

УДК 00465:58.084

© Коллектив авторов

*Н.А. Кутовский, А.В. Нечаевский, Г.А. Ососков, А.В. Ужинский,  
М.В. Фронтасьева*

# ОБЛАЧНАЯ ПЛАТФОРМА УПРАВЛЕНИЯ ДАННЫМИ БИОМОНИТОРИНГА ПРОЕКТА КОМИССИИ ООН ПО ВОЗДУХУ ЕВРОПЫ ICP VEGETATION



## Введение

Загрязнение воздуха не только самым негативным образом влияет на различные компоненты экосистемы и здоровье человека, но и ведет к значительным экономическим потерям. Не удивительно, что вопросы контроля загрязнения окружающей среды имеют чрезвычайное значение для стран Европы и Азии и курируются Комиссией ООН (UNECE ICP Vegetation), созданной в рамках Конвенции по дальнему трансграничному переносу воздушных загрязнений (CLRTAP).

Изучение атмосферных выпадений тяжелых металлов, азота, стойких органических соединений и радионуклидов базируется на анализе мхов-биомониторов, собираемых каждые 5 лет на территориях стран, участвующих в ICP Vegetation [1, 2]. С 2014 года за координацию программы отвечает Лаборатория нейтронной физики ОИЯИ. В настоящее время общее число стран-участниц достигло 39. За последние два года к программе присоединились Казахстан, Молдова, Азербайджан, Армения, Грузия, Монголия и Вьетнам. Респонденты предоставляют данные об образцах, собранных в десятках тысяч различных мест Европы и Азии. Исследование тяжелых металлов ведется с 1990 года, азота – с 2005 года, стойких органических соединений (POPs) – с 2010 года и радионуклидов – с 2015. Целью этих исследований является определение наиболее загрязненных областей, создание региональных карт загрязнений и улучшение понимания природы долгосрочных трансграничных загрязнений [3].

Сбор образцов происходит по единой методике, утвержденной Комиссией ООН [4]. Для определения концентрации химических элементов наряду с инструментальным нейтронным активационным анализом (INNA), проводимым в ЛНФ ОИЯИ, используются такие аналитические методы, как AAS, AFS, CVAAS, CVAFS, ETAAS, FAAS, GFAAS, ICP-ES, ICP-MS. В отчетах респондентов содержатся

данные о концентрации химических элементов, минимальный набор которых составляет As, Cd, Cr, Cu, Fe, Hg, Ni, Pb, V, Zn, Al, Sb, и N. Отдельные лаборатории проводят исследование на содержание стойких органических соединений и радионуклидов (<sup>210</sup>Pb and <sup>137</sup>Cs) во мхах-биомониторах. В общем случае результаты исследований представляют собой набор точек сбора образцов, их географическое положение (долгота, широта, высота над уровнем моря), ряд описательных параметров [4] и список концентраций элементов. Данные анализируются в различных статистических пакетах, а также используются для построения карт загрязнений в ArcGIS [5]. Примеры карт загрязнений приведены на рис. 1.

## Мотивация

Несмотря на огромную значимость программы UNECE ICP Vegetation, инструменты, используемые ее участниками, недостаточно формализованы и зачастую не соответствуют современным требованиям. Данные в основном хранятся в виде xls-файлов и передаются координаторам посредством электронной почты или физических носителей. Для проведения исследований требуется вручную объединять файлы от участников и использовать различные пакеты для их обработки. Серьезной проблемой является отсутствие единой базы данных, что существенно ограничивает возможности коллективной работы и анализа временных трендов. Подобное положение вещей негативным образом сказывается на скорости и эффективности исследований.

Таким образом, основной целью нашей работы стало создание облачной платформы, которая, используя современные аналитические, статистические, программные и организационные методы, предоставила бы участникам проекта UNECE ICP Vegetation удобные инструменты для управления и обработки данных биомониторинга.

