Машинное обучение в прикладных задачах, решаемых в Лаборатории информационных технологий им. М.Г. Мещерякова

Александр Ужинский auzhinskiy@jinr.ru

18.10.2023 ОИЯИ

О машинном обучении и его применении к задачам физики высоких энергий. Ососков Г.А. 15.10

Применение машинного обучения в задаче идентификации частиц. Папоян В.В.

1. Распознавание болезней и проблем в развитии растений

2. Контроль загрязнения тяжелыми металлами

Background



Employment in agriculture

(% of total employment, 2005 and 2020)





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.

Where to find data?

https://rshbdigital.ru/startups

Витрина стартапов

Тщательно отобранная и структурированная самая полная база агротехнологических стартапов, которая будет полезна инвесторам, инновационным лидерам и самим стартапам.

Мы оцениваем стартапы по собственной уникальной методике рейтингования, основываясь на глубокой банковской экспертизе





https://t.me/svoe_fermerstvo https://t.me/agrotech_startup https://t.me/fermerznaet

ОБЗОР РЫНКА АГРОТЕХНОЛОГИЙ

...

https://data-economy.ru/reports



Эффективные отечественные практики на базе технологий искусственного интеллекта в сельском хозяйстве

01.06.2023

(ВКЛЮЧАЯ FOODTECH) ПО ИТОГАМ 2022 ГОДА



РоссельхозБанк

From Top to bottom – soil & crop management

https://www.aerospace-agro.com/

We study territories and cultures using information and aerospace technologies in order to increase the yield and revenue of the enterprise.



Commercial programs

https://onesoil.ai/

Helps to remotely monitor the condition of land, crops, increase yields, reduce the cost of seeds and fertilizers



Open programs. 10 -15 m resolution

"Teleagronom" - InnoGeoTech, Innopolis University, etc TERRATECH presented a complex of geoservices "DigitalEarth"

Geoscan

Geoscan is a Russian group of companies engaged in the development and production of unmanned aerial vehicles (UAS), as well as the development of software for photogrammetric data processing and 3D visualization





Sputnik AgroGeoinformation system focused on solving the problems of precision farming.









AssistAgro

AssistAgro is a digital platform for effective agribusiness management



AssistAgro allows one to solve the following tasks:

- forecast of emergence and development of weeds;
- calculation of the density of the condition and the quality of the placement of crop plants;
- forecast of occurrence and development of diseases of agricultural plants;
- forecast of the emergence and development of pests of agricultural plants;
- obtaining recommendations on the optimal dates for agroscouting, pesticide application, tank mix composition and consumption rate;
- yield forecast;
- independent field status control



Sure, there are lots of foreign companies with same functional!

https://www.taranis.com/







https://www.fermata.tech/



And many others





SiberianTiger – field robot (project stopped)







Датчик влажности почвы

Датчик солёности почвы



P.s. Метод обучения с подкреплением. Baseline на keras с YoloV3

Autonomus agro-robots

https://ravenind.com/ omnipower-3200







Case II Magnum (concept)

Cognitive Technologies (https://cognitive.ru/)



ПЕТЕРБУРГСКИ

Harvesting robots



Controlled environment farming

iFarm (https://ifarm.fi/)

GALAD Green Line (http://npcsvet.ru/)





https://agrotechfarm.com









https://city-farmer.ru/),

Healthy garden (https://healthygarden.ru/)



Foreign companies: GreenLabs, Plenty, Pure Harvest Smart Farms iFarm и т.д.

https://farm.bot – plants printer



What about LIT activities?

Specialization



- Choosing the optimal architecture and mechanisms to solve the problem
- Developing a pre-built solution, training models, and implementing a software environment based on proven approaches and methods
- Choosing the equipment and creating devices to support the operation of the models

Palnts disease detection platform



PDDP consists of a set of interconnected services and tools developed, deployed and hosted with the help of the JINR cloud infrastructure. Our web-portal (pdd.jinr.ru – old. Doctorp.org - new), was developed with the Node.js and PHP. It provides not only a web-interface but also the API for third-party services. We have the Pytorch model in the Docker realized as a Tensorflow serving. The model can work at the virtual server, or at a GPU cluster.

We have a mobile App for Android that was developed using the Flutter, so we could build it for iOS, and Windows.

