An intelligent environmental monitoring platform

Alexander Uzhinskiy

Joint Institute for Nuclear Research, Joliot-Curie 6, 141980 Dubna, Russia, auzhinskiy@jinr.ru

auzhinskiy@jinr.ru, Tel.: +79057766865

9th International Conference "Distributed Computing and Grid Technologies in Science and Education" (GRID'2021) 5-9 July 2021

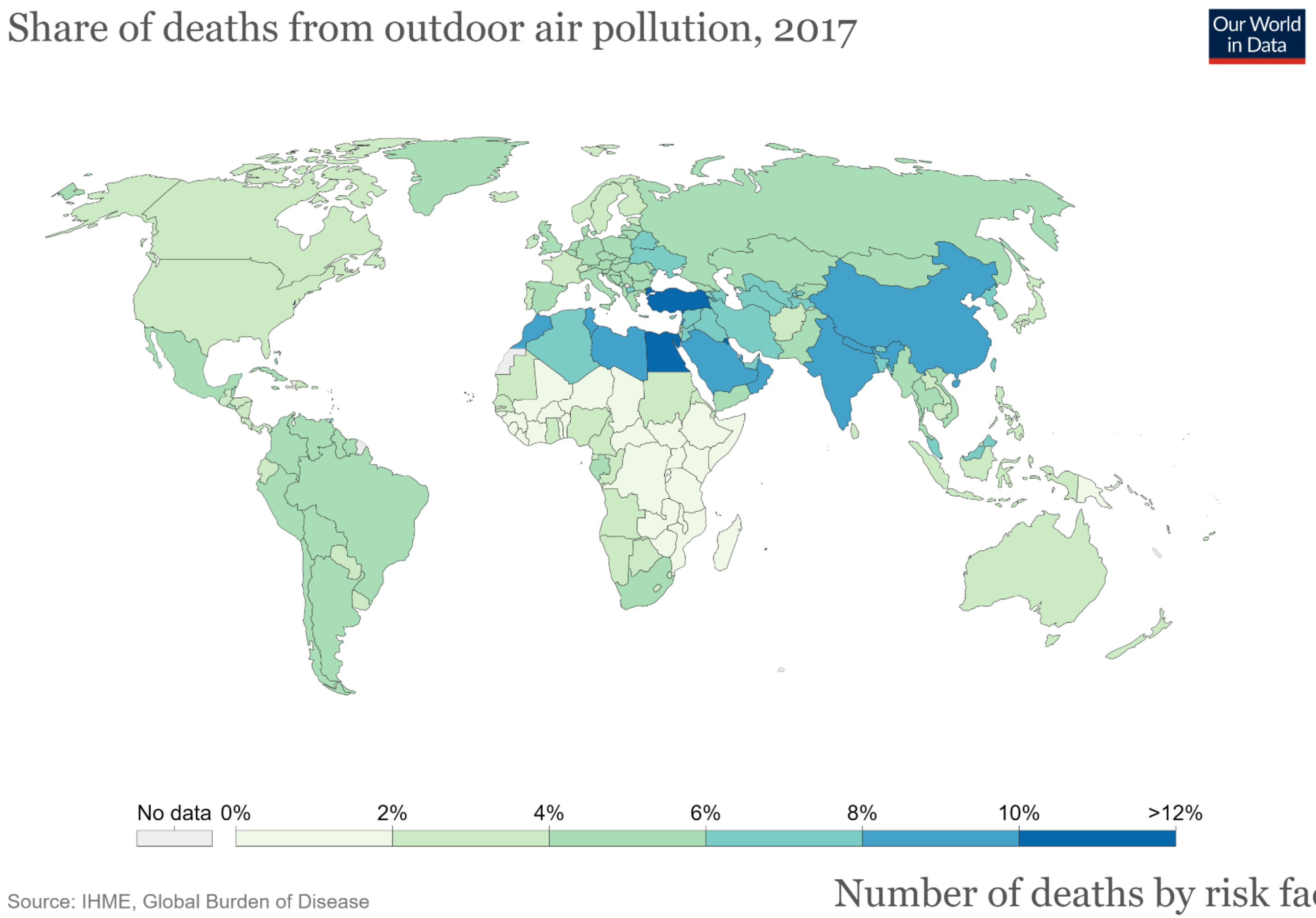
Introduction

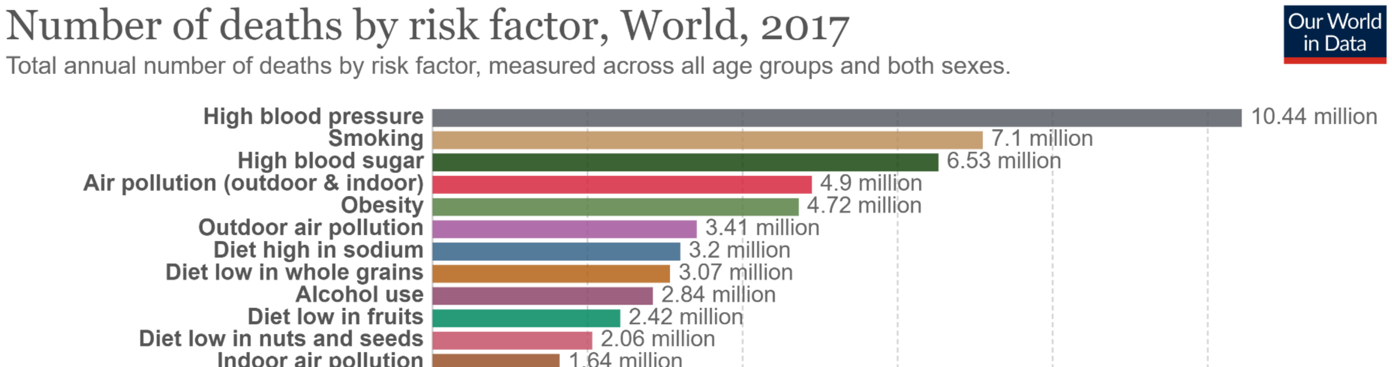
Air pollution has a significant **negative impact** on the various components of ecosystems, **human health**, and ultimately, cause significant **economic damage**.

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).

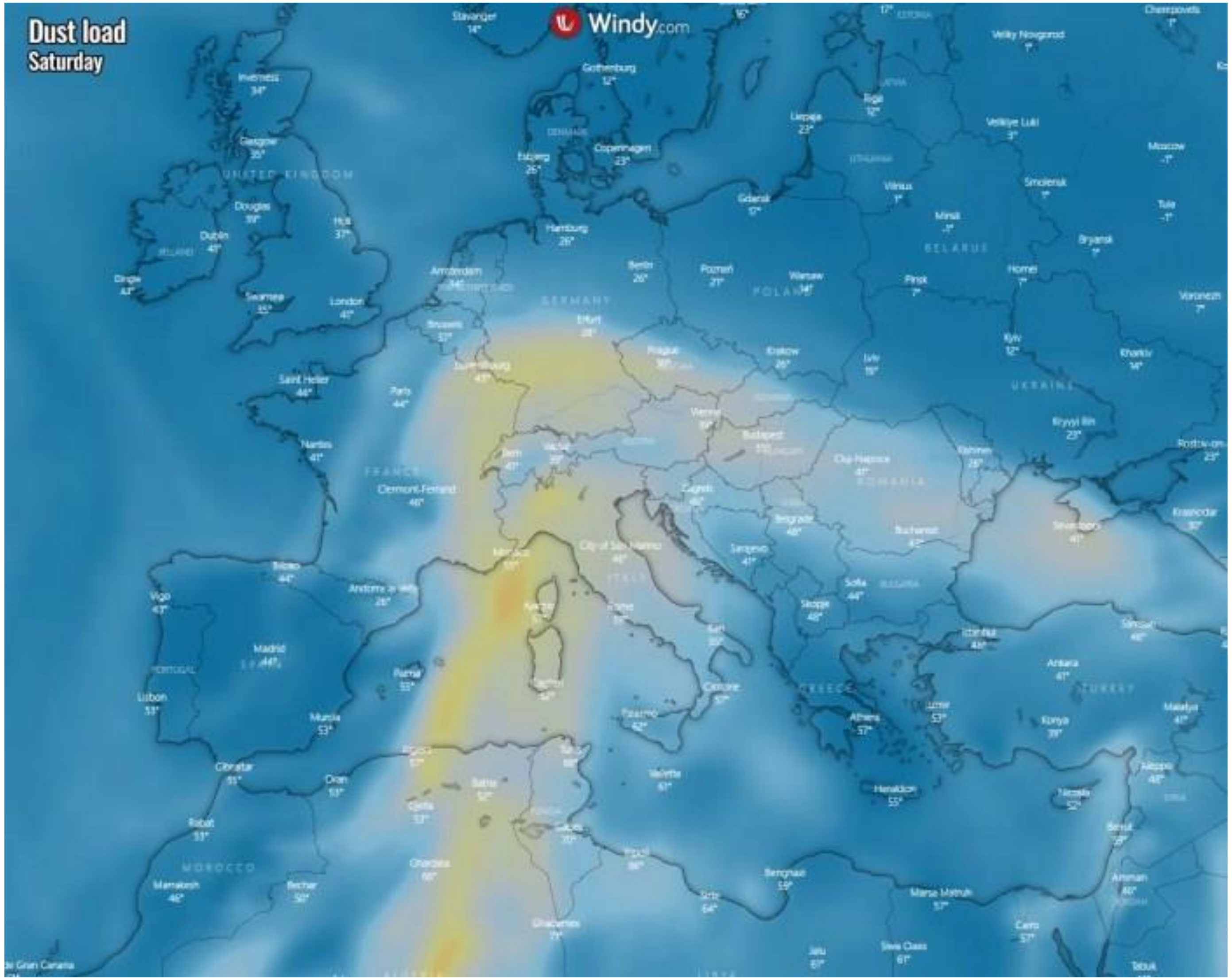
There are an regional and international **environment control programs**. They use different techniques and tools but as a result, they all want to understand **what is the current situation** and how it will evolve.