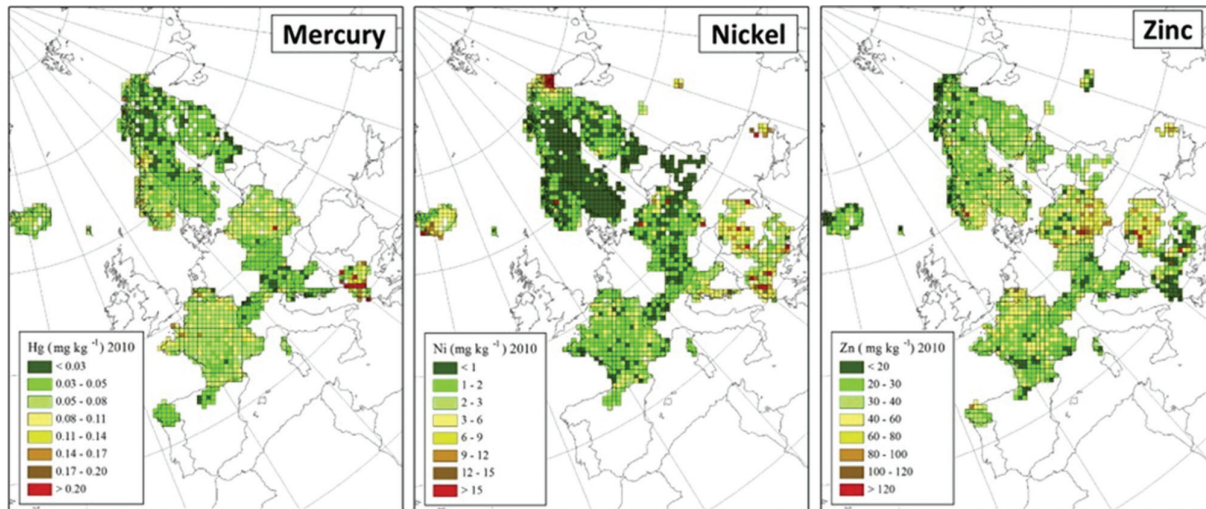


Рис. 1. Примеры карт загрязнений [3]

Элементы платформы должны позволить упростить, а по возможности и автоматизировать все этапы мониторинга, начиная с выбора мест для сбора образцов и заканчивая генерацией карт распределения загрязнений и прогнозированием изменений в окружающей среде. В рамках платформы должны быть реализованы механизмы для коллективной работы участников, статистической проверки достоверности предоставляемых данных и анализа результатов современными методами. Открытая часть платформы должна предоставлять возможность своевременного информирования общественных организаций, предприятий и органов управления о проблемах и изменениях в окружающей среде.

Важнейшим аспектом биомониторинга является анализ и интерпретация полученных данных. Современный подход к подобным задачам заключается в применении многофакторной статистической и интеллектуальной обработки данных с целью определения общих тенденций и прогнозирования изменений в наблюдаемых параметрах. Популярным направлением в данной области является включение в процесс исследований не только топографических и геологических метрик, но и данных дистанционного зондирования земли, космоснимков, аэрофото съемки и прочих данных, способных качественно расширить возможности анализа. Подобные решения характеризуются использованием больших вычислительных мощностей, высокой точностью и динамичностью получения результатов. Задача интеграции различных типов данных тесно связана с разработкой новых методов и алгоритмов их обработки. Широкое распространение получили нейронные сети [6] и самоорганизующиеся карты [7]. Таким образом, еще одной задачей нашего проекта является разработка современных инструментов,

ориентированных на применение нейронных сетей и космоснимков, для многомерного статистического и интеллектуального анализа данных.

#### Архитектурные принципы и используемые технологии

Для достижения требуемого уровня надежности и масштабируемости было решено реализовать платформу в виде ряда взаимосвязанных сервисов и приложений, развернутых на базе облачной инфраструктуры ОИЯИ [8]. Подобный подход позволил оптимизировать использование ресурсов и предоставил удобный механизм для расширения и распространения решения. На рис. 2 представлена архитектура платформы и используемые технологии.