PDDP database

В настоящий момент у нас есть отдельные модели для следующих культур: яблоки, барбарис, вишня, хлопок, пшеница, кукуруза, конопля, огурцы, смородина, виноград, орхидеи, томаты, клубника.



anthurium, barberry balsam mango aglaonema balsam dracaena glaonema balsam dracaena glaonema balsam dracaena glaonema balsam rosemary blackberry polatoes bamboo rosemary blackberry polatoes bamboo monstera chrysanthemum die grape onies cornio www.cabbage.iolet pepper & cornio www.cabbage caclus peonies cornio begonia fern violet pepper & so do to begonia fern violet begonia fern violet begonia fern violet begonia fern violet basil blueberry of the spathiphyllum basil blueberry of the spathiphyllum carrot dieffenbachia marigold carrot thuja petunia avocado coleus lavender dahlias

General model > 65 classes Corpse model > 70 classes Specialized models > 25 classes

Web-portal, telegram-bot, API, app

Побеги, которые больны, вырезают. При дальнейшем

распространении болезни растение уничтожают

0{

"error":0,

⊕{...}.

⊕{...},

∃ { ... }

⊕{...},

8

Ξ{

],



- The Plant Care Service from the Russian Agricultural Bank
- Andijan Institute of Agriculture and Agrotechnics of the Republic of Uzbekistan have utilized our API in their respective applications, which focus on describing plant diseases and pests. 20

Mobile App

Google play: "DoctorP". * Only android version is available



The user has the opportunity to photo a diseased plant and get a prediction for the disease and treatment suggestions. It is possible to download images from device store.

The application requires access to the Internet to work.

We can run the model on the mobile device directly (we have tried it) but models changing too often.

Server side



What's under the hood



We conducted research in the field of optimizing neural network architectures for models.

We tested various state-of-the-art neural network architectures, explored methods for automatically selecting optimal data augmentation policies (autoaugmentation), and experimented with different loss minimization functions (including contrastive, triplet, arcface, cosface, and sphereface).

We identified the optimal approaches for training the models

MobileNet/ ConvNeXt + No augmentation + No quantization + Triplet loss General model -> corpse model -> custom models Validation accuracy > 97%

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 ٠ development of the plant.
- Determination of the weight group of the plant.







Salads classification











Object detection – 1 class Classification – 6 classes

Accuracy > 99%

Set of images

Data collection was carried out in automatic mode using raspberry pi 4, a digital camera with manual focus, and sonar.





A set of images in 5 days:



446 images



More than 150 different classes of objects



1656923471.7066

3.jpg

968.jpg

1656923493.6070

764.jpg

1656923519.4312

954.jpg





742.jpg

1656923550.5826

333.jpg





1656922186.3365 1656922208.2808 576.jpg 664.jpg

1656922225.5775 263.jpg





1656923183.0592 258.jpg







11.jpg

1656923576.7198 1656923598.0400 126.jpg

Final results

Yolo v5s		Yolo v6s		Yolo v7		Yolo + Triplet	
<u>Metrics results</u>		Metrics results		<u>Metrics results</u>		<u>Metrics results</u>	
Recall	0.856	Recall	0.805	Recall	0.901	Pocall	0.005
Precision	0.723	Precision	0.687	Precision	0.845	Precision	0.995
mAP	0.906	mAP	0.854	mAP	0.925	I I CCISION	0.994

Examples



Potatoes disease (Doka-Gennyye Tekhnologii)



Potatoes disease (Doka-Gennyye Tekhnologii)



Hyperspectral cameras

RGB

Automated analysis of the state of plants in the greenhouse complex



Контроль загрязнения тяжелыми металлами

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.

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, PMs, and rarely about heavy metals. To get detailed information samples should be processed in laboratories.

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

In most cases such kind of researches are limited, both spatially and temporally






Modeling - Motivation & Benefits

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



Figure 1. 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

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.



Figure 2. 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 5 in winter season

Machine learning (Supervised learning)



Satellite programs



Differences

Task: Map all change between 2000 – 2010 over a specific forest

Old solution:

- Select Area of Interest
- Find WRS path/row(s) assume an area of 4 scenes

Data Prep:

- Download and store all Landsat during growing season (1 GB / zipped scene)
- Extract and layer stack all Landsat (1.75 GB / scene)
- ~48 scenes per year * 11 years = ~528 scenes, or 924 GB
- Apply atmospheric correction (924 more GB)
- Normalize scenes
- Apply FMASK or similar to remove clouds and shadows
- Create composite and mosaic by year
- Generate vegetation index per year (NDVI and/or NBR)

Analysis:

- Build spatial model in ERDAS to compare year pairs
- Generate change layer
- Classify pixels > certain value as "change"
- Build spatial model to apply year attribute to each pixel
- Repeat for each year pair
- Build spatial model to stack all attributed change layers into a single raster image, with the most recently changed pixel on top
- Apply color ramp visually demonstrating change



The new method took an experienced geospatial programmer about an hour and ~100 lines of code to generate a raster layer showing the extent of landscape change thematically colored by year of change (in this case, yellow changed closer to the year 2000 and red colors changed closer to 2010.