Air pollution

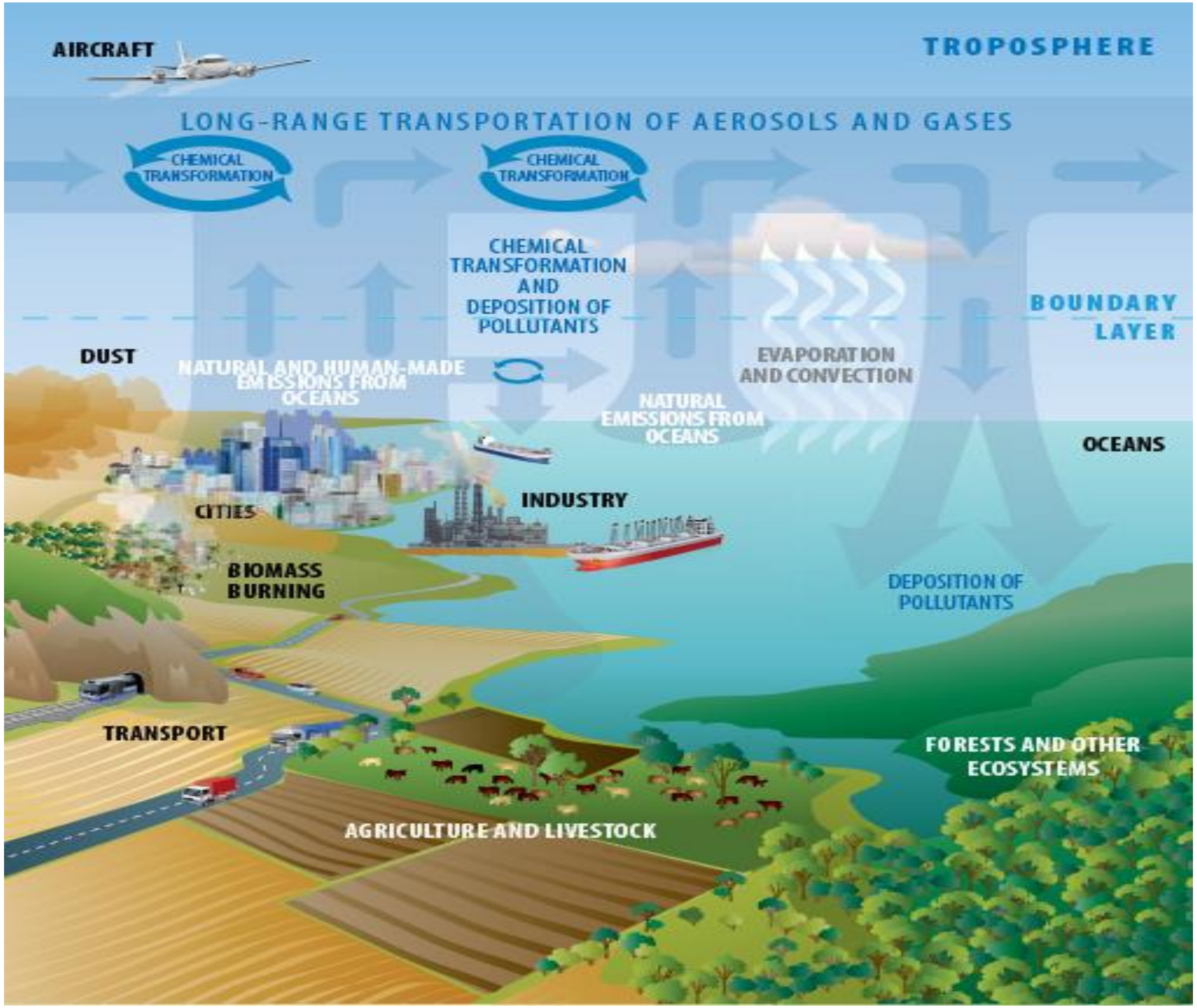




Transboundary Air Pollution



Sources of air pollution



Who cares?

The Air Pollution Control Act of 1955 was the first federal legislation involving air pollution. This Act provided funds for federal research in air pollution. The Clean Air Act of 1963 was the first federal legislation regarding air pollution *control*. It established a federal program within the U.S. Public Health Service and authorized research into techniques for monitoring and controlling air pollution.

The legal authority for federal programs regarding air pollution control is based on the 1990 Clean Air Act Amendments (1990 CAAA). These are the latest in a series of amendments made to the Clean Air Act (CAA). This legislation modified and extended federal legal authority provided by the earlier Clean Air Acts of 1963 and 1970

The Geneva Convention on Long-Range Transboundary Air Pollution establishes a system allowing governments to work together with the aim of protecting health and the environment from air pollution that is liable to affect several countries. The convention was signed in 1979 in Geneva within the framework of the United Nations Economic Commission for Europe (UNECE) and entered into force in 1983.

Local programs are also being implemented that monitor the state of the environment on the scale of countries, regions, cities or enterprises.

Approaches

Generally, studies are based on the data obtained at the sampling sites in manual or automatic mode. The collected material is analyzed using various techniques in the field or in special laboratories. Air quality (AQ) monitoring stations provide information about regulatory air pollutants such as gaseous pollutants, PM, but rarely about heavy metals. To get detailed information samples should be processed in laboratories.

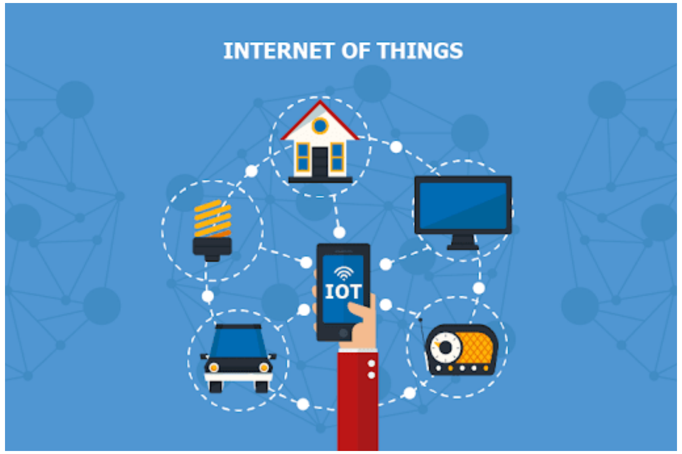
After collection the data are aggrigated and interpreted, and quite often the results are ambiguous and require the involvement of experts.

The level of automation and adoption of information technology in environmental monitoring programs is constantly increasing, although it lags far behind areas where the use of modern technology can lead to rapid economic impact.

Introduction

In last decade, various modern technologies are used in environmental pollution control projects, which make it possible to provide a new level of service, as well as the quality and speed of obtaining results. Now we can talk about intelligent platforms capable of generating new knowledge based on incoming and available data and, in some cases, making decisions that previously required the competence of an expert.

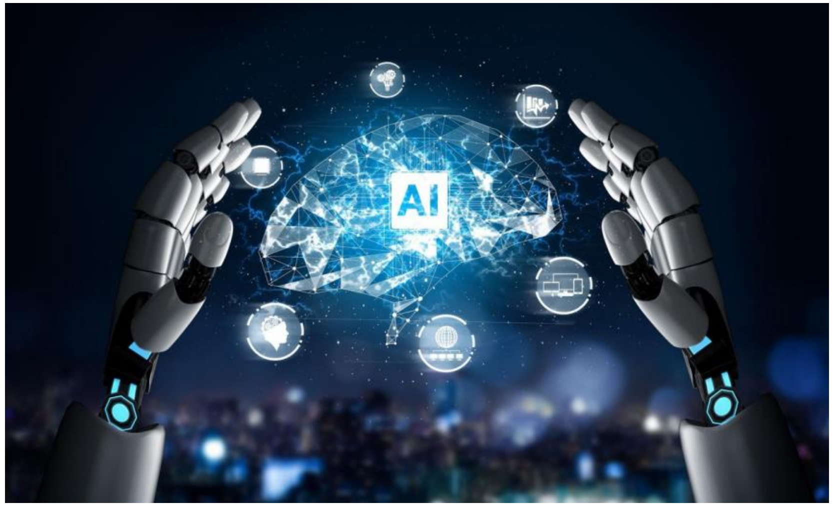
Here are only few examples of such technologies



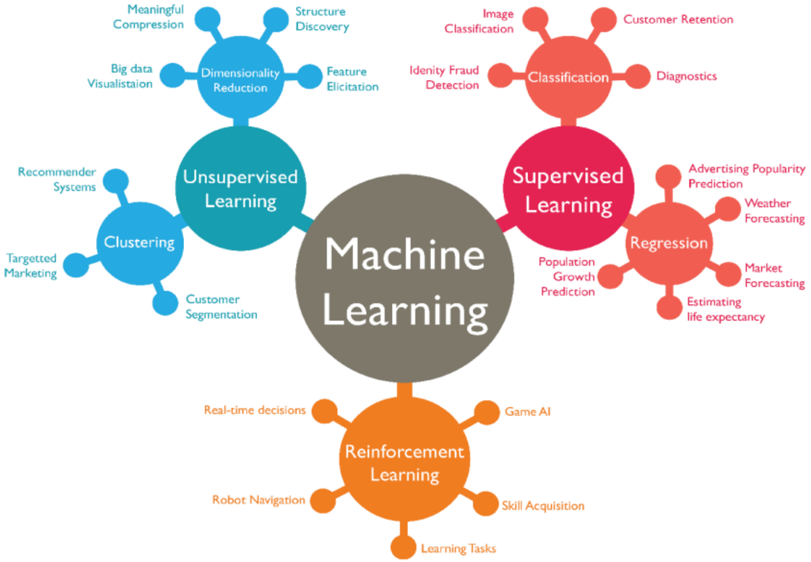
The Internet of things (IoT) specify the principles of connection and exchanging data between physical objects that are embedded with sensors and another objects, programs and systems. Many platforms use IoT technologies to organize sensor network and process environmental monitoring data. That allow to minimize number of errors, automate routine processes, and speed up data gathering routines.