Предварительный анализ данных показал, что при описании точек пробоотбора и результатов исследований образцов может использоваться от 10 до 80 различных параметров. Данные межлабораторных исследований и анализа стойких органических соединений также имеют слабо формализованную структуру. Кроме того, с развитием платформы возникнет необходимость хранить данные других экспериментов и проектов, которые также будут иметь различный состав. Использование традиционных реляционных СУБД в подобных условиях представляется неэффективным, в следствие чего для хранения была предложена нереляционная MongoDB. Разработка серверной части системы ведется на PHP, который используется в том числе и для первичного анализа данных, вычисления статистических характеристик, биологических индексов и факторов. Например, появились возможности расчета коэффициента CF (contamination factor) (1) и индекса накопления загрязнений (geo-accumulation index) I (geo) (2). В качестве веб-сервера используется

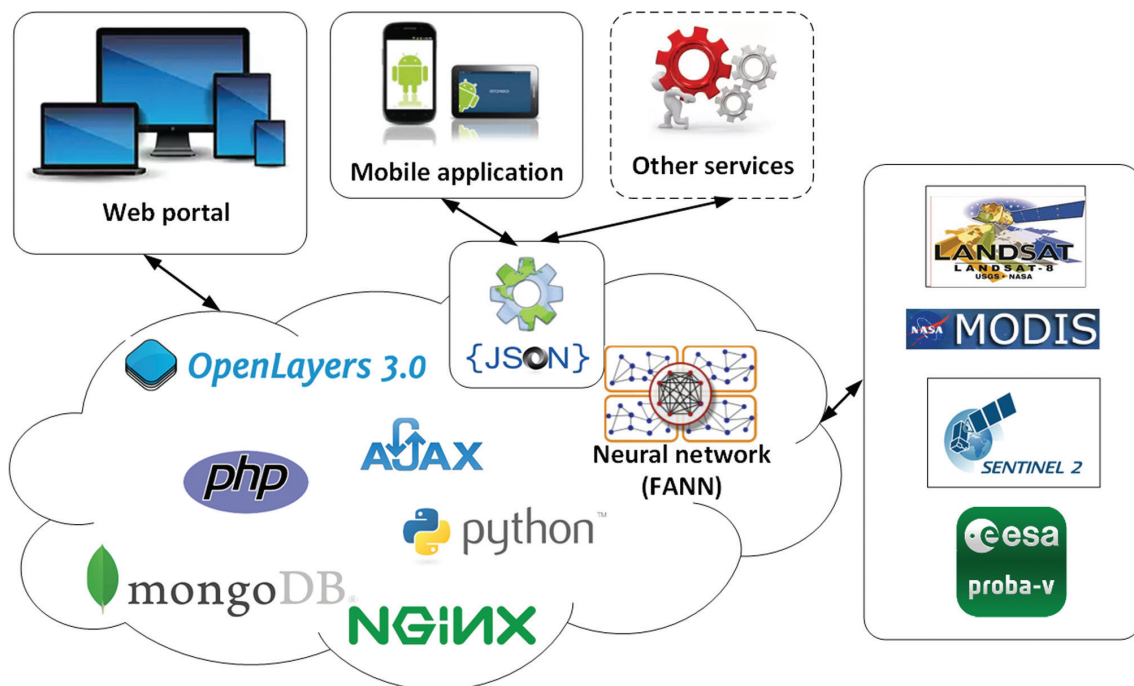


Рис. 2. Архитектура системы и используемые технологии

Nginx, который способен обеспечить необходимый уровень производительности.

$$CF = \frac{Md(element)}{\text{фоновый уровень концентрации элемента}}, \quad (1)$$

$$I(geo) = \log_2 \left( \frac{2CF}{3} \right). \quad (2)$$

Веб-портал платформы реализован в адаптивном дизайне, что позволяет взаимодействовать с ним как с обычного компьютера, так и с любого мобильного устройства. Существуют общедоступная и закрытая части портала. Первая содержит общую информацию по проекту и используется для отображения открытых данных о собранных образцах. Закрытая часть доступна только для зарегистрированных участников программы и используется для управления и анализа данных. Примеры интерфейсов веб-портала приведены на рис. 3.

На текущий момент пользователи могут добавлять данные по пробоотбору, межлабораторным исследованиям и анализу стойких органических соединений. Существует возможность импорта/экспорта данных из/в xls для минимизации проблем при переходе на новую схему работы с данными. Пользователи могут анализировать статистические показатели и корреляцию между элементами, а также рассчитать различные биологические факторы и индексы. Данные могут быть отображены на картах в виде точек пробоотбора либо в виде кругов, чья площадь зависит от концентрации (т.н. «graduated symbol map»). В большинстве случаев участнику

программы доступны только его собственные данные, однако существуют и исключения. Например, исследователей часто интересует информация по средним показателям содержания элементов в граничащих странах и областях для построения различных диаграмм и графиков (рис. 4). В своем профиле пользователь может разрешить или запретить подобные операции с его данными.

