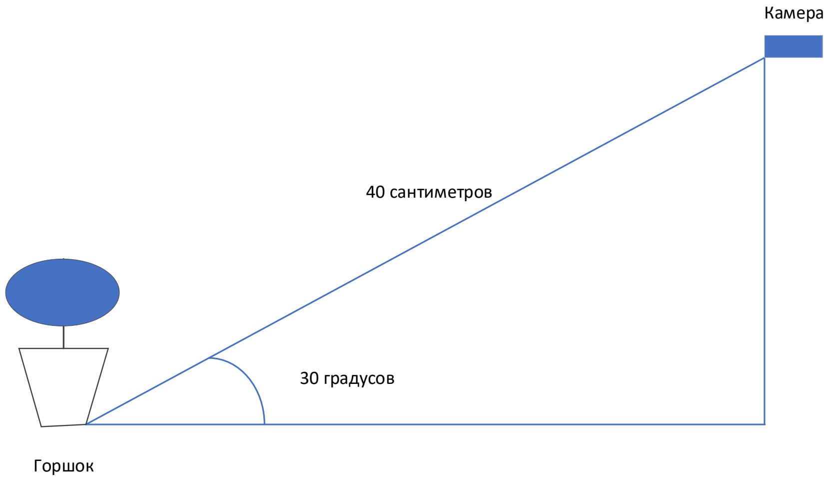
Plants doctor


```
{
  "error":0,
  "general_predictions":[
    {...},
    {...},
    {...}
  ],
  "custom_predictions":[
    {...},
    {
      "type":"розы",
      "prediction":[
        {
          "name":"Мозаика",
          "sample":"http://usrbb.ru/botainika/diseases/rmv1.jpg",
          "description":"<p>Это наиболее распространенное заболевание вирусного типа. Узнать его очень просто – листья покрываются хлоротичными пятнами и узорами, которые могут перейти в прожилковый хлороз. Листья деформируются (узколистность, курчавость, морщинистость) и постепенно опадают. Рост побегов замедляется, и они не вызревают. Побеги, которые больны, вырезают. При дальнейшем распространении болезни растение уничтожают
```

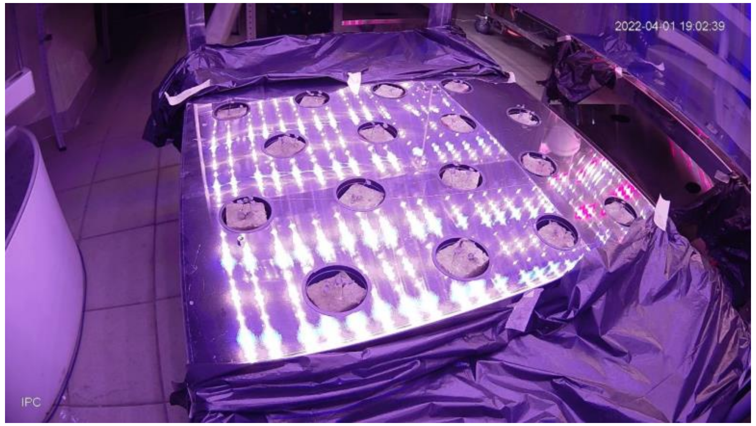

One-shot learning with triplet loss for vegetation classification tasks
A.V. Uzhinskiy, G.A. Ososkov, P.V. Goncharov, A.V. Nechaevskiy, A.A. Smetanin, Computer Optics, ISSN:ISSN 0134-2452, 2021

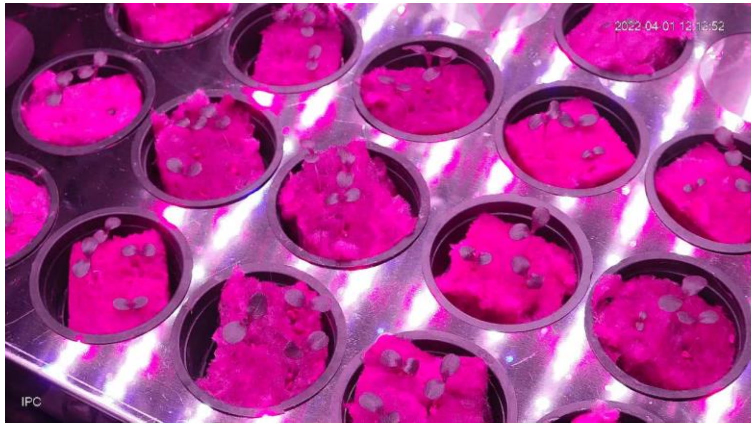
Plant state tracking

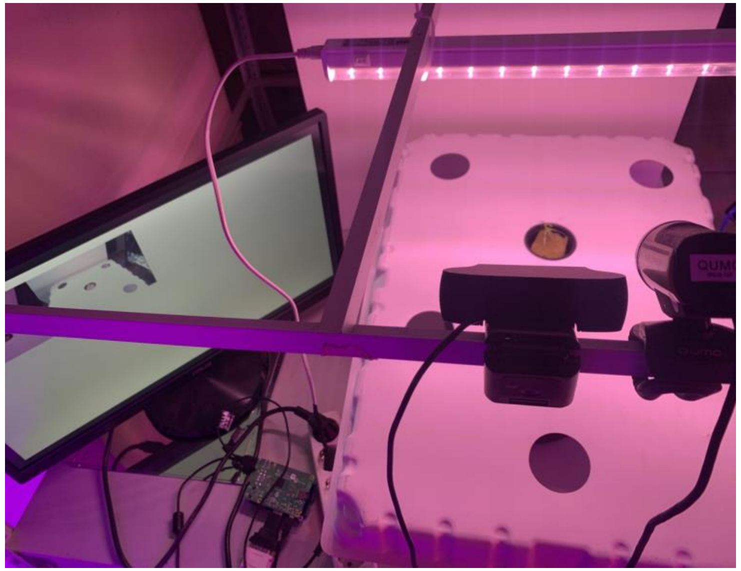
Joint project with the Temiryazev Academy within the framework of the project World-class Scientific Center “Agrotechnologies of the Future”