The Big data is a field that treats ways to analyze, systematically extract information from, or otherwise deal with data sets that are too large or complex to be dealt with by traditional data-processing application software. In case of the environmental monitoring the data we have to work with could be both large if we dial with huge sensor network and complex if we dial with sampling sites meta-data.

Introduction



Artificial intelligence (AI) is a wide-ranging branch of computer science concerned with building smart machines capable of performing tasks that typically require human intelligence. In environmental monitoring there are always operations requiring expert opinion. AI technologies could execute primely analysis and save expert time.

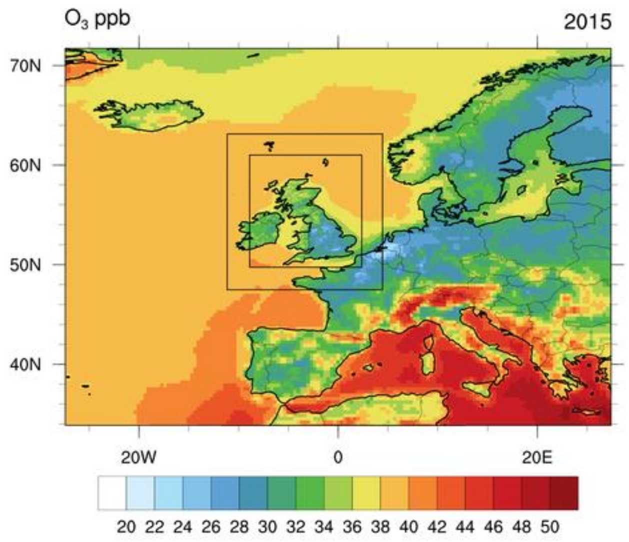


Machine learning is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention. Both classification and prediction tasks of ML are very useful for environmental monitoring.

Robotics, remote sensing, drones etc.

Introduction

Here are few examples of such platforms:



MegaFon offers a platform for environmental monitoring based on the Internet of Things - MegaFon.Ecology.

SimpliVity promotes Omnicube, a universal smart monitoring solution that allows to effectively control various aspects of enterprise operations, including environmental parameters.

Rostec is implementing projects in the field of intelligent environmental monitoring systems.

There are solutions that combine weather stations of various levels into a single infrastructure.

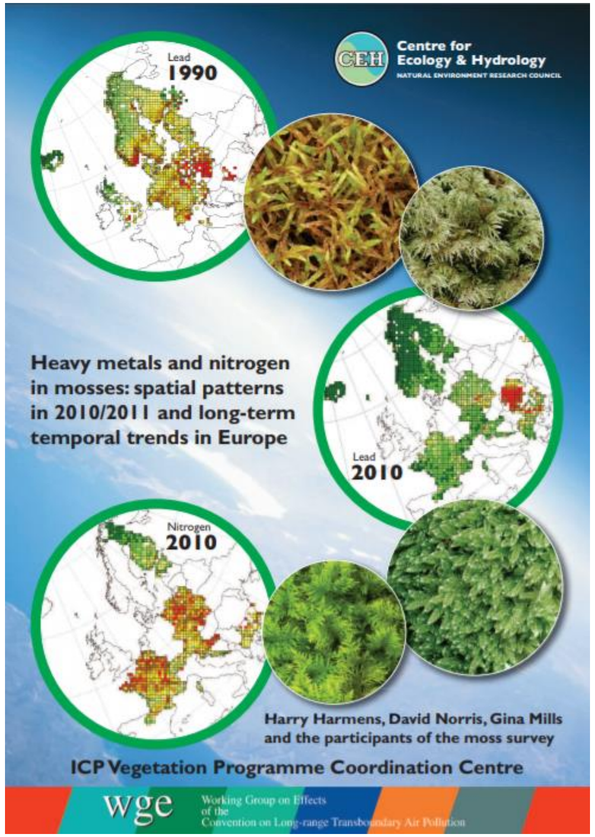
Naturally, there are also foreign projects, mainly based on IoT technologies. For example, the platform of EXM and Libelium companies, designed to improve the efficiency of environmental monitoring, or solutions from the Filippetti Group or Novolyze, providing similar functionality.

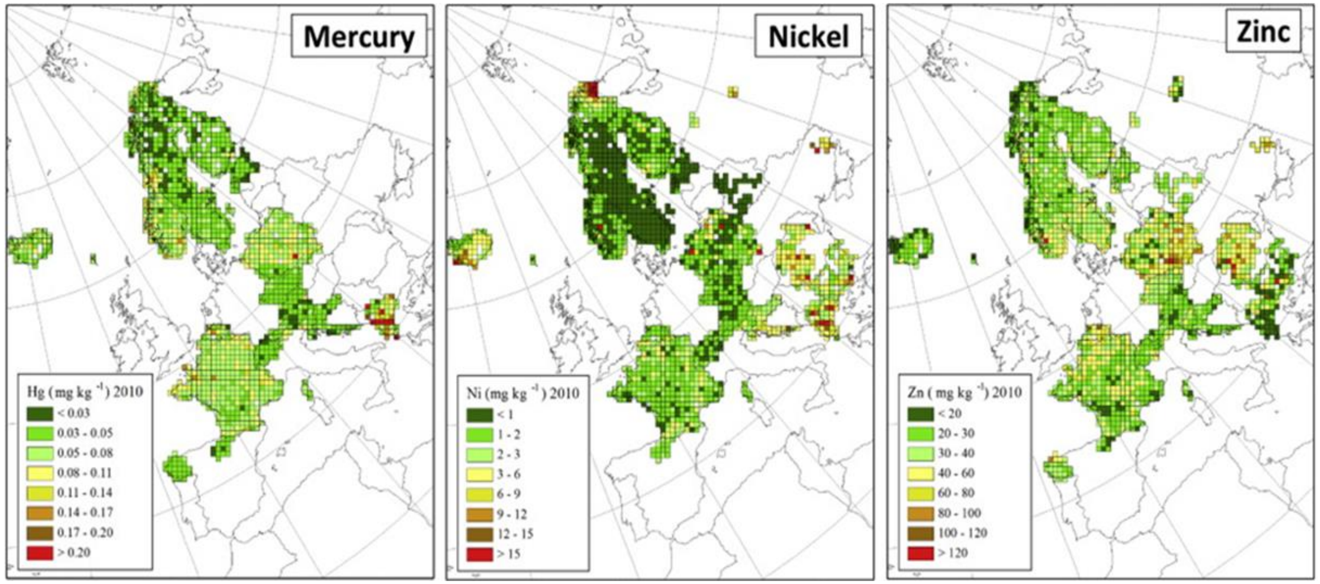
EMEP (*European Monitoring and Evaluation* Programme) use transport models and Air control station data for atmospheric transport and deposition modeling.

Our intelligent environmental monitoring platform

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

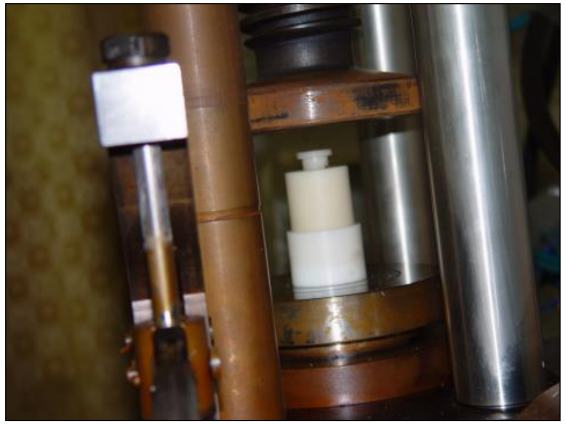
Moss biomonitoring

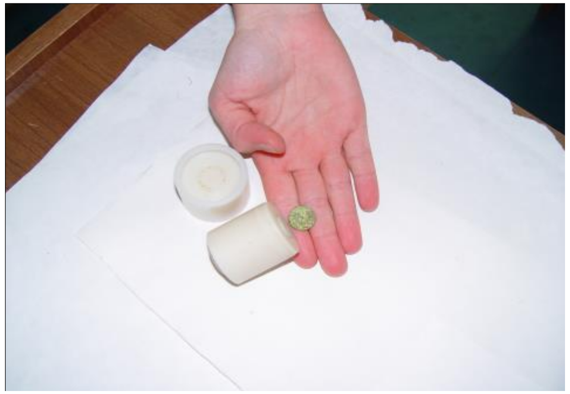
Annual segments

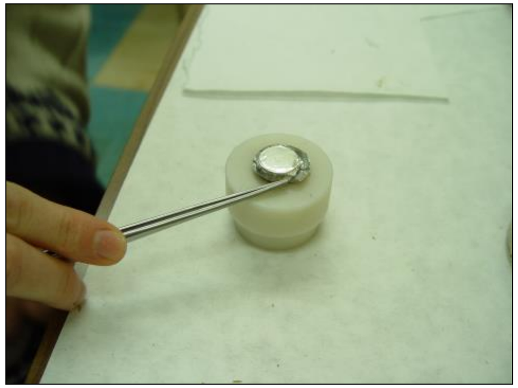
- Heavy Metals (~1970)
- Nitrogen (1980-…)
- Persistent Organic Pollutants (2010)
- Radionuclides (1980 - 2015)
- Microplastic (?)