У пользователей есть возможность открыть часть своих данных для публичного ознакомления, тогда карты или статистические выкладки будут постоянно доступны по специальной ссылке. Координаторам доступны данные всех пользователей, и они могут совершать с ними групповые операции.

Кроме графического веб-интерфейса, в рамках платформы планируется реализовать RESTful веб-сервис, который будет использоваться для предоставления информации мобильному приложению и прочим сервисам, заинтересованным в данных биомониторинга.

Прогнозирование изменения отслеживаемых параметров – важная и сложная задача, которая также будет решаться в рамках платформы. После накопления достаточного количества данных за исторически значимый период наблюдений будет реализована возможность вычисления различных трендов в процессах загрязнения воздуха. Однако наибольший интерес представляет задача прогнозирования концентрации элементов и построения карт распределения. Ранее для этих целей использовался ArcGIS [5] и различные методы интерполяции. Современные исследования показывают, что существует



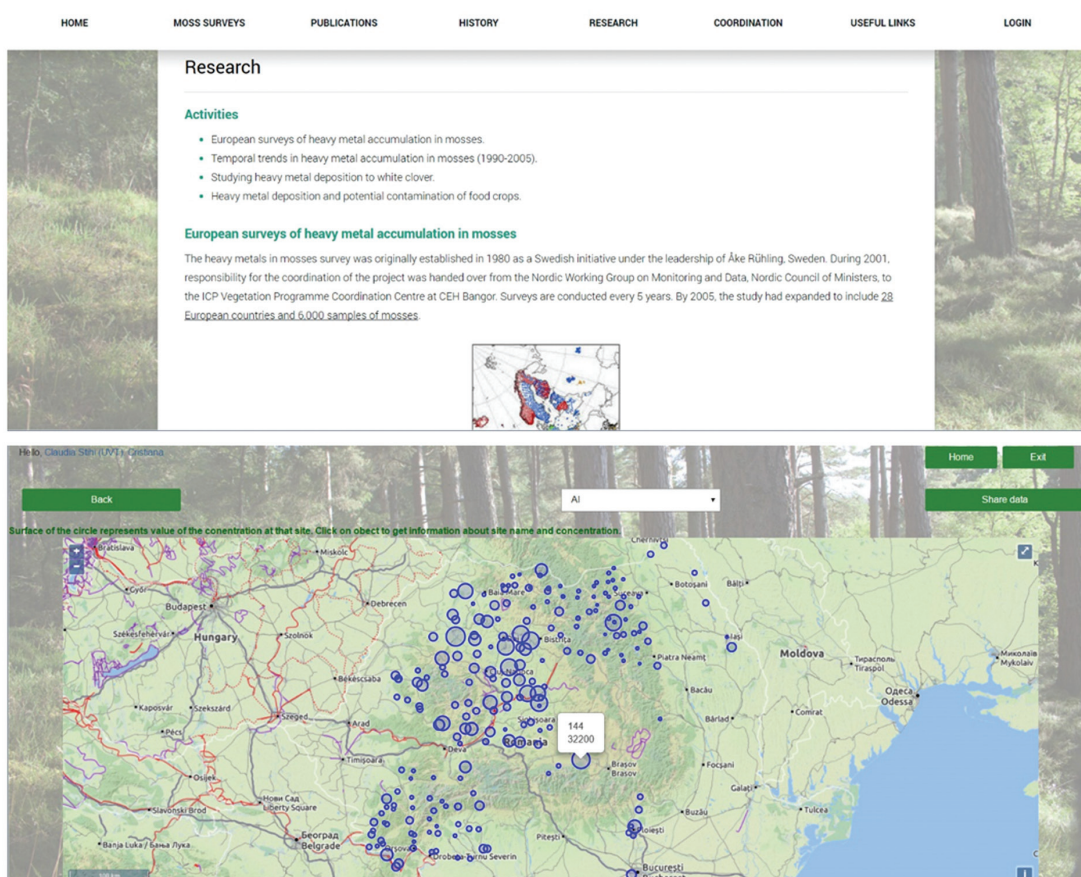


Рис. 3. Интерфейсы. Сверху – общая часть портала. Снизу – карта отображения концентраций

