MANAGEMENT OF ENVIRONMENTAL MONITORING DATA: UNECE ICP VEGETATION CASE

A. Uzhinskiy[a], G. Ososkov, M. Frontasyeva

Joint Institute for Nuclear Research, 6 Joliot-Curie, Dubna, Moscow region, 141980, Russia

E-mail: [a]auzhinskiy@jinr.ru

Air pollution has a significant negative impact on various components of ecosystems, human health and ultimately causes significant economic damage. Air pollution is the fourth largest threat to human health, behind high blood pressure, dietary risks and smoking. The aim of the UNECE International Cooperative Program (ICP) Vegetation in the framework of the United Nations Convention on Long-Range Transboundary Air Pollution (CLRTAP) is to identify the main polluted areas of Europe, produce regional maps and further develop the understanding of long-range transboundary pollution. The program is realized in 39 countries of Europe and Asia. Mosses are collected at thousands of sites.
Here we present the Data Management System (DMS) of the UNECE ICP Vegetation consisting of a set of interconnected services and tools deployed and hosted at the Joint Institute for Nuclear Research cloud infrastructure. DMS provides its participants with a modern unified system of collecting, analyzing and processing biological monitoring data and facilitates IT aspects of all biological monitoring stages, starting from the choice of sampling sites and finishing with the generation of pollution maps of a particular area or a long-term state-of-environment forecast.

Keywords: environmental monitoring, data management, cloud platform, intellectual data processing, UNECE ICP Vegetation, air pollution, mosses, heavy metals, machine learning, neural networks

1. Introduction

The aim of the UNECE International Cooperative Program (ICP) Vegetation in the framework of the United Nations Convention on Long-Range Transboundary Air Pollution (CLRTAP) is to identify the main polluted areas of Europe, produce regional maps, and further develop the understanding of long-range transboundary pollution. The atmospheric deposition study of heavy metals, nitrogen, persistent organic compounds (POPs) and radionuclides is based on the analysis of naturally growing moss samples through surveys carried out every 5 years [1]. Due to the intense activity of the Joint Institute for Nuclear Research (JINR) as a coordinator of moss surveys since 2014, Armenia, Azerbaijan, Georgia, Kazakhstan, Moldova, Mongolia and Vietnam joined the survey for 2015/2016. At present, the UNECE ICP Vegetation program includes 39 countries of Europe and Asia. Mosses are collected at thousands of sites across Europe, and their contents of heavy metals (since 1990), nitrogen (since 2005), POPs (persistent organic compounds, a pilot study in 2010) and radionuclides (since 2015) are determined [2]. A total of 13 elements are reported for the Atlas (As, Cd, Cr, Cu, Fe, Hg, Ni, Pb, V, Zn, Al, Sb, and N). Individual results are reported for a number of sampling sites, as well as minimum, maximum and median concentrations in mg/kg. The data interpretation is based on the Multivariate statistical analysis, the description of sampling sites (MossMet information package) and distribution maps for each element.

The UNECE ICP Vegetation program had a serious drawback related to its weak adoption of modern informational technologies. Information on collecting and processing samples was carried out manually or with minimum automation. Until 2014, data were mostly stored in Excel files, aggregated manually by the coordinator and preprocessed in specific packages like ArcGIS, MATLAB, etc. Examples of distribution maps from the 2010-2011 Atlas are presented in Figure 1. 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 meet the modern standards for quality, effectiveness, and speed of research and demands to develop a unified web platform in order to provide a comprehensive solution for biological monitoring and forecasting tasks. Our project was aimed to develop DMS using modern analytical, statistical, programmatic and organizational methods to provide the ICP Vegetation community with a unified system of gathering, storing, analyzing, processing, sharing and using biological monitoring data.