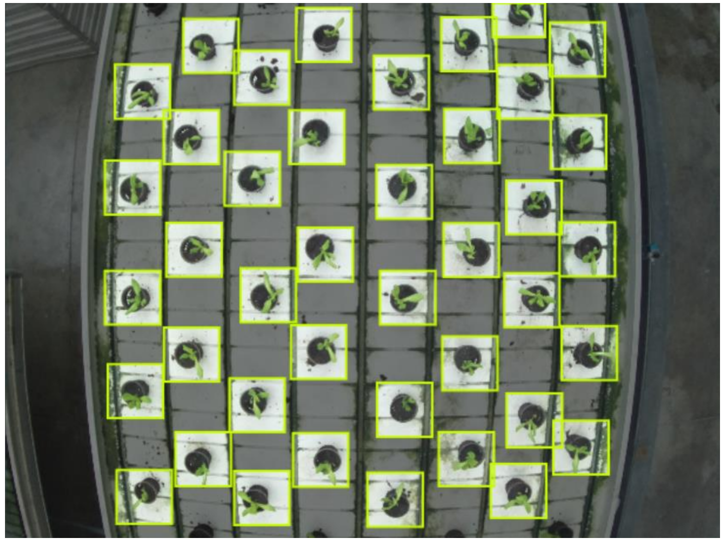
- **Classification of the degree of development of the plant.**
- **Determination of the weight group of the plant.**

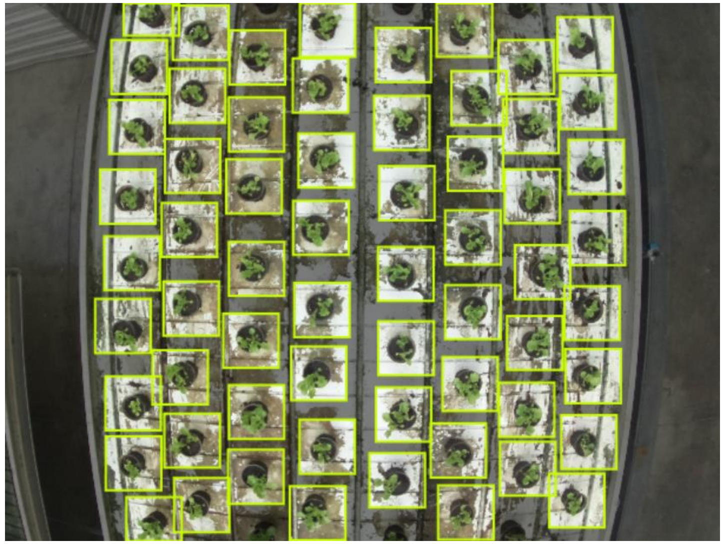


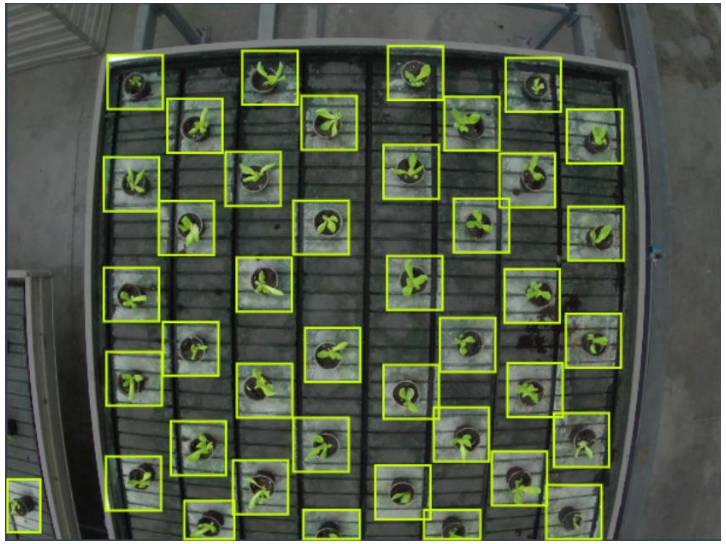


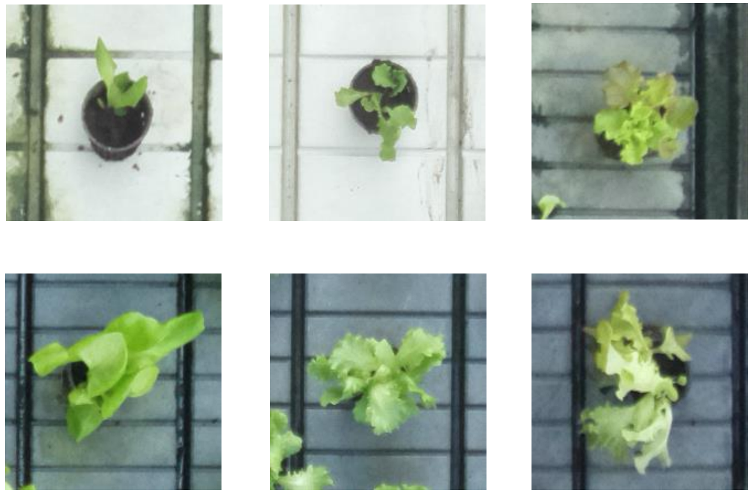


Salads classification



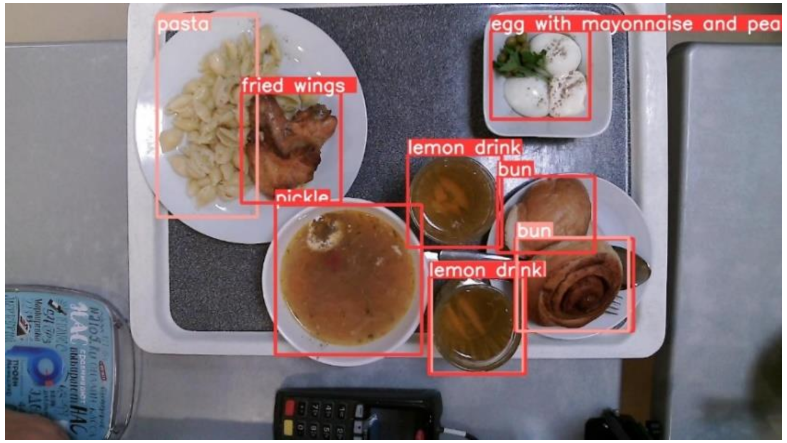






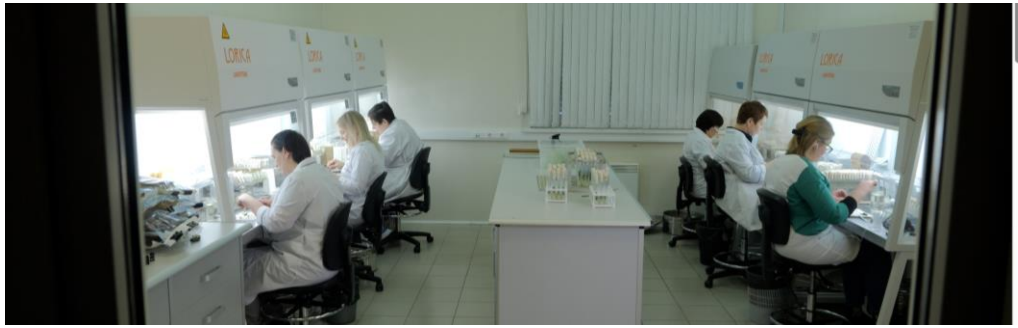
Object detection – 1 class
Classification – 6 classes