Sampling

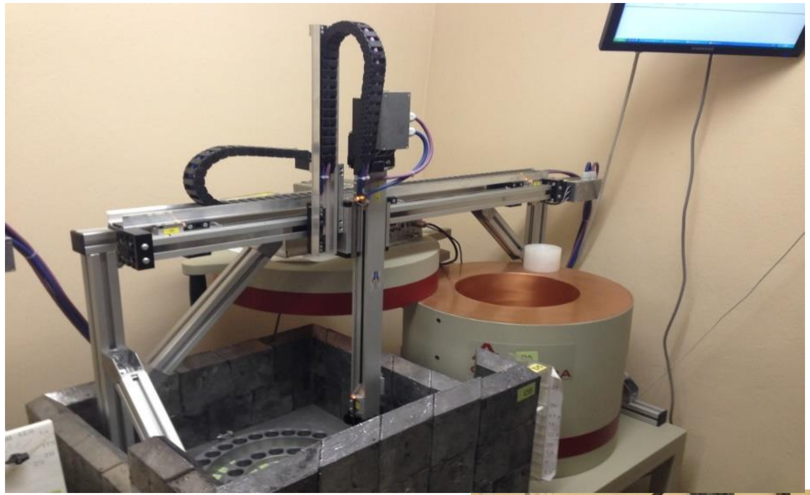
Processing







Analysis



Three sample changers were installed

Each sample changer consists of:

- ❖ two axes liner movement device M202A (DriveSet, Germany)
- ❖ Rotated disk with 40 cells for samples (JINR)
- ❖ Three axes Xemo Motion controller with software and cables (Systec GmbH, Germany)

Abilities

NAA + AAS

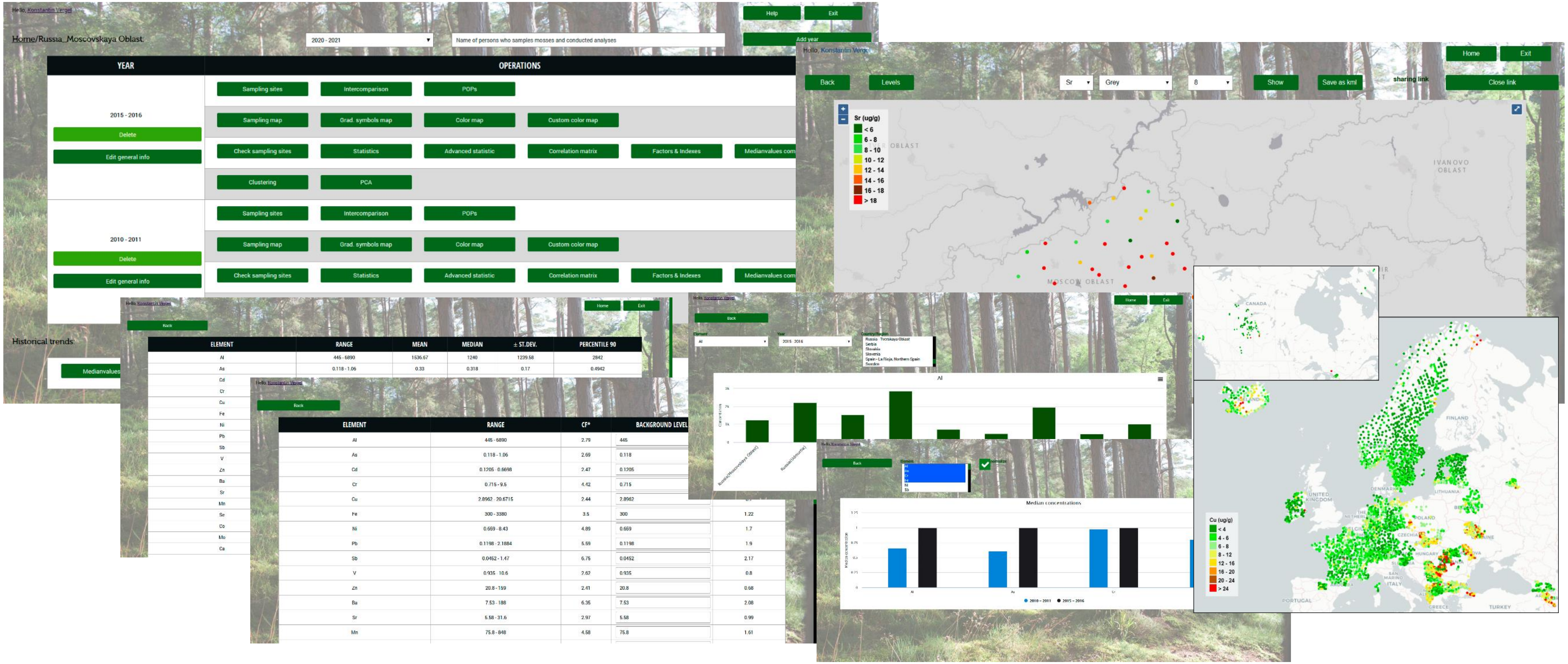
55

H																	He
Li	Be											B	C	N	O	F	Ne
Na	Mg											Al	Si	P	S	Cl	Ar
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
Cs	Ba	La*	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn
Fr	Ra	Ac**											Rf	Db	Sg	Bh	Hs
	*	Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu		
	**	Th	Pa	U	Np	Pu	Am	Cm	Bk	Cf	Es	Fm	Md	No	Lw		

ATLAS: Al As Al Cd Cr Cu Fe Hg Ni Pb S Sb V Zn

DMS

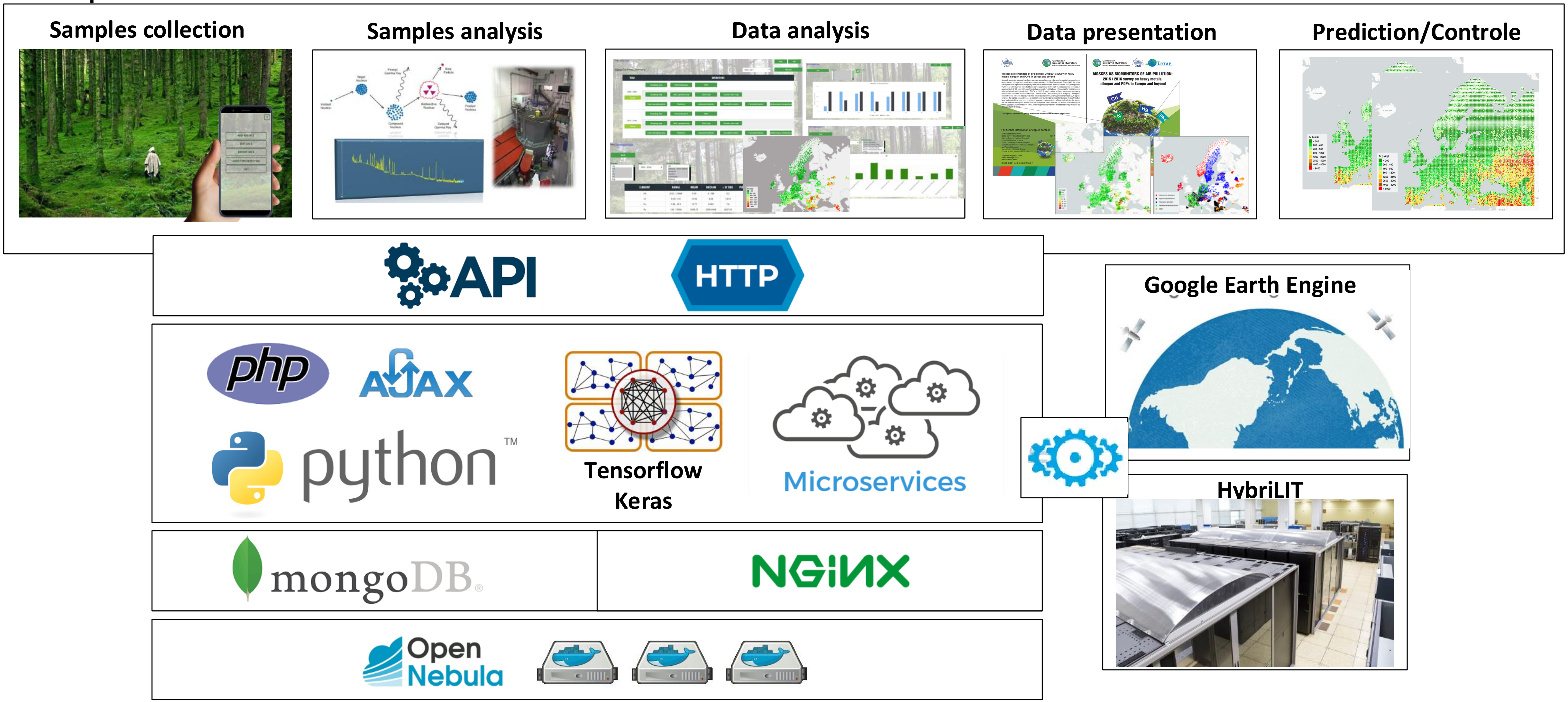
The Data Management System (DMS) of the UNECE ICP Vegetation was developed at the Laboratory of Information Technologies and consists of a set of interconnected services and tools deployed and hosted at the Joint Institute for Nuclear Research (JINR) cloud infrastructure. DMS is intended to provide its 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.

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.