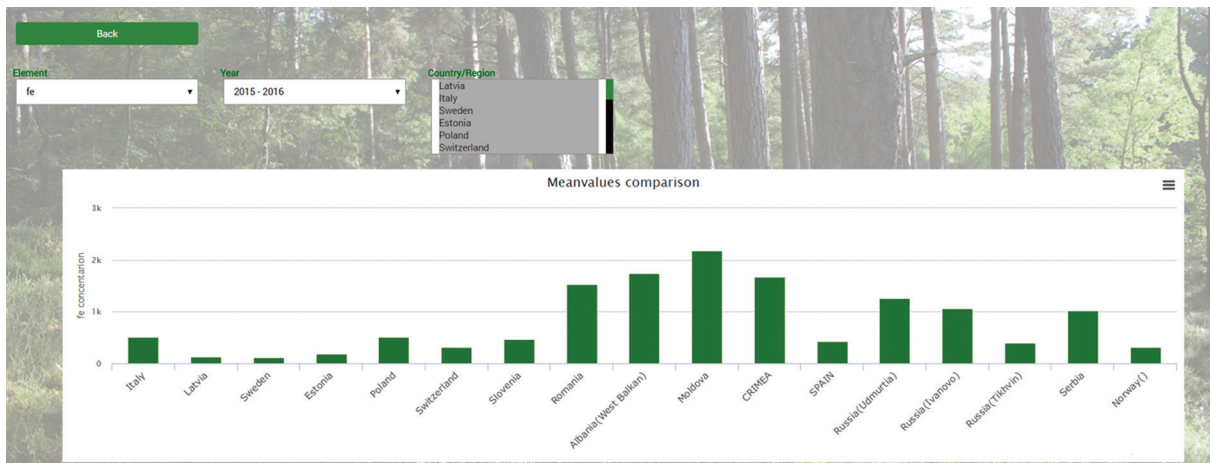


Рис. 4. Сравнение средних показателей концентрации железа по странам и регионам

возможность решения задачи прогнозирования изменения отслеживаемых параметров с использованием нейросетей. Для апробирования такого подхода мы использовали сеть на основе многоуровневого персептрона с 12 входными параметрами (2 координаты и 10 вторичных свойств), одним выходным параметром – концентрацией элемента и двумя промежуточными слоями 25 и 15 нейронов. В качестве функции активации был использован гиперболический тангенс. Сеть обучалась методом градиентного спуска

на модельных данных. В результате нам удалось получить достаточно четкую картинку распределения концентраций с небольшой невязкой. Наиболее доступным источником дополнительной информации для прогнозирования являются космоснимки, сделанные в различных спектрах. Существуют спутниковые программы, позволяющие бесплатно использовать их данные – LandSat, MODIS, Sentinel. Следует заметить, что нашей целью является не только решение поставленной задачи, но и предостав-

ление данных возможностей пользователям в виде сервиса. Мы изучили различные методы получения и обработки космоснимков и пришли к выводу, что единственный подходящий для нас вариант – использование специализированной платформы Google Earth Engine. Этот продукт предоставляет широкие возможности для анализа и обработки данных от более чем 100 различных поставщиков, удобный интерфейс разработчика на JavaScript (рис. 5) и программный интерфейс на Python.

Следующим этапом стало получение ряда количественных показателей в точках сбора образцов для последующего использования их совместно с информацией о концентрациях элементов при обучении нейронной сети. Успешная реализация данной задачи позволила бы, имея только дополнительные данные, прогнозировать концентрации в любой точке исследуемой области.

Используя возможности Google Earth Engine, мы разработали программный модуль, позволяющий получить количественные показатели с космоснимков различных спектров в заданных координатах и рассчитать их корреляцию с концентрациями элементов в точках сбора образцов. Для анализа мы использовали данные нескольких европейских стран и регионов России. Были апробированы различные поставщики данных: MODIS, LandSat 8, Sentinel 2. К сожалению, на данный момент подобрать показатели, подходящие для обучения нейронной сети и получения достаточно точного прогноза, не получилось. Для некоторых элементов использовать подобный подход

вообще не представляется возможным, т.к. степень связанности их концентраций с получаемыми из космоснимков величин близка к нулю. В табл. 1 приведены результаты вычисления корреляции между концентрациями элементов в 612 точках в Швеции с показателями, полученными из различных спектров снимков программы Sentinel 2.