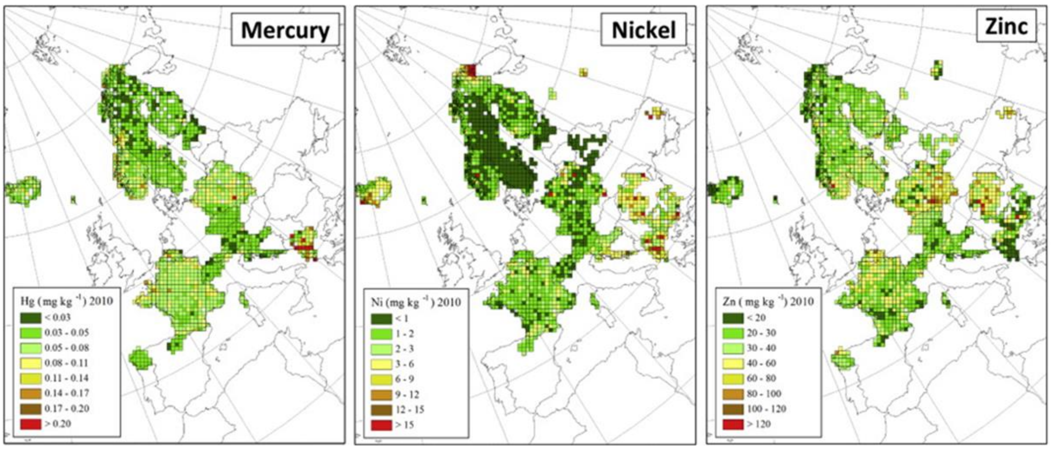

Figure. 1. Examples of distribution maps [2]

2. ICP Vegetation data management system (DMS)

DMS was developed at the Laboratory of Information Technologies of the Joint Institute for Nuclear Research [3]. It consists of a set of interconnected services and tools developed, deployed and hosted in the JINR cloud infrastructure [4]. Such an approach also allows scaling cloud resources up and down according to the service load. When some particular component requires more resources, the cloud can provide them without affecting other components. It increases the efficiency of hardware utilization, as well as reliability and availability of the service itself for end users. The platform provides ICP Vegetation participants with a modern unified system of collecting, analyzing and processing biological monitoring data. More than 6000 sampling sites from 47 regions of different countries are presented at DMS for the 2015-2016 Moss Survey.