Interesting tasks (Storing of complex data + geodata)

```
{
  "_id": ObjectId("60a124f83c549b478a4b4d9e"),
  "user_id": ObjectId("58981bdf9e7ba441018b4dca"),
  "project_id": "58981c239e7ba443018b53ab",
  "year_id": "5be5331f9e7ba476718b4926",
  "site name": "31. Aiviekste",
  "longitude": 25.9442,
  "latitude": 56.6528,
  "sample date": "2020-09-22",
  "altitude (m)": 83,
  "land cover": "Forests-coniferous",
  "topography": "plain",
  "distance to the nearest projection of the tree canopy (m)": 3,
  "further details": "sunny",
  "moss species": "Pleurozium schreberi",
  "cd": 0.123,
  "cr": 0.3991,
  "cu": 6.3528,
  "fe": 123.8097,
  "ni": 0.2276,
  "pb": 0.6013,
  "v": 0.3835,
  "zn": 24.7611
...
}
```

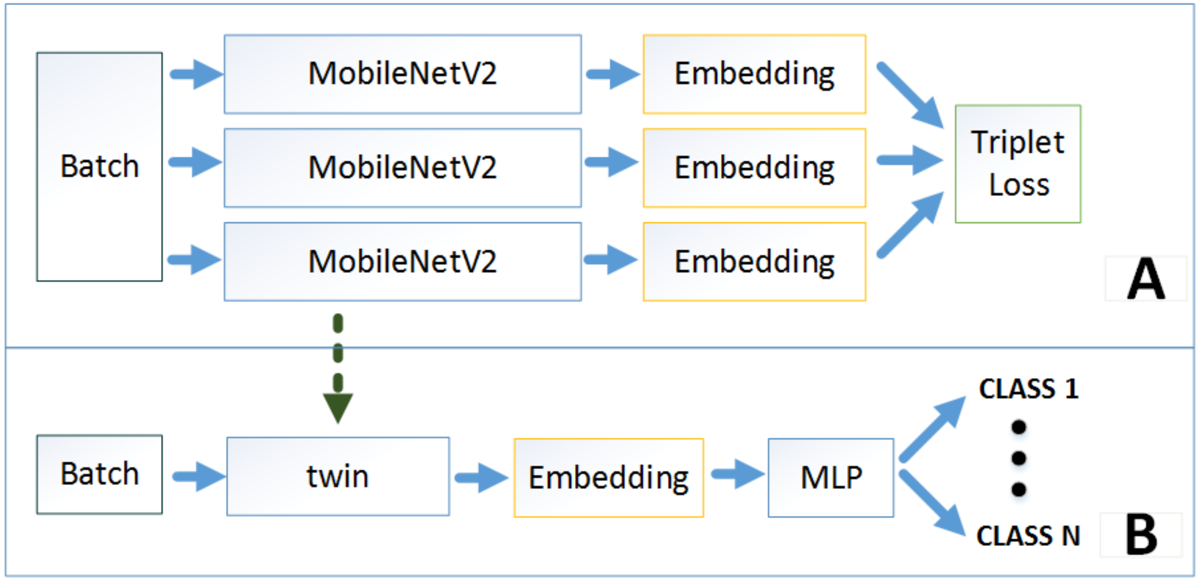

```
db.Address.find({
  "mapLocation": {
    $geoWithin: {
      $box: {
        [ [bottom left coordinates], [top right coordinates] ]

      }
    }
  }
})
```

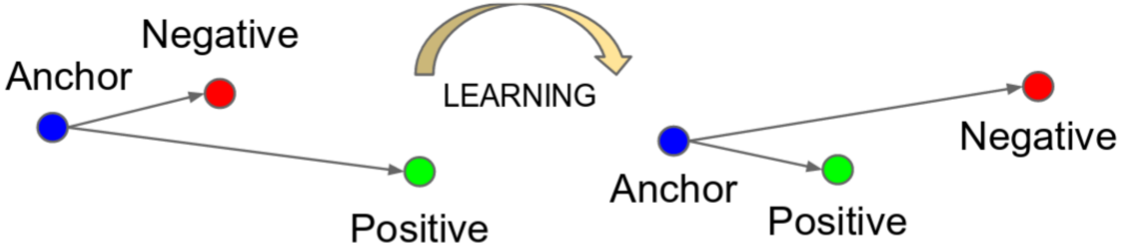
- Different collections (sampling sites, PoP's, Intercomparison etc)
- Tens to hungered parameters
- Geospatial data
- Satelite imagery indexes – hungered thousands to millions objects

Interesting tasks (Moss species classification))

- 599 images
- 5 moss species
- **97.6% classification accuracy**



Siamese networks with triplet loss function



$$\boxed{L = \max(d(a,p) - d(a,n) + margin, 0)}$$

- **"d"** is some kind of function for calculating the distance between vectors, for example, Euclidean distance.

A **"a"** is an anchor image which we want to identify

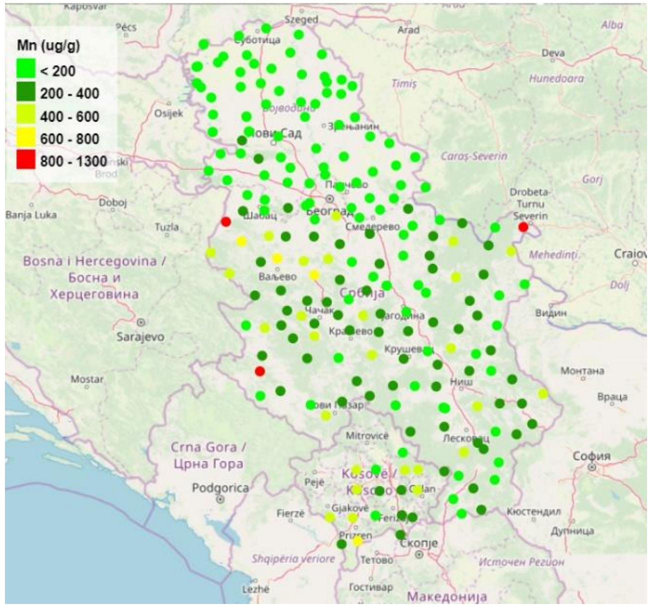
P **"p"** image the same class as anchor

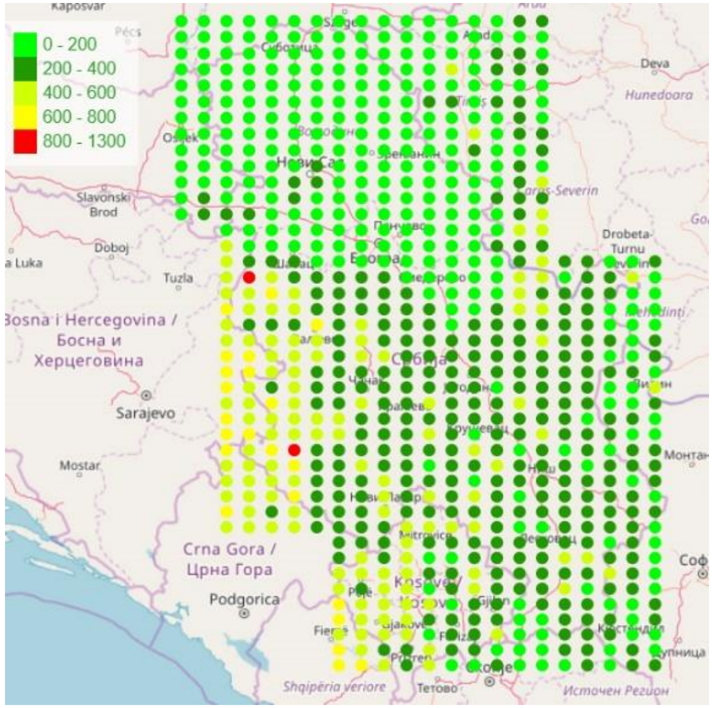
N **"n"** image of another class not matching the anchor

Interesting tasks (prediction)

Regulatory monitoring of air pollution by potentially toxic elements (PTEs) are limited, both spatially and temporally.

Modelling of air pollution can be a good option for overcoming gaps in the data gathering, while moss bag biomonitoring has been recognized as a technique for highly spatially resolved measurements of PTE air pollution.





Modelling allows us to:

- monitor the evaluation of situation when it needed,

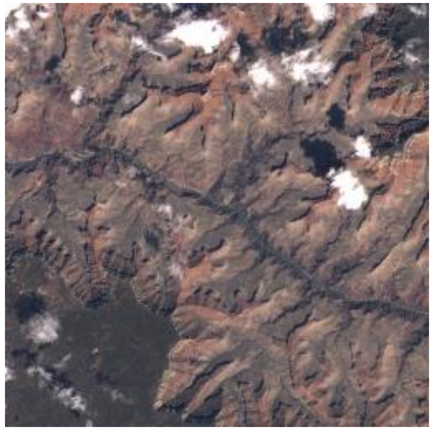
- get detailed information about areas of interests,

- check the situation at the cross border areas,

- partly automate the environment control process.

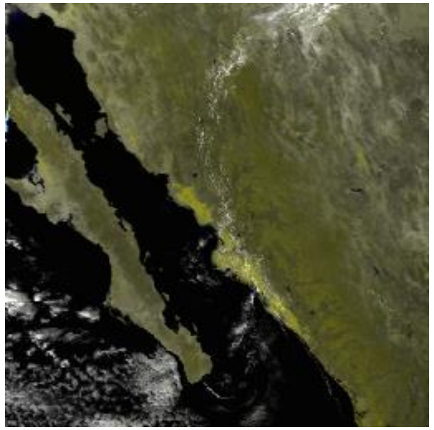
Google Earth Engine

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.