Accuracy > 99%



M. Gerasimchuk,A. Uzhinskiy, Food Recognition for Smart Restaurants and Self-service cafes, Physics of Particles and Nuclei Letters, 2024, Vol. 21, No. 1, pp. 79–83

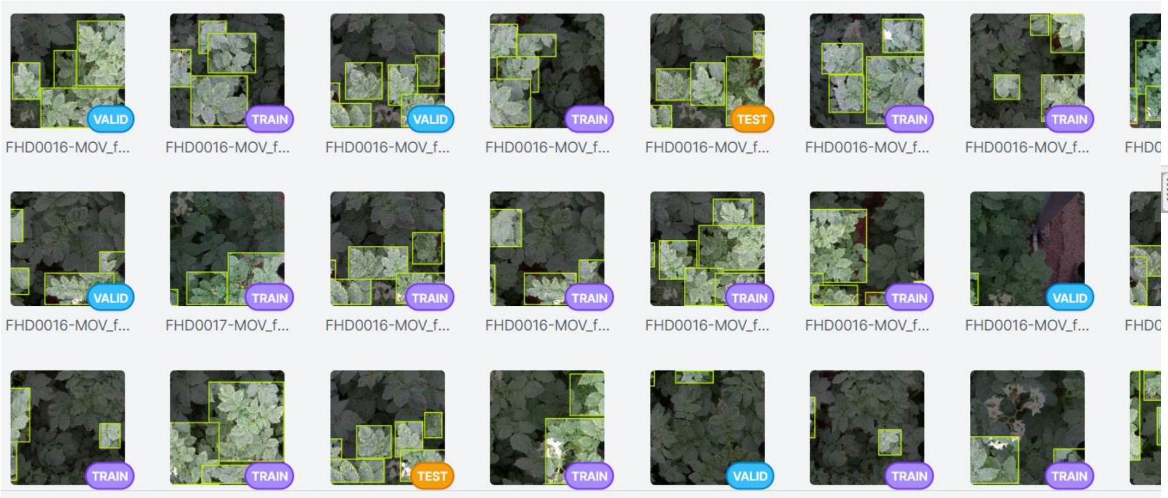

DOKAGENE

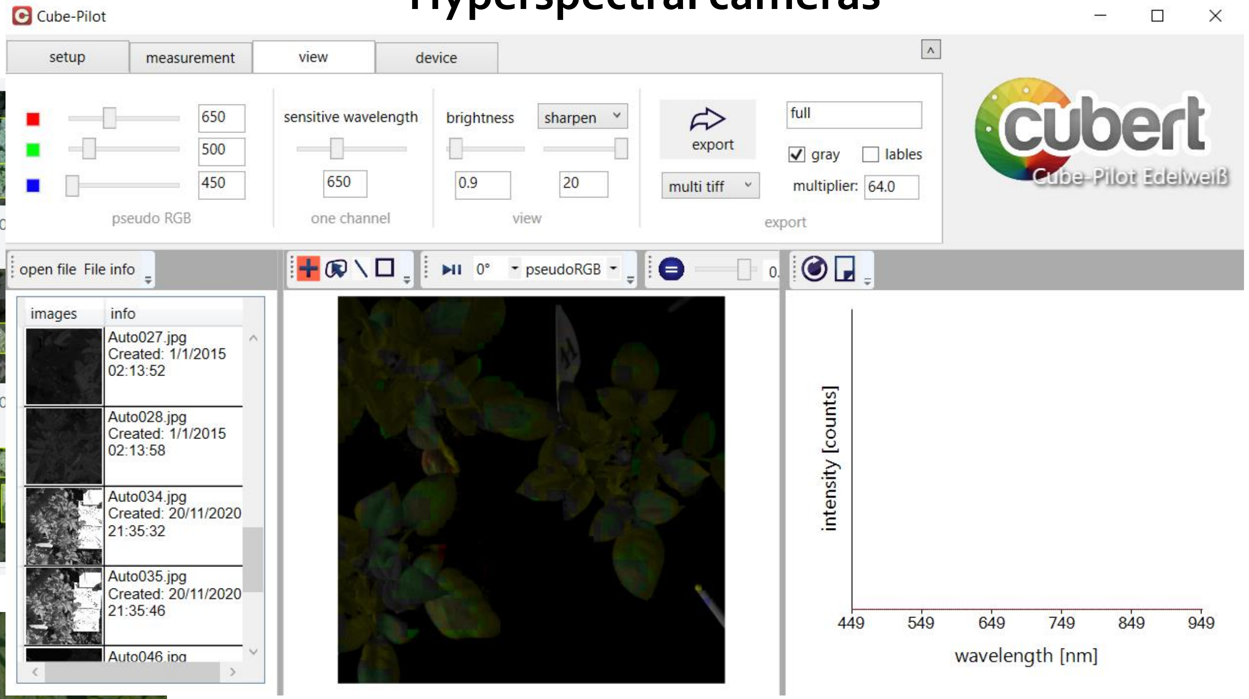


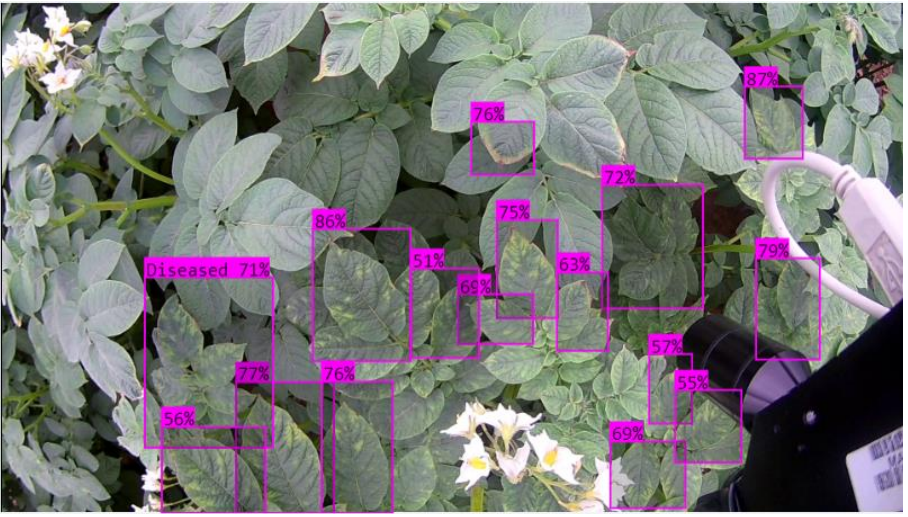
Массивы данных для обучения и тестирования формируются как в условиях производственных участков (RGB камеры размещаются на инспекционных машинах агрономов), так и в контролируемых условиях на стенде (RGB, гиперспектральная камера и VisNIR спектрометр).