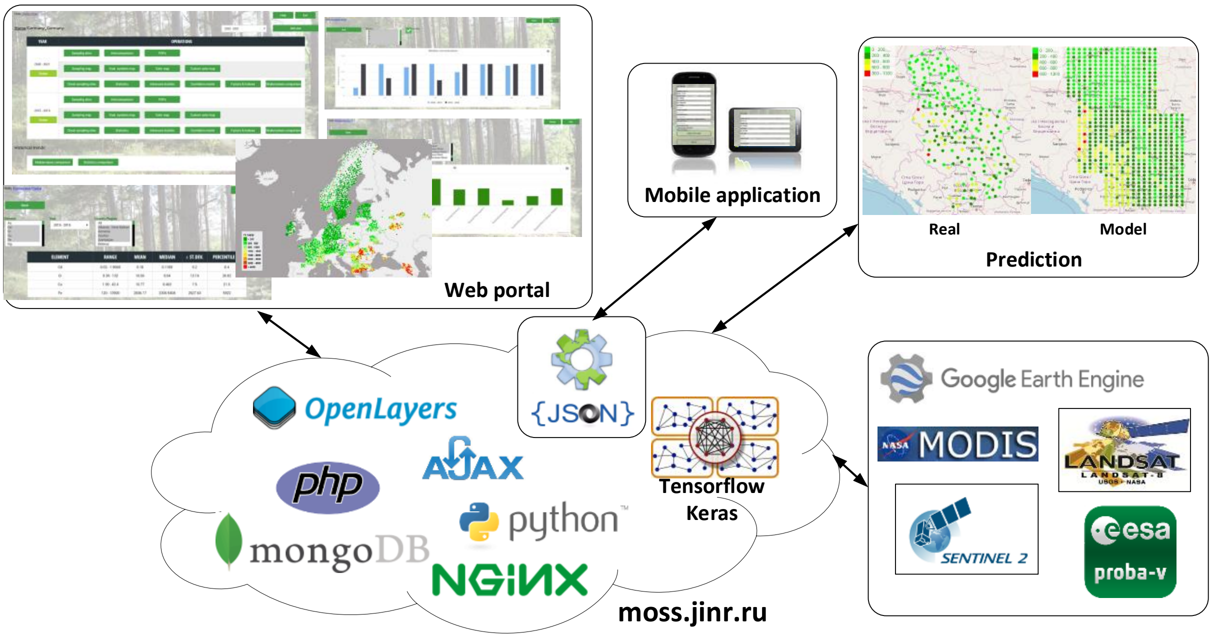

Figure 2. Architecture of DMS and technologies used

Users communicate with DMS through a web portal, a mobile application or web services. We have to deal with four main types of data: sampling sites data, MossMet descriptions, interlaboratory comparison, and POPs. All of them have tens to hundreds of different parameters. To store such variegated data structures we use the non-relational MongoDB. The portal backend was developed with PHP that is also used for calculation of factors, indexes and other statistic parameters. There are two parts of the portal – public and private. General information about the project and the platform is presented in the public part. The private part can be accessed only by authorized contributors and is used for data management and analysis. Data can be imported from and exported to Excel files. Verification of imported data is done automatically and allows finding most of the human-made mistakes. Participants can get some basic and advanced statistic parameters of their data, calculate correlations between element concentrations, contamination factors, geo indexes and so on.

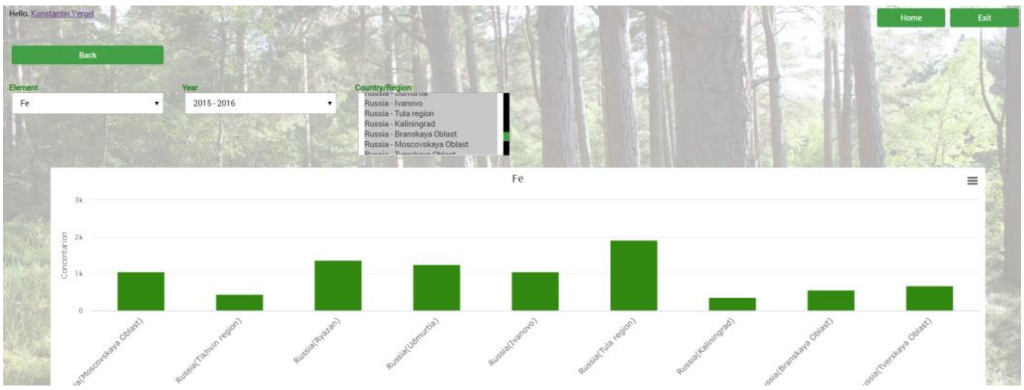

Figure 3. Comparison of the median concentrations of Fe in some regions of Russia

We use the OpenLayers library and several backgrounds to generate maps. Data can be represented as a sampling map, a color map (where color represents concentration) and a graduated symbol map. Contributors can share their maps or statistic metrics, so it can be accessed with no credentials. Mostly participants can access only their own data, but in some cases they can use special tools to get general information from other contributors. For example, it is useful to execute the mean values comparison of median elements concentration between neighboring countries and regions (Fig. 3). Contributors can always forbid joint operation with their data at their profile. As soon as we have information for several atlases, it is possible to get historical reports and build historical charts and graphs.

Coordinators of the program can access any contributor's data, get some general information about completeness and quality of the data, export any type of data, execute group operation and build special combined maps (Fig. 4).