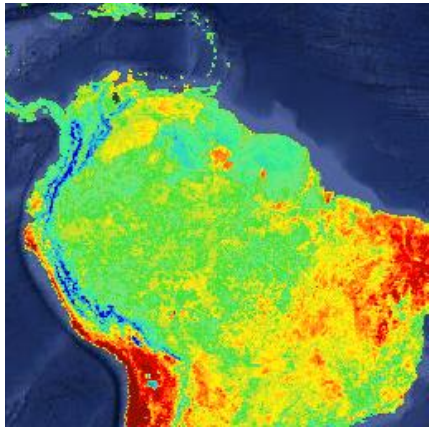


Landsat (15-30m Resolution)

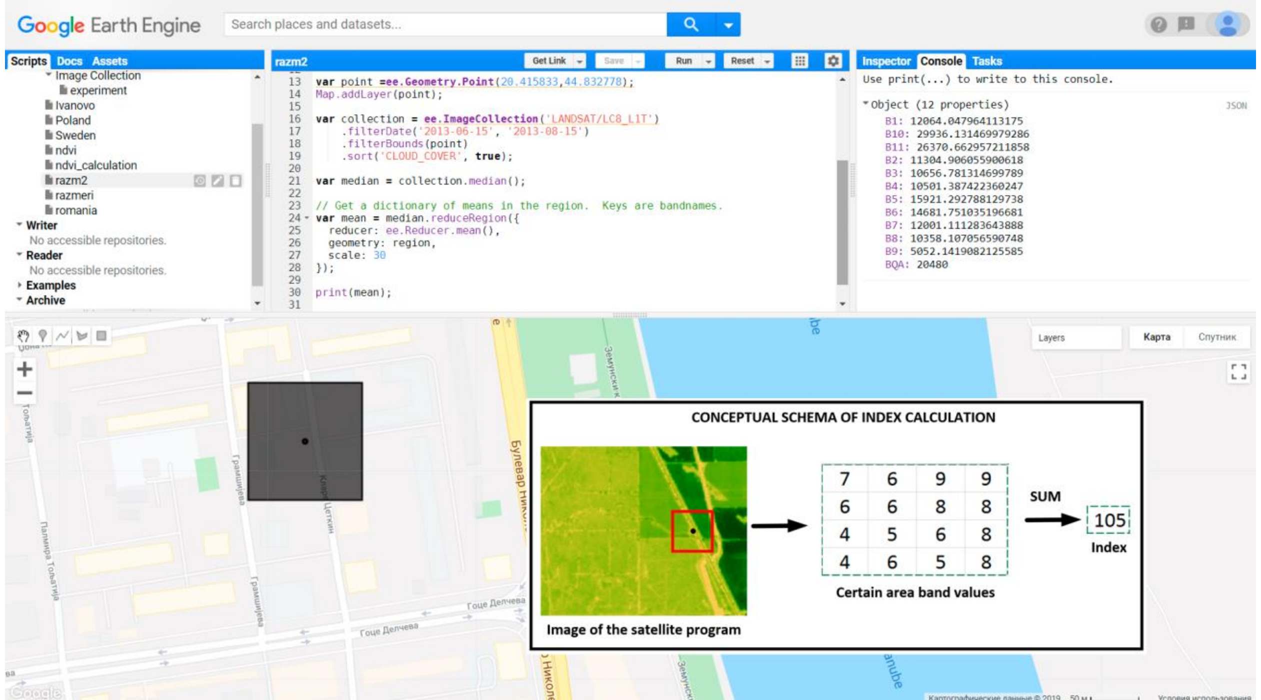
Modis (250-500m Resolution)



Sentinel (250-500m Resolution)



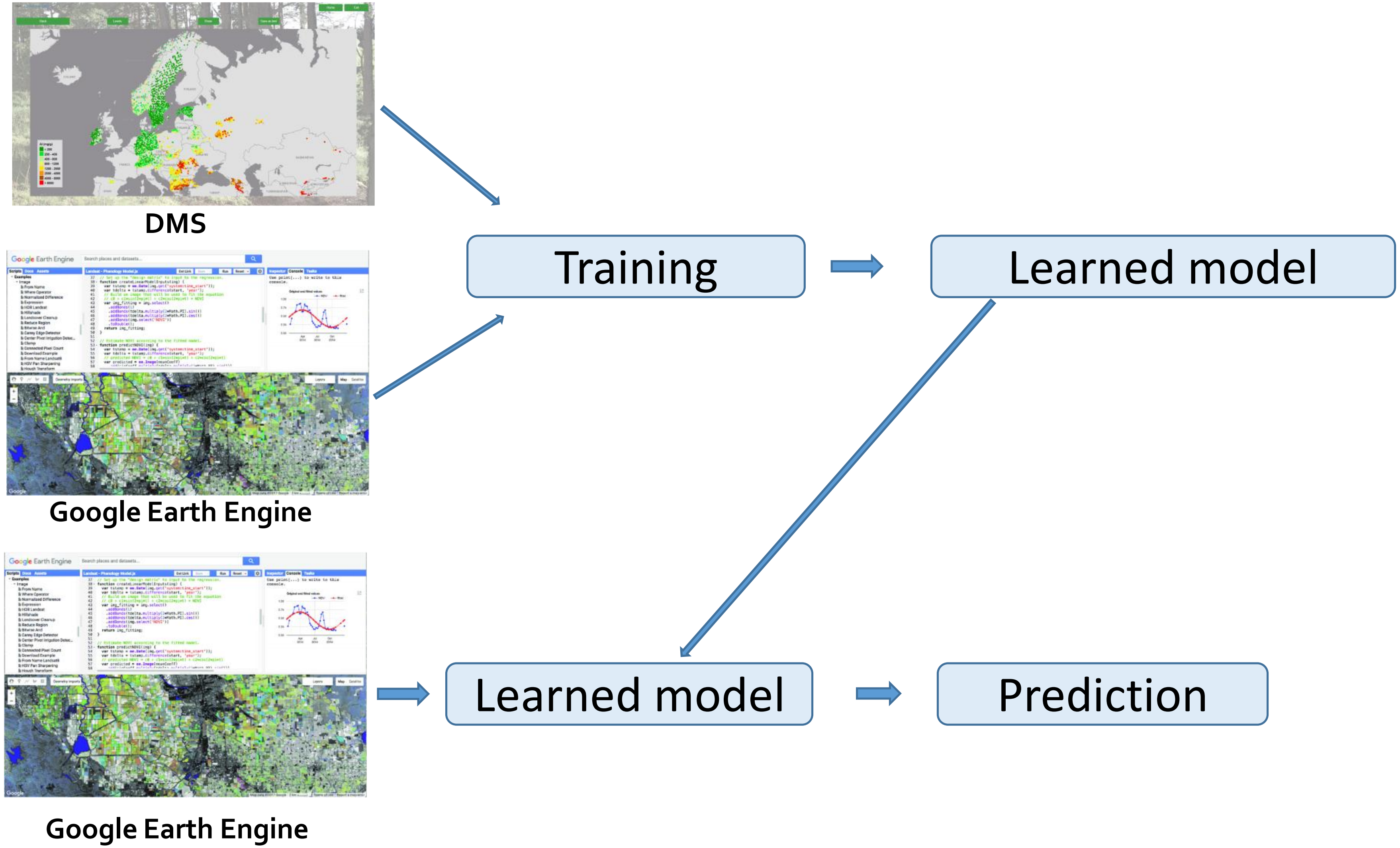
The MOD11A2 V6 average 8-day land surface temperature (LST) in a 1200 x 1200 kilometer grid.



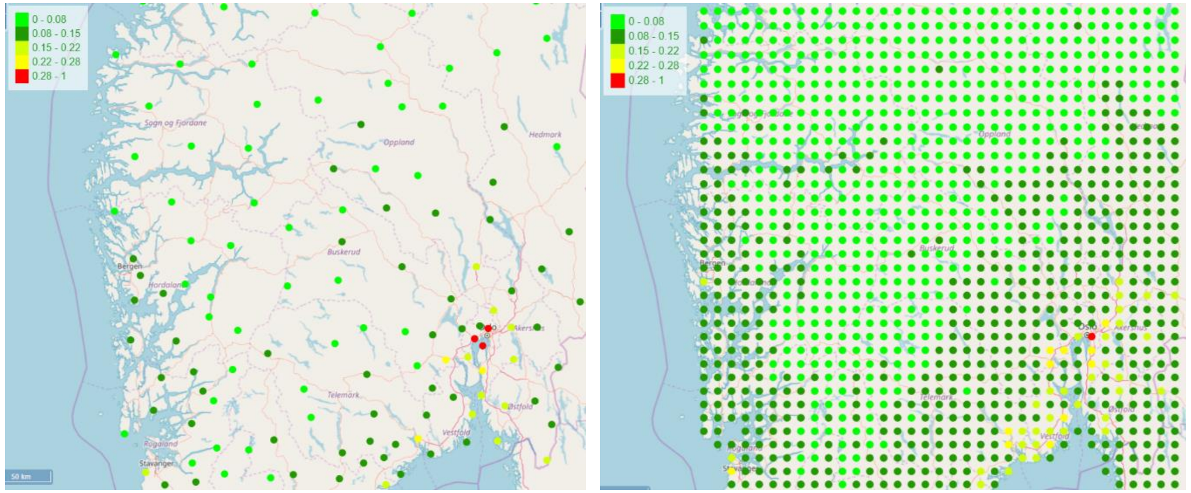
Specify program and time-period to get a collection of images, for example, program – "MODIS/006/MOD09A1" from 2013-06-15 to 2013-08-15 (the period relevant for in situ biomonitoring). Then, define the analyzed area, for example, a square kilometer, with center at the coordinates where sampling was performed. 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.

Schema

We use **satellite imagery data** and the **artificial neural network** to **predict concentration**. The general idea is to use data that we can get from satellite images together with sampling data from DMS to learn NN and then use only data from satellite images to predict concentration.