Potatoes disease (Doka-Gennyye Tekhnologii)

Hyperspectral cameras







YOLOv7

YOLO-NAS

ultralytics
YOLOv8

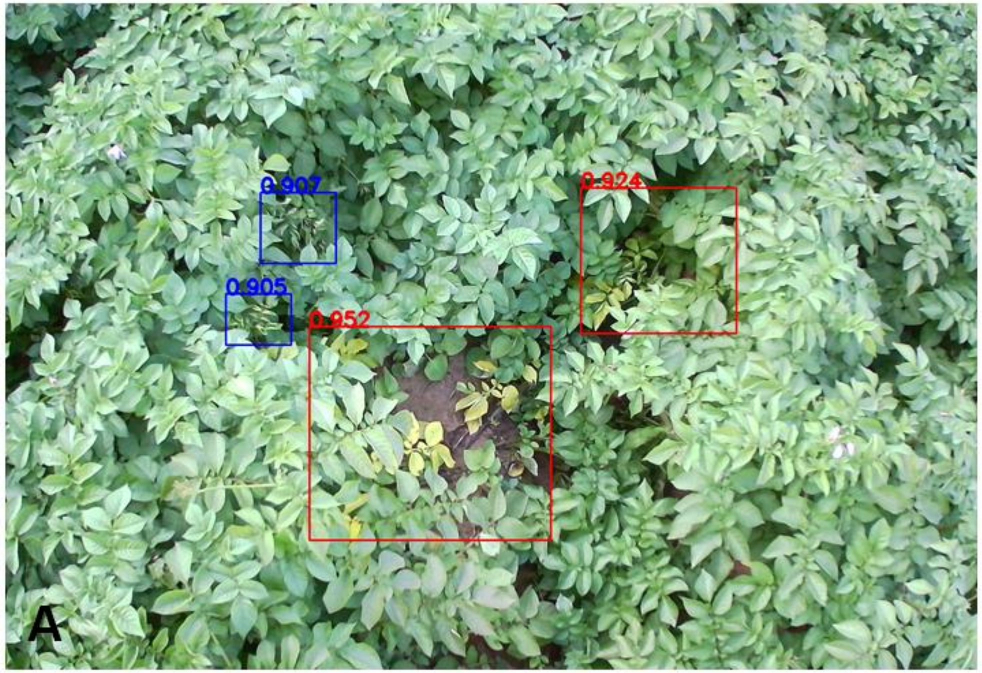
Тестировалась на различных моделях семейств YOLO V5, YOLO V7, YOLO V8, YOLO NAS

1я модель 92 изображения, 200 объектов
2я модель 92 изображения из нового сезона, 170 объектов
3я модель (объединенная 1я и 2я)

Модель	Precision	Recall	mAP50
V1	0.685	0.417	0.516
V2	0.635	0.531	0.586
V3	0.728	0.588	0.631

1я модель 124 изображения, 194 объектов

	Precision	Recall	mAP50
V1	0.827	0.567	0.703



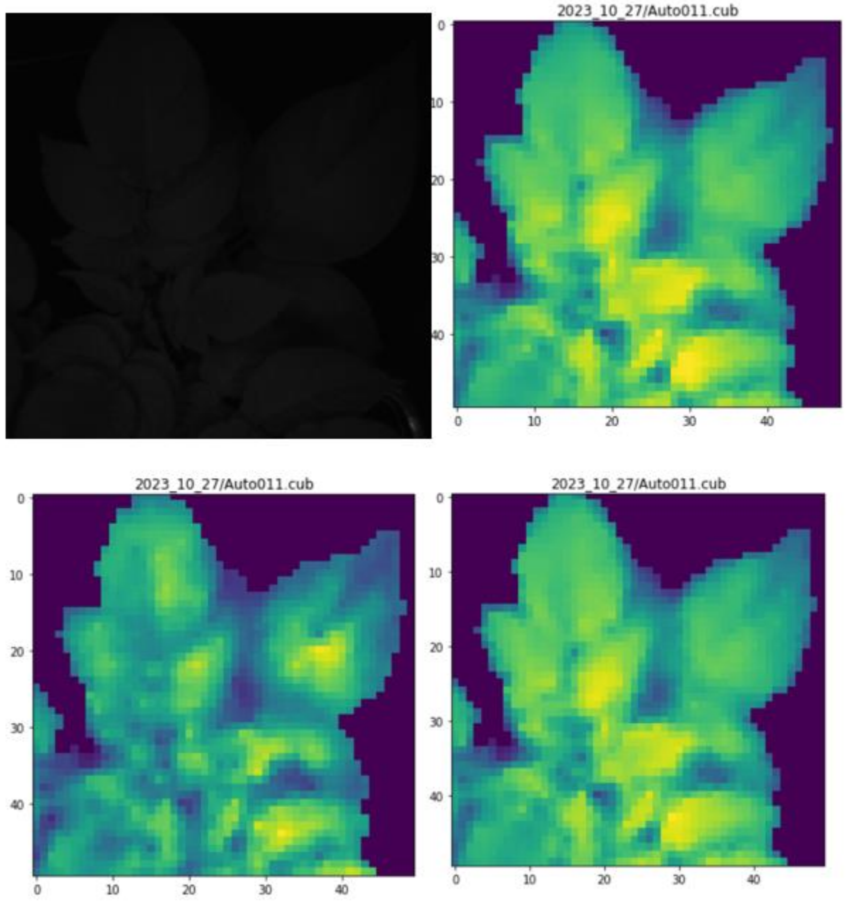


Рисунок 1. Изображение 1000x1000 (слева сверху), 50x50 в 80м канале (справа сверху), 50x50 в 60м канале (слева снизу), 50x50 в 110м канале (справа снизу)

```
Mean cross-validation score: 0.79
              precision    recall  f1-score   support

           0       0.96      0.98      0.97     21642
           1       0.98      0.96      0.97     22338

    accuracy                           0.97     43980
   macro avg       0.97      0.97      0.97     43980
weighted avg       0.97      0.97      0.97     43980

[[21200   800]
 [  442 21538]]
```