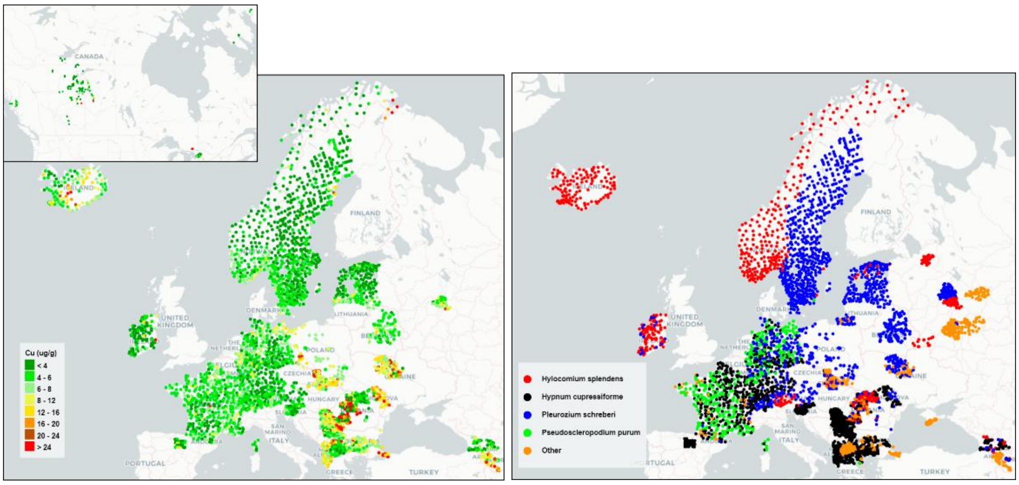

Figure 4. Distribution map of Cu and moss species for the 2014-2015 Survey

3. Mobile application

During data processing for the 2015-2016 Atlas, we experienced misspelling of moss names, wrong coordinates, negative concentrations, and many other problems with data. Now for the 2020-2021 Atlas, we have a mobile application that allows filling in information about sampling sites as required by the UNECE ICP Vegetation manual. The application automatically sets longitude and latitude of the sampling site, controls the correctness of input data and allows capturing photos of moss samples and the nearest area. The application is integrated with DMS, and all information about sampling sites can be imported to the system.

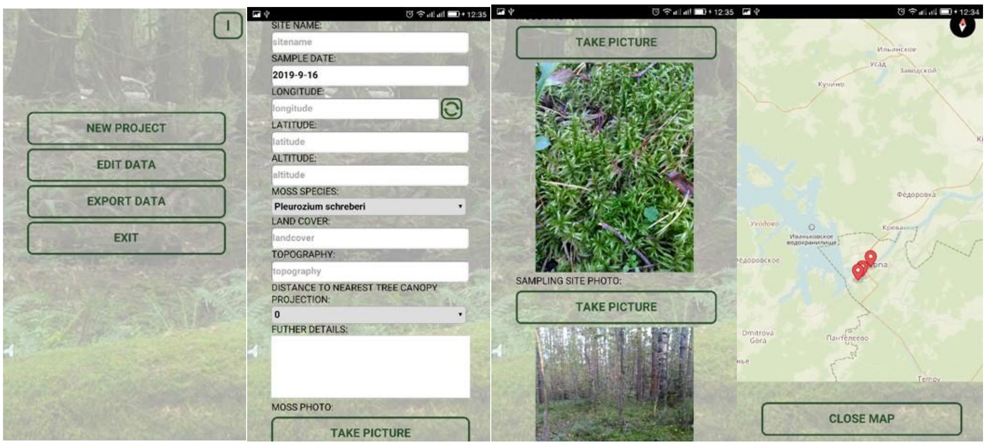

Figure 5. Interfaces of the Moss ICP Vegetation mobile application

We developed the mobile application using Apache Cordova, so we can build it for Android, iOS, and Windows. Currently, we deployed only the Android version that can be found at Google Play under the "Moss ICP Vegetation" name.

4. Prediction

Prediction is an important step in data analysis in any ecological survey. We try to implement machine learning methods to predict heavy metal concentrations. The general idea is to use data that we can get from satellite images together with sampling data from DMS to learn a statistical model or a neural network and then use only data from satellite images to predict concentration. It is clear that getting indexes from satellite images is a much easier process than field sampling. Sampling and analysis of an area like the Moscow region can take 4-6 months. Gathering of indexes for the same area can be done within several days. We use the Google Earth Engine (GEE) [5] – a cloud-based platform for the planetary-scale environmental data analysis to get satellite image indexes. The purpose of the Earth Engine is to perform highly interactive algorithm development at a global scale, push the edge of the envelope for big data in remote sensing, enable high-impact, data-driven science and make a substantive progress on global challenges that involve large geospatial datasets. There are more than 100 satellite programs and modeled datasets. We have several successful studies in this direction, for example, prediction of some heavy metals for Norway, Romania, and Serbia [6]. See Fig. 6 for an example of the model work for Cu in Belgrade. The realization of this idea allows researchers to monitor the evaluation of the situation when it is necessary, get detailed information about areas of interest, and check the situation in cross-border areas. In the future, it will be possible even to partly automate the environment control process. We will be able to automatically run the model and get a notification when the contamination level in some regions is higher than the critical level.