Дальнейшие исследования будут проводиться в области использования альтернативных методов обработки снимков, предоставляемых Google Earth Engine, подключения дополнительных источников данных и применения различных статистических методов структурирования информации. Очевидно, что механизм прогнозирования будет доступен только для ряда элементов и будет включать в себя большое количество этапов предварительного анализа и подготовки данных. Тем не менее, мы надеемся на положительный результат в этом направлении.

### Заключение

Контроль, анализ и определение источников загрязнения окружающей среды – одна из важнейших проблем экологии. Целью проекта является создание облачной платформы, которая, используя современные аналитические, статистические, программные и организационные методы, предоставила бы участникам программы удобные инструменты для управления и обработки данных биомониторинга. Работа проводится в рамках программы Комиссии ООН по воздуху Европы UNECE ICP Vegetation. На текущий момент реализованы все основные

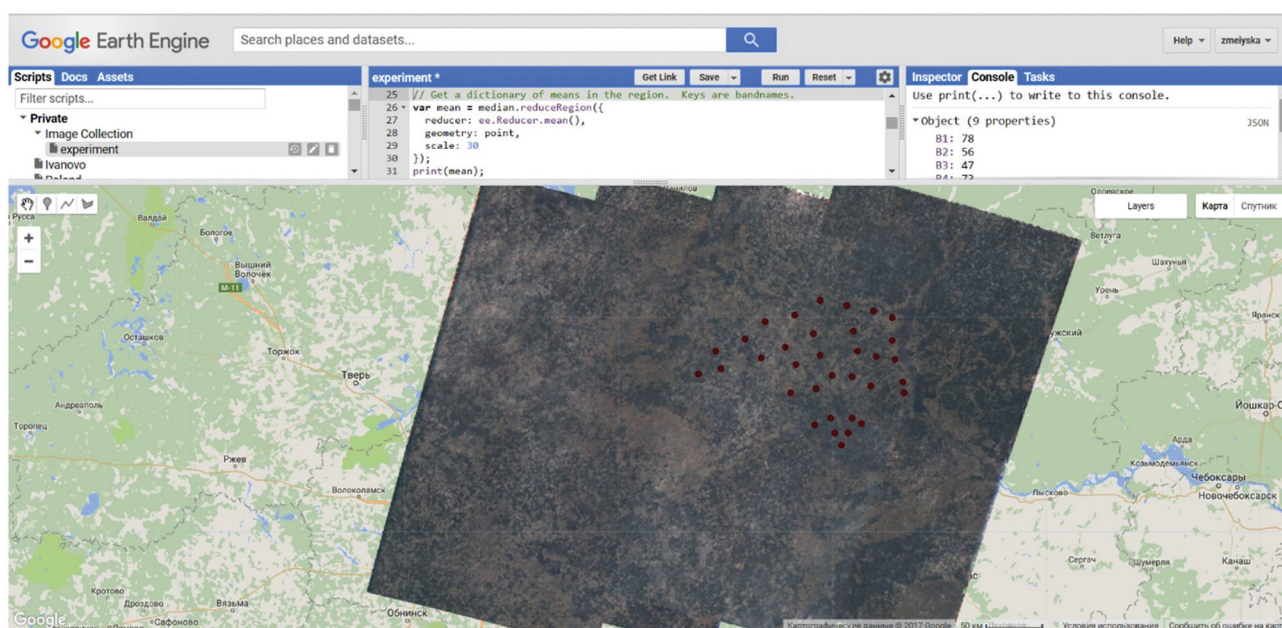


Рис. 5. Интерфейс разработчика Google Earth Engine



Таблица 1

## Связь концентраций элементов в Швеции с данными, полученными от Sentinel 2

	Al	As	Cd	Cr	Cu	Fe	Hg	Ni	Pb	Sb	V	Zn	N	Ba	Sr	Mn	Se	Co
B1	0,28	0,21	0,21	0,19	0	0,23	0	0,25	0	0,17	0,25	0,14	0		0,33	0,16	0,02	0,24
B2	0,28	0,22	0,22	0,19	0	0,23	0	0,25	0	0,16	0,26	0,15	0		0,32	0,16	0,02	0,24
B3	0,28	0,22	0,23	0,19	0	0,23	0	0,25	0	0,16	0,26	0,15	0		0,29	0,15	–0	0,24
B4	0,29	0,23	0,24	0,2	0	0,24	0	0,26	0	0,18	0,27	0,17	0		0,3	0,15	0,01	0,25
B5	0,29	0,23	0,25	0,2	0	0,24	0	0,267	0	0,19	0,27	0,19	0		0,31	0,16	0,01	0,26
B6	0,32	0,25	0,27	0,22	0	0,26	0	0,29	0	0,23	0,3	0,2	0		0,34	0,16	0,02	0,28
B7	0,33	0,27	0,29	0,24	0	0,28	0	0,31	0	0,26	0,32	0,22	0		0,36	0,17	0,03	0,3
B8	0,33	0,28	0,3	0,24	0	0,28	0	0,31	0	0,28	0,32	0,24	0		0,37	0,17	0,04	0,31
B9	0,44	0,38	0,42	0,33	0	0,38	0	0,43	0	0,4	0,44	0,36	0		0,4	0,22	0,04	0,42
B10	0,23	0,22	0,38	0,23	0	0,25	0	0,33	0	0,23	0,31	0,47	0		0,29	0,09	0,13	0,3
B11	0,1	0,18	0,2	0,08	0	0,1	0	0,14	0	0,31	0,11	0,18	0		0,15	0,08	0,07	0,13