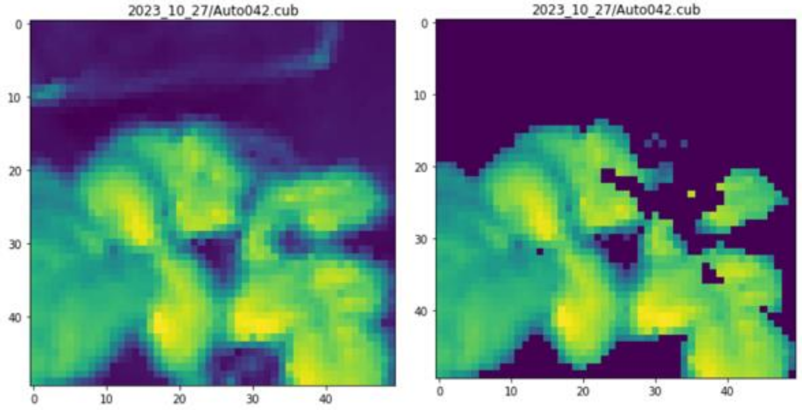


Рисунок 2. Изображение до применения фильтра в 80м канале слева, и после применения фильтра – справа.

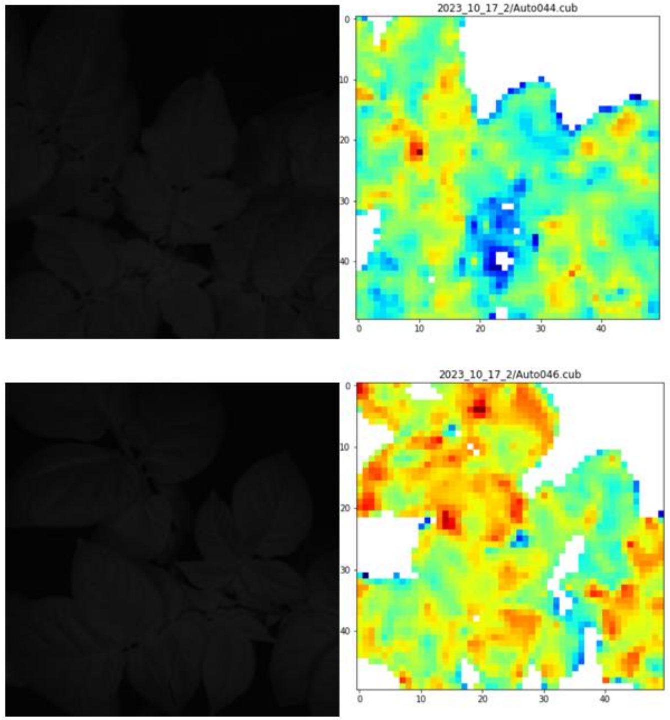


Рисунок 3. Больное растение (верхние 2 фото) и здоровое растение (нижние 2 фото)

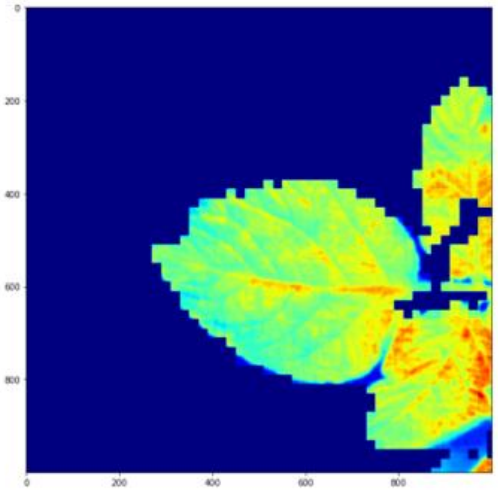
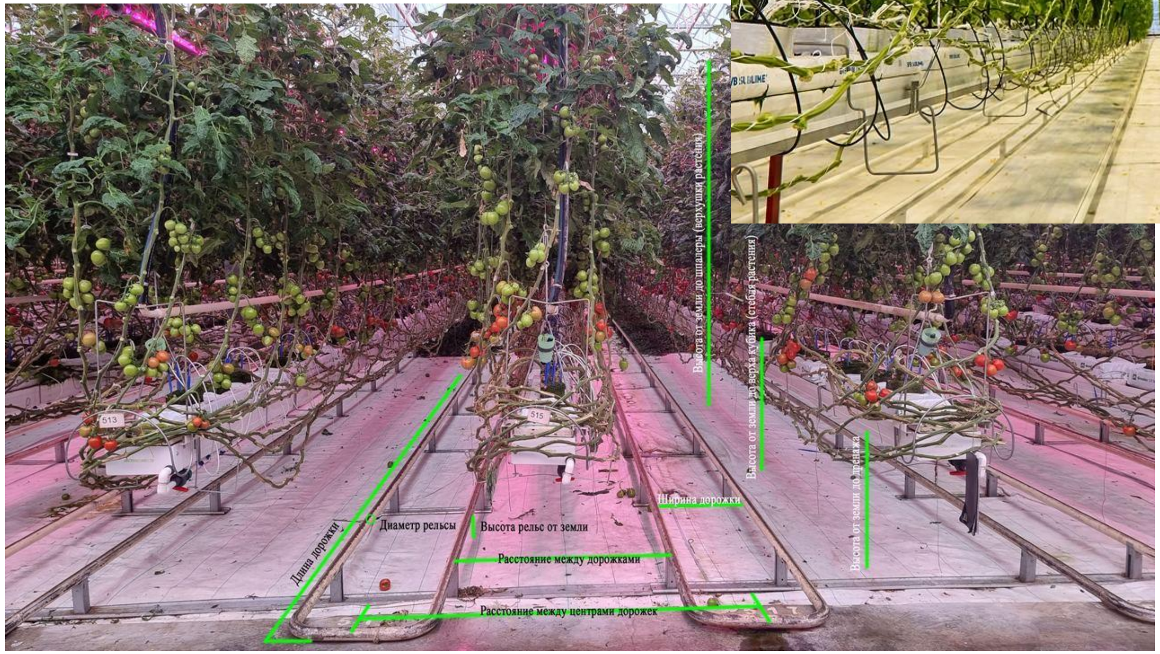
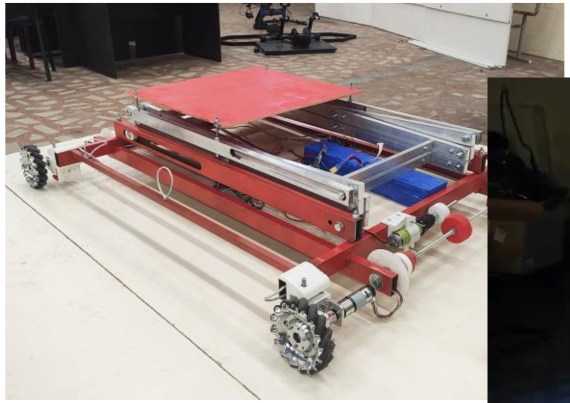
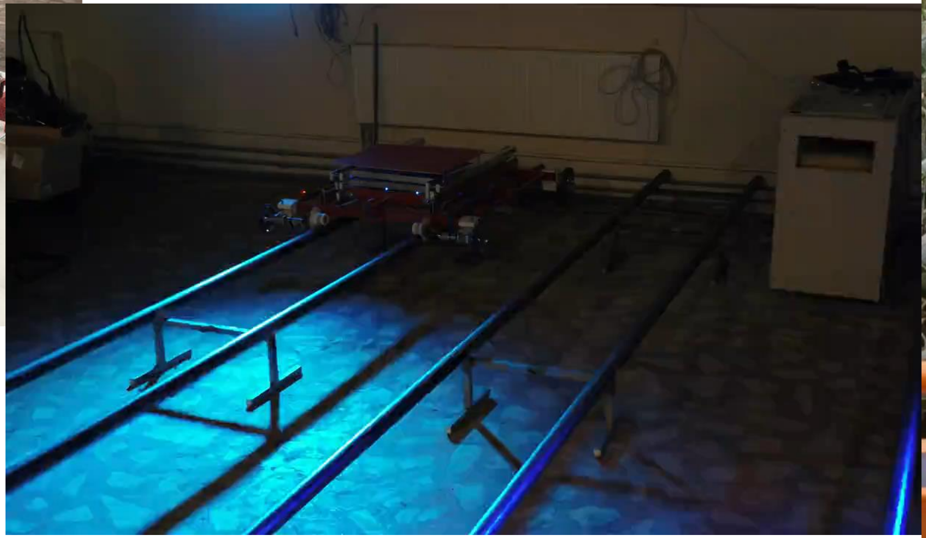