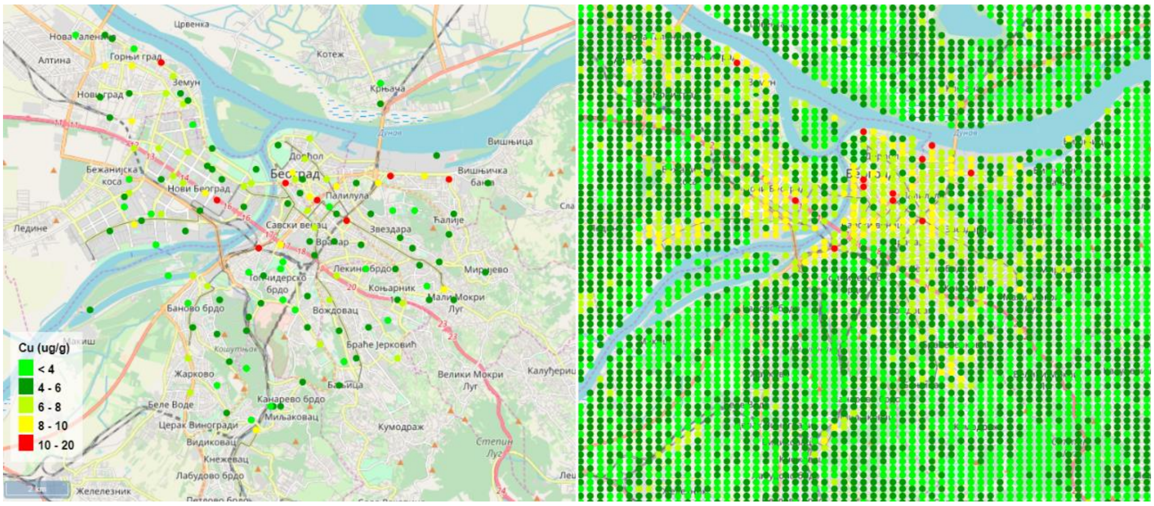

Figure 6. Concentration of Cu in Belgrad in summer. Left – real measurements, right - prediction

5. Conclusion

The study of migration and deposition of toxic pollutants, which include heavy metals, persistent organic pollutants, and radionuclides, and the influence of pollutants on various components of natural and urban ecosystems are key problems of modern biogeochemistry and environmental research. DMS uses modern analytical, statistical, programmatic and organizational methods to supply the ICP Vegetation community with a unified system of collecting, analyzing and processing biological monitoring data. Currently, all key elements of DMS have been implemented. We are going to adjust the functionality of the platform with cluster and factor analysis. A mobile application integrated with DMS was developed for better organization of the samples collection for the 2020–2021 Atlas. The application allows filling in all information about sampling sites required by the UNECE ICP Vegetation manual. Experiments on predicting heavy metal concentrations using machine learning and satellite imagery are in progress.

References

[1] UNECE International Cooperative Programme on Effects of Air Pollution on Natural Vegetation and Crops [Electronic resource]: http://icpvegetation.ceh.ac.uk/ (Accessed 1.10.2019)

[2] Harmens H., Norris D.A., Sharps K., Mills G. … Frontasyeva M., et al. Heavy metal and nitrogen concentrations in mosses are declining across Europe whilst some "hotspots" remain in 2010. Environmental Pollution. pp. 93-104, 2015

[3] Ososkov G., Frontasyeva M., Uzhinskiy A., Kutovskiy N., Nechaevsky A., Cloud platform for data management of the environmental monitoring network: UNECE ICP Vegetation case // CEUR Workshop Proceedings, Vol. 1787, pp. 224-229, 2016

[4] Korenkov V. , Balashov N., Kutovskiy N., Dimitrov V., Kouzmov K., Hristova R., Hristov S., Clouds of JINR, University of Sofia and INRNE — current state of the project, CEUR Workshop Proceedings, vol. 2267, pp. 248-251, 2019

[5] Google Earth Engine project home page [Electronic resource]: https://earthengine.google.com/ (Accessed 1.10.2019)

[6] Uzhinskiy, A., Ososkov G., Goncharov P., Frontsyeva M. Combining satellite imagery and machine learning to predict atmospheric heavy metal contamination, Proceedings of the 8th International Conference of Distributed Computing and Grid technologies in Science and Education. CEUR Workshop Proceedings 2267: 351- 358, 2018