Users can change the AOI very simply and run this same process anywhere in the world, then export the results to a raster TIFF image.

This represents an extremely dramatic improvement in efficiency. In fact, it allows us to ask new questions.

Google Earth Engine

Google Earth Engine combines a multi-petabyte catalog of satellite imagery and geospatial datasets with planetaryscale 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.

The Earth Engine Data Catalog



... and upload your own vectors and rasters

> 200 public datasets

> 5 million images

> 4000 new images every day

> 80 petabytes of data

The Earth Engine Data Catalog

retrieves the cloud fraction using

measurements in the UV/VIS spectral ...

Earth Engine Data Catalog Q Search Ф Русский • Войти Datasets tagged climate in Earth Engine Filter list of datasets Sentinel-5P NRTI CLOUD: Near Sentinel-5P OFFL CH4: Offline Sentinel-5P OFFL CLOUD: Near **Copernicus** Atmosphere ERA5 Daily Aggregates - Latest Climate Reanalysis Produced by **Real-Time Cloud** Methane Real-Time Cloud Monitoring Service (CAMS) Global Near-Real-Time ECMWF / Copernicus Climate NRTI/L3_CLOUD This dataset provides OFFL/L3_CH4 This dataset provides OFFL/L3_CLOUD This dataset provides The Copernicus Atmosphere Monitoring ERA5 is the fifth generation ECMWF near real-time high-resolution imagery of offline high-resolution imagery of offline high-resolution imagery of cloud Service provides the capacity to atmospheric reanalysis of the global climate. Reanalysis combines model cloud parameters. The TROPOMI/S5P methane concentrations. Methane parameters. The TROPOMI/S5P cloud continuously monitor the composition data with observations from across the cloud properties retrieval is based on (CH4) is, after carbon dioxide (CO2), the properties retrieval is based on the of the Earth's atmosphere at global and the OCRA and ROCINN algorithms regional scales. The main global nearworld into a globally complete and most important contributor to the OCRA and ROCINN algorithms currently currently being used in the operational anthropogenically enhanced greenhouse being used in the operational GOME and real-time production system is a data consistent dataset. ERA5 replaces its GOME and GOME-2 products. OCRA effect. Roughly three-quarters of GOME-2 products. OCRA retrieves the assimilation and forecasting suite predecessor, the ERA-Interim reanalysis.

cloud fraction using measurements in

the UV/VIS spectral regions ...

methane emissions are anthropogenic

and as such it is important to continue

providing two 5-day forecasts per day

for aerosols and chemical ...

ERA5 DAILY provides aggregated values

for each day for ...

11

JavaScript code editor https://code.earthengine.google.com/



* Python API too and also Google colab!

Example of map



Cloud free map



Data Types and Geospatial Processing Functions

- Image band math, clip, convolution, neighborhood, selection ...
- Image Collection map, aggregate, filter, mosaic, sort ...
- Feature buffer, centroid, intersection, union, transform ...
- Feature Collection aggregate, filter, flatten, merge, sort ...
- Filter by bounds, within distance, date, day-of-year, metadata ...
- **Reducer** mean, linearRegression, percentile, histogram
- Join simple, inner, outer, inverted ...
- Kernel square, circle, gaussian, sobel, kirsch ...
- Machine Learning CART, random forests, bayes, SVM, kmeans, cobweb ...
- **Projection** transform, translate, scale ...

over 1000 data types and operators, and growing!

Reduce

Aggregate everything in a collection

"Reduction"

Examples

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



Alert System for Algal Bloom

Algae blooms occur when certain kinds of algae grow very quickly, forming patches, or "blooms," in the water. These blooms can be indicators of water degradation and emit powerful toxins that can endanger human and animal health.



Trends.Earth

Through this project, we will develop a cloud-based platform dedicated to mapping land degradation which identifies potential land restoration opportunities at national to regional scale, allowing communities to prioritize areas to protect, manage, and restore in order to achieve land degradation neutrality.