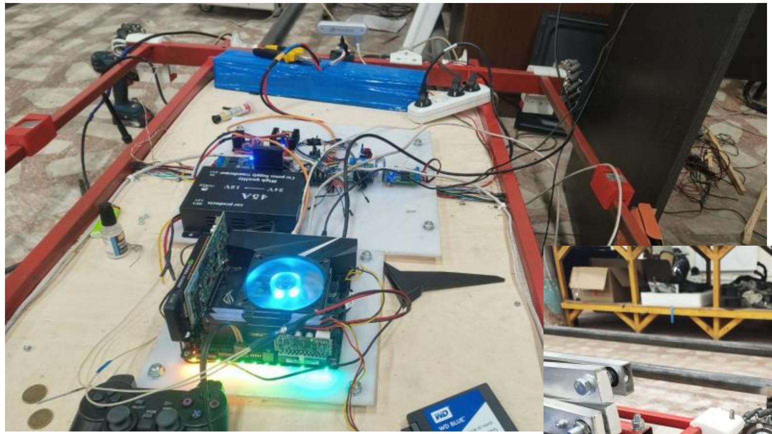
Рисунок 19. Изображение 1000x1000 помноженное на индекс (r=65, nir = 98)

N°	Компания	Регион	Площадь, га
1	УК «Рост»	г. Санкт-Петербург	443
2	Агропромышленный холдинг «ЭКО-Культура»	г. Москва	304,2
3	АО «Агрокомбинат «Южный»	Карачаево-Черкесская Республика	144
4	**ООО «Агро-Инвест»**	г. Москва	105
5	ООО «Агрокультура Групп»	Московская область	100
6	ГК «Горкунов»	Новосибирская область, Ярославская область, Смоленская область, Республика Крым	89,8
7	ООО «ТК «Зеленая линия»	Краснодарский край	85
8	Компания «Ботаника»	Волгоградская область	68
9	ООО «Родина»	Воронежская область	61
10	ГГУП ВОСП «Заря»	Волгоградская область	47

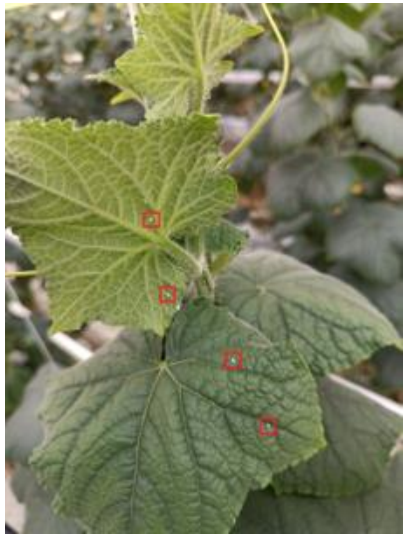
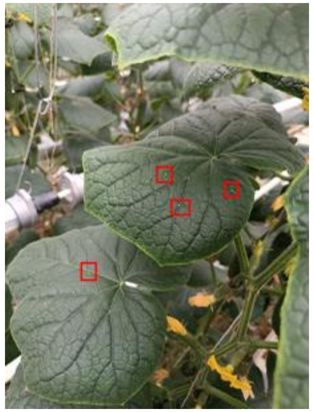
Площадь > 1300 Га
Объем > 1300 тыс. тонн овощей
огурцы — 48%, томаты - 46%,
салаты, зелень и пр. - 6%