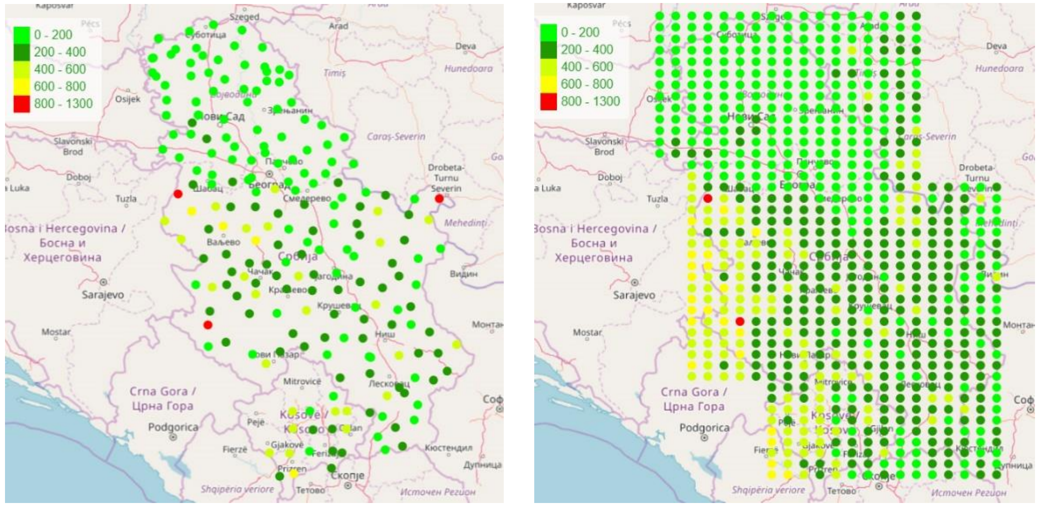


Results on the regional level

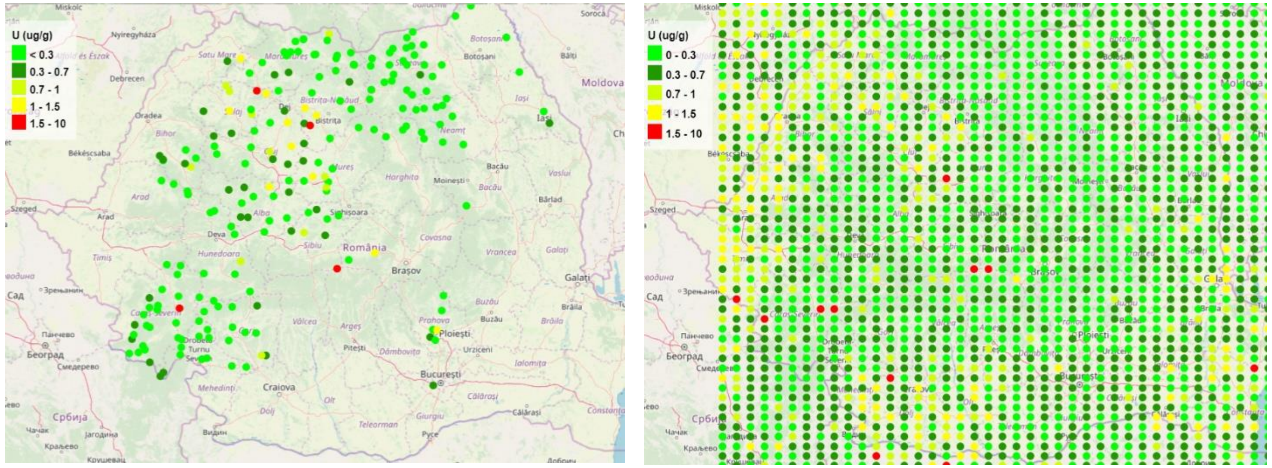


Sb at Norway. Left – real life, right - prediction

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

Last research

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.

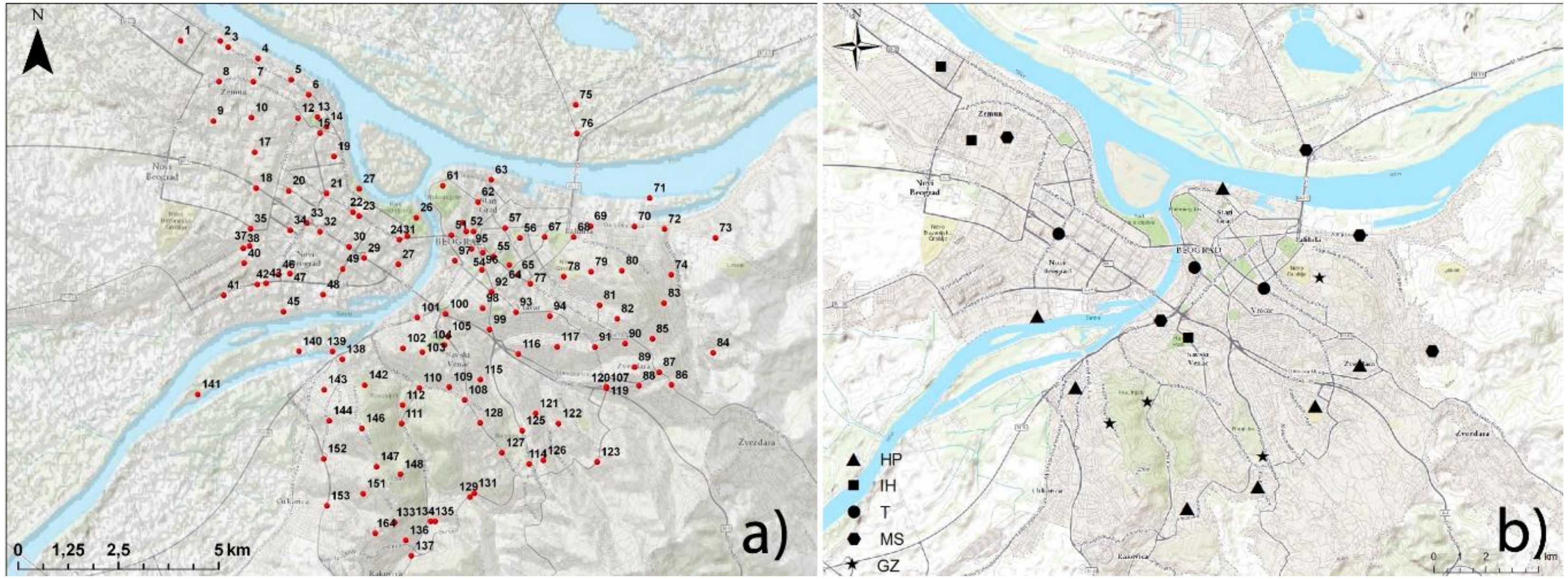


Figure 1. Moss bag biomonitoring across the Belgrade urban area; maps of the sampling sites during two seasons: (a) summer (urban, suburban and green zones) and (b) winter (U–urban sites, GZ–green zones)

Results

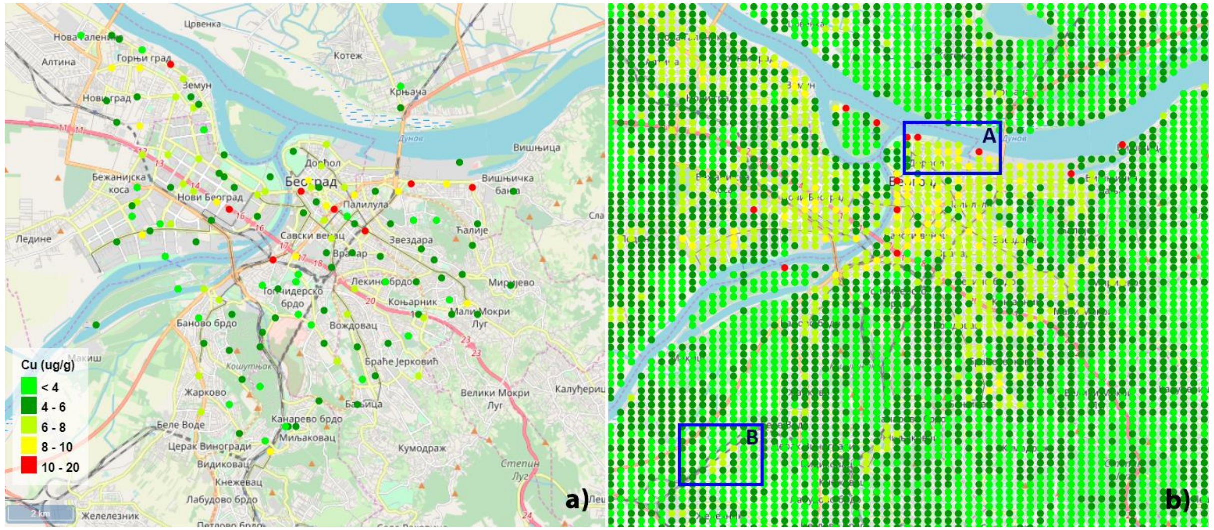


Figure 3. Concentration of Cu in the summer of 2013 (Belgrade): a) real measurements, and b) prediction values; area A represents central part of Old Belgrade with permanently high traffic flow; area B represents a large railway terminal