Essential Biodiversity Variables - ScaleUp

We implemented a workflow for species distribution modelling in GEE that includes importing species occurrence data into the GEE platform, selecting and preparing predictor variables, and performing model fitting with spatial or temporal split-block cross-validation techniques.



The determination of different gaseous and PMs based on satellite imagery



- Tongshu Zheng, Michael H. Bergin, Shijia Hu, Joshua Miller, David E. Carlson. (2020). Estimating ground-level PM2.5 using micro-satellite images by a convolutional neural network and random forest approach. Atmospheric Environment, Volume 230, 117451, ISSN 1352-2310, <u>https://doi.org/10.1016/j.atmosenv.2020.117451</u>.
- Ibrahim, S.; Landa, M.; Pešek, O.; Brodský, L.; Halounová, L. Machine Learning-Based Approach Using Open Data to Estimate PM2.5 over Europe. Remote Sens. 2022, 14, 3392. <u>https://doi.org/10.3390/rs14143392</u>
- Elshorbany, Y.F.; Kapper, H.C.; Ziemke, J.R.; Parr, S.A. The Status of Air Quality in the United States During the COVID-19 Pandemic: A Remote Sensing Perspective. Remote Sens. 2021, 13, 369. <u>https://doi.org/10.3390/rs13030369</u> Seham S. Al-Alola, Ibtesam I. Alkadi, Haya M. Alogayell, Soha A. Mohamed, Ismail Y. Ismail, Air quality estimation using remote sensing and GIS-spatial
- technologies along Al-Shamal train pathway, Al-Qurayyat City in Saudi Arabia, Environmental and Sustainability Indicators, Volume 15,

2022, 100184, ISSN 2665-9727, https://doi.org/10.1016/j.indic.2022.100184.



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 **39 countries of Europe and Asia**. Mosses are collected at thousands of sites



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



Examples of distribution maps in old Atlas

Moss biomonitor





Annual segments

Sampling



















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)



Scanning electron microscope images of captured particles on the moss surface and corresponding spectrograms



1 - Fe particle with Mg impurity;
2 - Spherule of pure iron;
3 - Al-Fe cluster particle with impurities of Zn, Cu, and Ti;
4 - Diatomic alga

ICP Vegetation (Past)



The UNECE ICP Vegetation program had a serious drawback related to its **weak adoption of modern informational technologies**. Information on collecting and processing of samples was carried out **manually** or with minimum automation.

Until 2016, data mostly was stored in Excel files. It was aggregated and processed in different packages (ArcGIS, MATLAB, etc.) manually by the coordinator.

Files from respondents were usually passed to the coordinator **by email**. There were **no common standards** in data transfer, storing and processing software.

Such a situation does not meeat the modern standards for quality, effectiveness, and speed of research and demands developing a **single platform** to provides a comprehensive solution for biological monitoring and forecasting tasks

ICP Vegetation (First steps)

The idea was to create a cloud platform for data management to facilitate IT aspects of all biological monitoring stages starting from a choice of collection places and finishing with generation of pollution maps of a particular area or state-of-environment forecast in the long term to more complicated one:

• optimization of the sample collection spatial distribution

- advanced mathematical methods for multi-level intelligent statistical analysis,
- geostatistical analyses,
- atlases and reports creation
- and others.

We was going to move from simple tasks:

- storing and manipulating with data,
- processing of data,
- calculation of basic statistics,
- creation of simple maps
- etc.

The platform now



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.

WHY IT'S USEFULL

Fast verification of data structure and it completeness



Easy way to find human made mistakes



- Nice tool to analyze data. Access to yours data from anywhere. Online processing and results.
- Ability to store historical data and analyze trends

Workflow



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





'Mosses as biomonitors of air pollution: 2015/2016 survey on heavy metals, nitrogen and POPs in Europe and beyond'

Naturally-occurring mosses have been sampled across Europe and beyond to monitor the deposition of heavy metals, nitrogen and persistent organic pollutants (POPs) from the air. Since 1990, the moss survey has been repeated at five-yearly intervals for heavy metals. Since 2005 and 2010, nitrogen and POPs respectively were included too in some countries. In 2015/2016, mosses were collected at approximately 5,100 sites in 34 countries for heavy metals, 1,500 sites in 12 countries for nitrogen and at selected sites in eight countries for POPs. In 2015/16, participation in the moss survey has greatly increased in countries in Eastern Europe, Caucasus and Central Asia (EECCA region). The highest concentrations of heavy metals were often observed in South-Eastern Europe and the EECCA region, whereas the highest concentrations of nitrogen were found in parts of central Europe. In countries that have participated in at least four out of the six surveys, the concentration of lead and cadmium in mosses has declined the most (81% and 64% respectively since 1990) and the concentration of mercury has hardly changed (2% decline since 1995). The nitrogen concentration in mosses has hardly changed too since 2005 (5% decline).