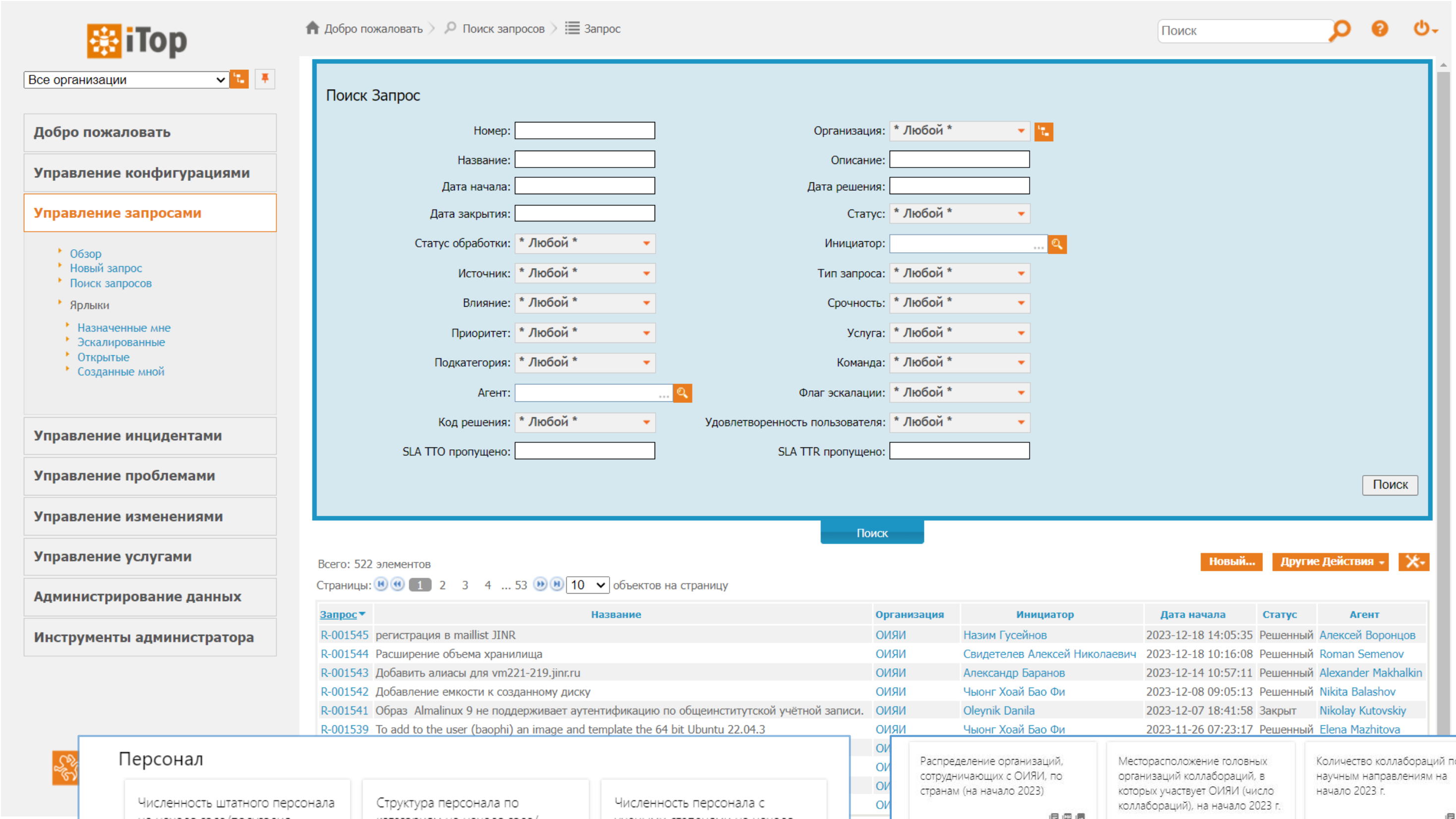


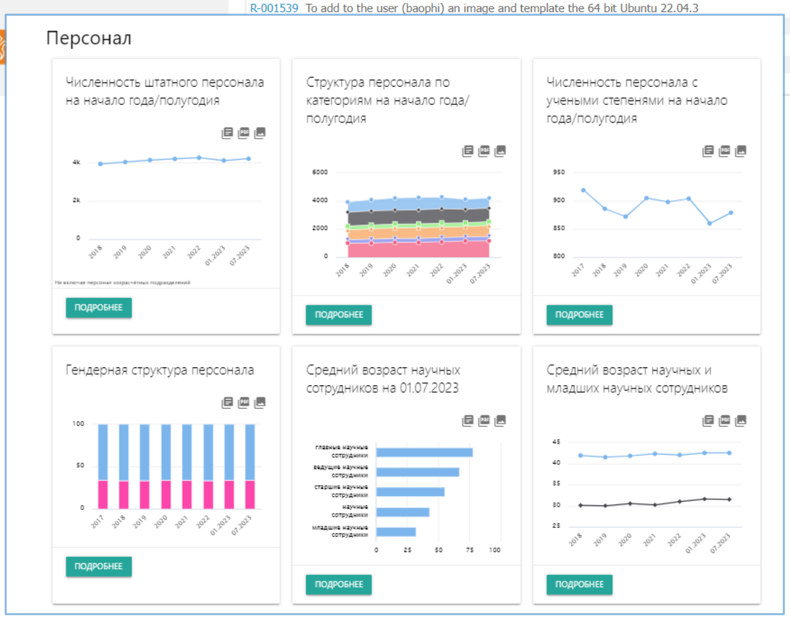


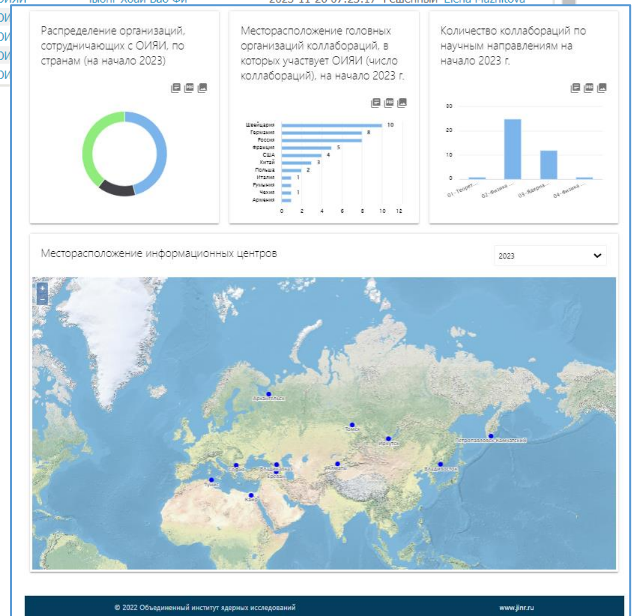
0




Планы: продолжение исследований в области использования современных средств автоматизации и машинного обучения для нужд сельского хозяйства