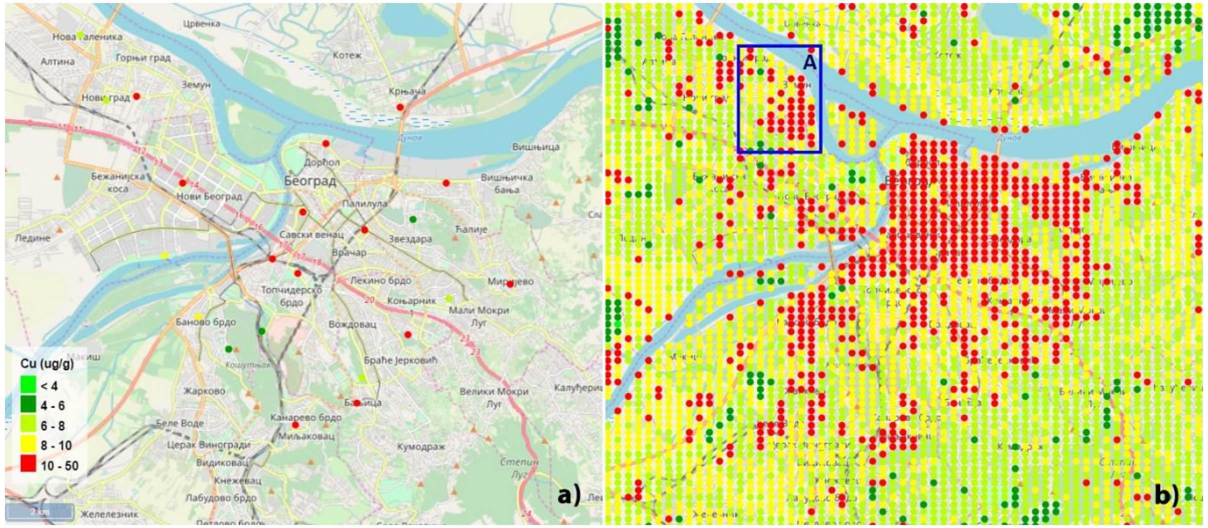


Figure 4. Concentration of Cu in the winter season 2013/2014 (Belgrade): a) real measurements, and b) prediction values; area A represents an old city core highly polluted in winter season

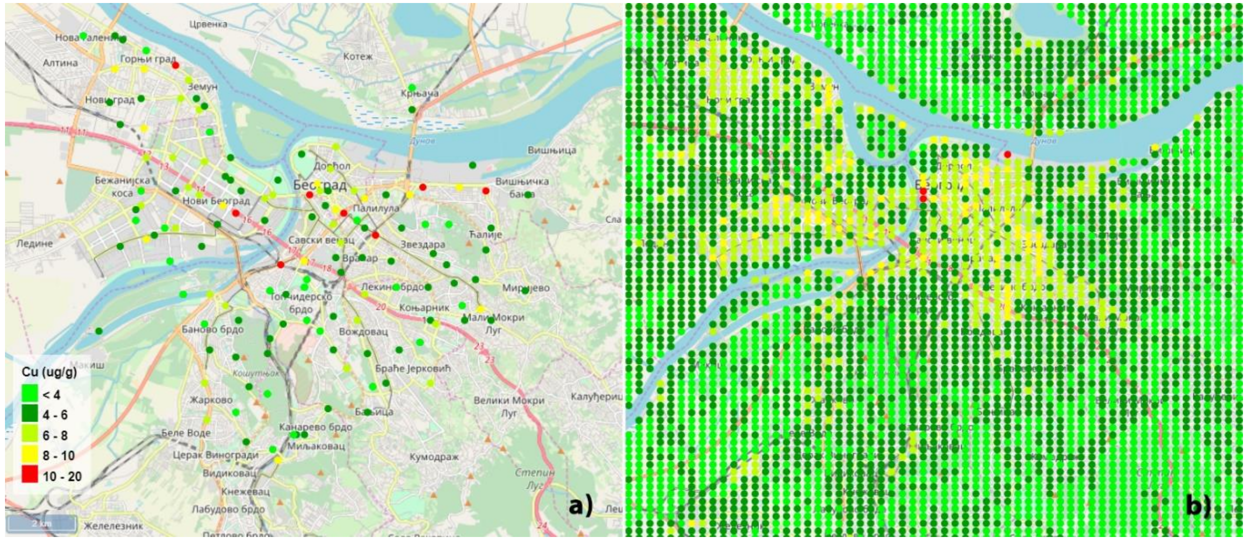
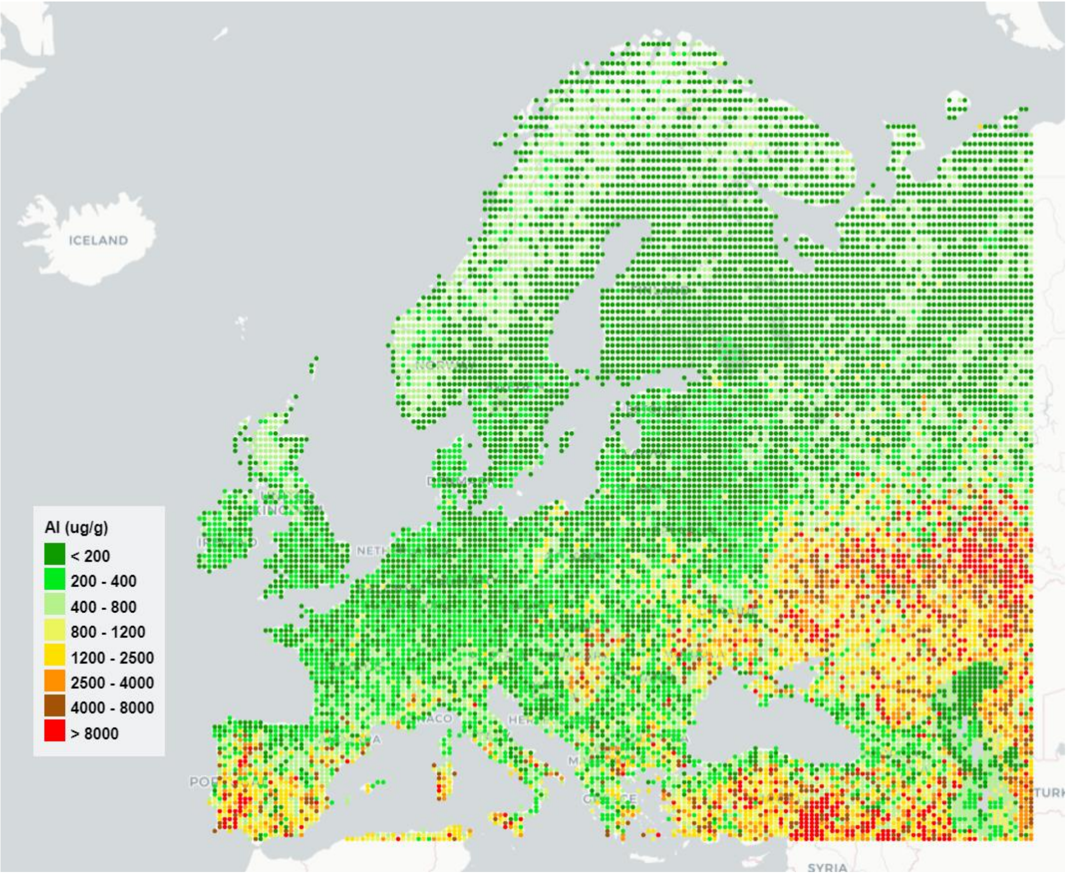
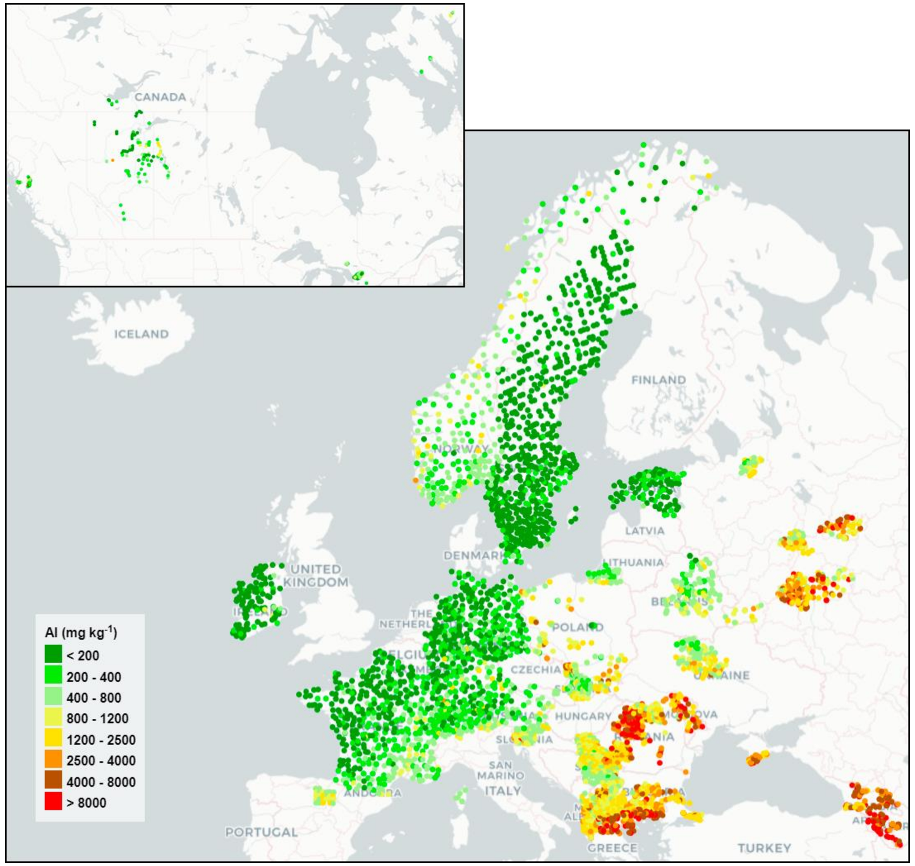


Figure 5. Concentration of Cu in Belgrade: a) biomonitoring measurements in the summer of 2013, and b) prediction for 2018

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.

Conclusion

The obtained results motivate few potential projects to consider our platform as a solution for their tasks.

The platform will not only improve current functionality but also provide new abilities.

- One of the major tasks is the automation of the environmental monitoring process based on modeling.

- Integration of the data about PM and gaseous pollutants from air quality monitoring stations to the platform is considered.

- We are working on the mechanisms of collecting and importing to the platform data about citizens' health. That will allow comparing contamination levels and human diseases in some areas.

Thank you for your attention!