This report is for scientists, policy makers and others with an interest in air pollution.

For further information or copies contact:

Dr Marina Frontasyeva Moss Survey Coordination Centre Joint Institute for Nuclear Research Division of Nuclear Physics Frank Laboratory of Neutron Physics Department of Neutron Activation Analysis and Applied Research Joliot-Curie, 6, Dubna, Russia Dubna 141980, Russian Federation

ISBN: ISBN 978-5-9530-0508-1

Telephone: +7 (496)21 65609 Email: marina@nf.jinr.ru Website: moss.jinr.ru Dr Harry Harmens ICP Vegetation Coordination Centre Centre for Ecology & Hydrology Environment Centre Wales Deiniol Road Bangor Gwynedd LL57 2UW United Kingdom

> Telephone: +44 (0) 1248 374500 Email: hh@ceh.ac.uk Website: icpvegetation.ceh.ac.uk



MOSSES AS BIOMONITORS OF AIR POLLUTION:

2015 / 2016 survey on heavy metals, nitrogen and POPs in Europe and beyond



30

Google Earth Engine

Ivanovo

Poland

Sweden

la razm2

li razmer

li romania

Nrite

Reader

Examples

Archive

li ndvi

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.



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.



Google Earth Engine

Results on the regional level



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





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



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

Urban Level (Belgrade)

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)

Urban Level (Belgrade)



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

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)



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.

High spatial resolution of the SNN model prediction of Sb contamination

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.

Sentinel-5



Sentinel-5 is focused on air quality and composition-climate interaction with the main data products being O_3 , NO_2 , SO_2 , HCHO, CHOCHO and aerosols. Additionally Sentinel-5 will also deliver quality parameters for CO, CH_4 , and stratospheric O_3 with daily global coverage for climate, air quality, and ozone/surface UV applications.

The Sentinel-5 mission is part of the European Earth Observation Programme "Copernicus" which is a coordinated and managed by the European Commission (EC). The space component of the Copernicus observation infrastructure is developed under the aegis of the European Space Agency (ESA).



Sentinel-5



contains modified Copernicus Sentinel data (2019-21), processed by ESA

The Sentinel-5 mission consists of high resolution spectrometer system operating in the ultraviolet to shortwave infrared range with 7 different spectral bands: UV-1 (270-300nm), UV-2 (300-370nm), VIS (370-500nm), NIR-1 (685-710nm), NIR-2 (745-773nm), SWIR-1 (1590-1675nm) and SWIR-3 (2305-2385nm).

Resolution ~ 1114 m, Orbital cycle 16 days.

Sentinel 7 (CO2M - Copernicus Anthropogenic Carbon Dioxide Monitoring) 2025 - 2026



Nitrogen dioxide worldwide 12/03/2019

Commercial programs - www.planet.com



Formula 1 Miami Grand Prix

2019

Daily, less than 3m resolution, 4 spectral channels



Commercial programs - www.planet.com

Now



PlanetScope Always-on Monitoring

~200 Satellites	
3.7 m (3.0 NIIRS) GSD	
8 Spectral Bands	
Not required	



SkySat High-Resolution Tasking

21 Satellites 0.5 m (4.0-5.0 NIIRS) GSD RGB, Pan and NIR Spectral Bands Sub-daily

Tasking



Pelican Very High Resolution

~32 Satellites 0.3 m (>5.5 NIIRs at-nadir) GSD 7 Spectral Bands

Up to 12 revisits/day Tasking



Hyperspectral Broad Spectral Range

2 Satellites	
30 m GSD	
400-2500nm	
(5nm spacing) Spectral Bands	

Tasking Required Tasking

Commercial programs - maxar.com



1999 ~ 3 days, 3m resolution, 3 spectral channels

Commercial programs - maxar.com



Commercial programs - Airbus





Commercial programs - Airbus







puk

The total number of AI publications doubled, growing from 162,444 in 2010 to 334,497 in 2021.





Monitoring is the first step to control, so I look in the future with optimism and believe that Earth will become clearer and safer.

Если хотите попробовать порешать интересные задачи и поучаствовать в проектах: email: auzhiskiy@jinr.ru https://t.me/bigzmey

Спасибо за внимание!

Если осталось время – то будет небольшой туториал!