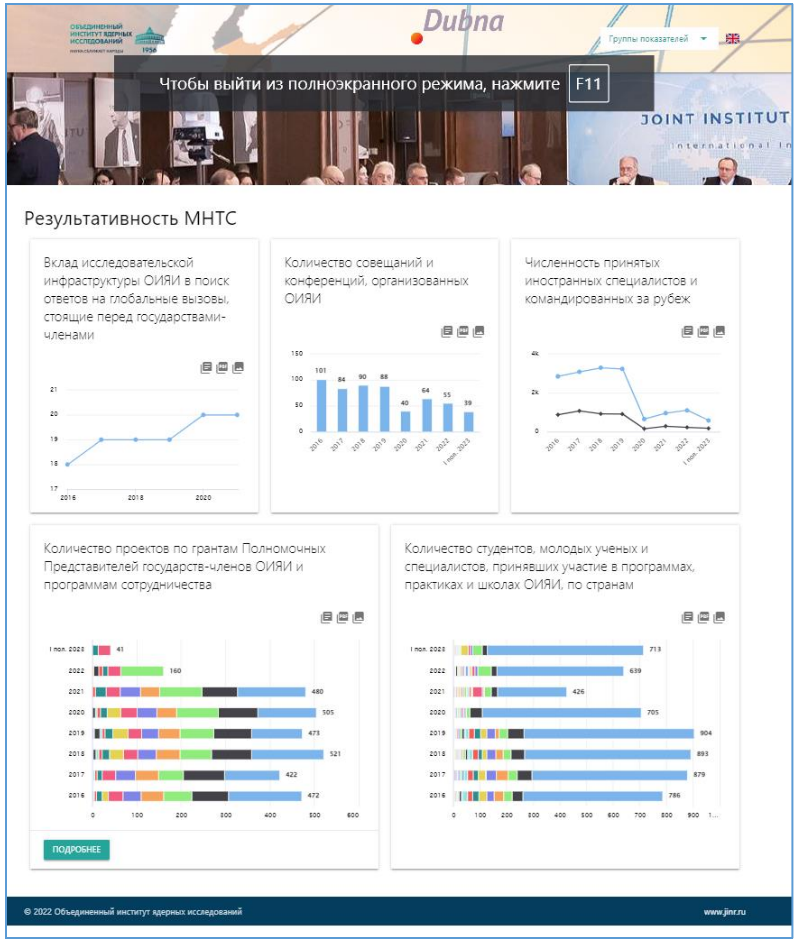
helpdesk.jinr.ru







monitoring.jinr.ru



19 публикаций за последние 5 лет.

- Advanced Technologies and Artificial Intelligence in Agriculture
  Alexander Uzhinskiy, AppliedMath, ISSN:2673-9909, Изд:MDPI, 2023
- Искусственный интеллект в сельском хозяйстве
  Александр Ужинский, Открытые системы, ISSN:1028-7493, Изд:Открытые системы, 3, 20-23, 2023
- Central Russia heavy metal contamination model based on satellite imagery and machine learning
  Uzhinskiy A., Vergel K, Computer Optics, ISSN:0134-2452, eISSN:2412-6179, Изд:Institution of Russian Academy of Sciences, Image Processing Systems Institute of RAS, 1, 47, 137-151, 2023
- Искусственный интеллект против болезней растений
  А.В. Ужинский, Открытые системы, ISSN:1028-7493, Изд:Открытые системы, 3, 29-31, 2022
- R-CCN Plant Diseases Detector Using Triples Loss and Siamese Neural Networks
  M. Gerasimchuk, A. Uzhinskiy, Physics of Particles and Nuclei Letters, ISSN:1547-4771, eISSN:1531-8567, Изд:MAIK Nauka/Interperiodica distributed exclusively by Springer Science+Business Media LLC., 19, 5, 570-573, 2022
- One-shot learning with triplet loss for vegetation classification tasks
  A.V. Uzhinskiy, G.A. Ososkov, P.V. Goncharov, A.V. Nechaevskiy, A.A. Smetanin, Computer Optics, ISSN:ISSN 0134-2452, eISSN:ISSN 2412-6179, Изд:Институт систем обработки изображений РАН, 45, 4, 608-614, 2021
- Интеллектуальная платформа экологического мониторинга
  Ужинский А.В., Открытые системы, ISSN:1028-7493, Изд:Открытые системы, 2, 2021
- Intelligent Environmental Monitoring Platform
  A. Uzhinskiy, CEUR Workshop Proceedings, ISSN:1613-0073, Изд:CEUR Workshop Proceedings, 3041, 424-428, 2021
- Deep learning methods for the plant disease detection platform
  Artem Smetanin, Alexander Uzhinskiy, Gennady Ososkov, Pavel Goncharov, and Andrey Nechaevskiy, AIP Conference Proceedings, ISSN:0094-243X, eISSN:1551-7616, Изд:American Institute of Physics, 2377, 060006, 2021

Review >20

Sensors, Plants, Mathematics, JMSE, Agronomy, Remote Sensing, HELIYON, etc

16 выступлений за последние 5 лет.

- 10. 34th Task Force Meeting of the ICP Vegetation, ICP Vegetation, ,
  PREDICTION OF AIR POLLUTION BY POTENTIALLY TOXIC ELEMENTS BY COMBINING SATELLITE IMAGERY, MOSS BIOMONITORING DATA AND MACHINE LEARNING: LIMITATION AND PERSPECTIVES, Uzhinskiy A., 2021
  11. 9th International Conference "Distributed Computing and Grid Technologies in Science and Education" (GRID`2021), JINR, Dubna, Russia
  Intelligent environmental monitoring platform, Uzhinskiy A., 2021
- The 35th ICP Vegetation Task Force Meeting, virtual meeting 21st - 23rd February 2022:
  - Moss Survey Data Management System
  - Central Russia heavy metal contamination model based on satellite imagery and machine learning
- The 6th International Workshop on Deep Learning in Computational Physics, Dubna July 6-8, 2022:

  - Google Earth Engine and machine learning for Earth monitoring
- 2nd Online UNEP Research Symposium on Air Pollution “Online Expert Group Meeting on Review and Analysis of Air Pollution Trends in the Asia-Pacific Region, virtual meeting 9.10.2022:

  - Remote sensing and machine learning advances for air quality monitoring
- The 36th ICP Vegetation Task Force Meeting, virtual meeting 13th-15th Feb, 2023:
  - The future of air quality monitoring
- GRID'2023 Distributed Computing and Grid-technologies in Science and Education 3-7 July 2023

  - Artificial Intelligence in Agriculture

- руководство диссертационным клубом ЛИТ ОИЯИ
- руководство проектной деятельность студентов университета Дубна и ИТ-школы ОИЯИ,
- чтение лекций, проведение демонстраций, туториалов и семинаров на научных школах и конференциях в различных университетах и институтах.

Спасибо за внимание!

Автоматизация в сфере продажи товаров и услуг с использованием электронных кассиров и консультантов

Методы и средства искусственного интеллекта и автоматизации для решения научных и прикладных задач

Интеллектуальные системы экологического мониторинга

Искусственный интеллект и организация данных в сельском хозяйстве

Автоматизация процесса продажи товаров и услуг с использованием электронных кассиров

Цифровизация и оптимизация управления данными в производстве картона

…