элементы платформы. Ведутся исследования в области прогнозирования концентраций элементов с использованием нейросетей и данных космоснимков.

**Ключевые слова:** биомониторинг, управление данными, облачная платформа, сервисы, анализ, мхи-биомониторы.

## ЛИТЕРАТУРА

1. United Nations Economic Commission for Europe International Cooperative Programme on Effects of Air Pollution on Natural Vegetation and Crops [Electronic resource]. – URL: <http://icpvegetation.ceh.ac.uk/> (date of access 20.04.2017).
2. Air Pollution: Deposition to and impacts on vegetation in (South-) East Europe, Caucasus, Central Asia (EECCA/SEE) and South-East Asia : Report prepared by ICP Vegetation (March, 2014) / Editors: H. Harmens, G. Mills. – UK : Bangor : ICP Vegetation Programme Coordination Centre, Centre for Ecology and Hydrology, 2014. – 72 p. [URL: [http://icpvegetation.ceh.ac.uk/publications/documents/CEHOzoneReport2014\\_web-highres.pdf](http://icpvegetation.ceh.ac.uk/publications/documents/CEHOzoneReport2014_web-highres.pdf)].
3. 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. – 2015. – P. 93-104.
4. Heavy metals, nitrogen and POPs in European mosses : 2015 Survey. – URL: <http://icpvegetation.ceh.ac.uk/publications/documents/MossmonitoringMANUAL-2015-17.07.14.pdf> (date of access 16.04.2017).
5. Buse A. et al. Heavy metals in European mosses : 2000/2001 survey. – UK : Bangor : UNECE ICP Vegetation Coordination Centre, Centre for Ecology and Hydrology, 2003.
6. Alijagić J. Application of multivariate statistical methods and artificial neural network for separation natural background and influence of mining and metallurgy activities on distribution of chemical elements in the Stavnja valley (Bosnia and Herzegovina) // PhD thesis. University of Nova Gorica, 2013.
7. Žibret, G., Šajn, R. Hunting for Geochemical Associations of Elements: Factor Analysis and Self-Organising Maps // Mathematical Geosciences. – 2010. – 42(6). – P. 681-703.
8. Baranov A.V., Balashov N.A., Kutovskiy N.A., Semenov R.N. JINR cloud infrastructure evolution // Physics of Particles and Nuclei Letters. – 2016. – Vol. 13. – No. 5. – P. 672-675. – ISSN 1547-4771 ; eISSN: 1531-8567 ; DOI: 10.1134/S1547477116050071.