

УДК 004.9:574

ЭКОЛОГИЧЕСКАЯ ИНФОРМАЦИОННАЯ СИСТЕМА «ОНЕГО»

КАЛИНКИНА
Наталья Михайловна *Федеральный исследовательский центр "Карельский научный центр РАН" Институт водных проблем Севера, cerioda@mail.ru*

КОРОСОВ
Андрей Викторович *Петрозаводский государственный университет, korosov@mail.ru*

ТЕКАНОВА
Елена Валентиновна *Федеральный исследовательский центр "Карельский научный центр РАН" Институт водных проблем Севера, etekanova@mail.ru*

Ключевые слова:
Онежское озеро
экология
информационная
система
ГИС
база данных

Аннотация: Онежское озеро, второй по величине пресноводный водоем Европы, является стратегическим источником пресной воды высокого качества. Важность водного объекта определяет необходимость постоянного контроля за его состоянием в условиях воздействия антропогенных факторов и климатических изменений. Сложность лимнической структуры водоема определила необходимость разработки специальной системы сбора и обработки разнообразной информации – экологической информационной системы «Онего» (ЭИСО). Первый блок системы создан в среде MS Access и представляет собой описание состава, структуры и функционирования базы данных по разнородным характеристикам Онежского озера, включая гидрохимические, гидрологические, гидробиологические, метеорологические и географические характеристики озерной экосистемы, а также источники и уровни антропогенного воздействия. В среде MS Access организованы взаимосвязи между отдельными блоками информации, обсуждаются типы запросов и способы их построения для извлечения выборок данных, подлежащих дальнейшей обработке. Во втором блоке информационной системы ЭИСО все данные связаны на основе картографической информации о размещении точек отбора проб (геоинформационная система), оформленной в среде QGIS. Информационная система позволяет решать как научные задачи (описание динамики экосистемы), так и помогает в решении практических задач, например оценке качества вод. С помощью ЭИСО рассмотрены закономерности распределения общего железа (маркера речных вод) в воде Петрозаводской губы Онежского озера, что позволило дать оценку влияния речных вод на водоем в новых климатических условиях. Изучены процессы эвтрофирования в Кондопожской губе Онежского озера, загрязняемой сточными водами целлюлозно-бумажного комбината. Обсуждаются перспективы развития ЭИСО для разработки модели формирования биопродуктивности Онежского озера.

© Петрозаводский государственный университет
Рецензент: Е. Ю. Косенко

Получена: 16 августа 2018 года

Подписана к печати: 24 марта 2019 года

Введение

Онежское озеро, один из крупнейших водоемов России, расположено в Северо-Западном регионе России. Площадь озера составляет 9720 км², средняя глубина 30 м, максимальная – 120 м, объем воды 295 км³, площадь водосборного бассейна 53100 км². Водоем характеризуется высокой степенью лимнической неоднородности. Северная часть озера расположена на Балтийском кристаллическом щите, для нее характерна высокая изрезанность береговой линии, большое количество заливов и губ. Эта часть озера характеризуется сложным рельефом дна и максимальными для озера глубинами. Здесь же сосредоточены источники загрязнения водоема. Южная часть озера расположена в пределах Русской равнины. Берега здесь слабо изрезаны, дно понижается плавно, глубины меньше, чем в северной части. Лимническую неоднородность озера определяет также наличие циркуляционных течений (Бояринов, Руднев, 1990) и неравномерный прогрев и остывание воды (Петров, 1990). Особые условия формируются в приступьевых участках, поскольку речные воды с высокой цветностью и повышенным содержанием органического вещества и железа резко отличаются по химическому составу от озерных. Влияние всех перечисленных факторов определяет своеобразие гидрологического, химического и биологического режимов в разных частях озера (Онежское озеро..., 2010). Сложность изучения водных сообществ Онежского озера определяется высокой временной и пространственной изменчивостью их показателей. Так, сезонная изменчивость планкtonных организмов связана с одновременным существованием разных фаз в развитии планктона в различных районах озера (Теканова, Сярки, 2015; Фомина, Сярки, 2016). Высокая пространственная вариабельность показателей глубоководного бентоса определяется различиями в геохимическом составе донных отложений (Калинкина и др., 2017а). Все это предопределяет сложности обработки и интерпретации весьма разнородной информации по состоянию экосистемы Онежского озера.

Цель работы состоит в объяснении того, каким образом решались интеллектуально значимые проблемы создания и функционирования ЭИСО. Конструкция информационной системы рассчитана на решение трех насущных задач: построение экспертной системы Онежского озера для характеристики

состояния водоема (Калинкина и др., 2017б), исследование влияния на озерную экосистему климатических изменений (Филатов и др., 2014; Назарова, 2015; Ladoga and Onego..., 2010) и оценка продуктивности озера в рамках реализации социально-экономических задач (Биоресурсы..., 2008).

Материалы

Информационная система «Онежское озеро» объединяет данные за период 1967–2018 гг. Базы данных по отдельным объектам экосистемы Онежского озера включают от сотен до нескольких тысяч записей. В качестве составной части ИСО организована географическая информационная система Онежского озера для переработки географической информации (ГИСО). ИСО и ГИСО непосредственно сопрягаются для обеспечения пространственной привязки атрибутивной информации.

Полный объем информационной системы будет включать следующие двенадцать блоков информации:

1. Контуры акватории и гидрологическая сеть
2. Рельеф дна
3. Макрофиты
4. Фитопланктон
5. Зоопланктон
6. Зообентос
7. Бактериопланктон
8. Рыбные запасы
9. Химический состав вод
10. Температура воды (по акватории)
11. Гидрологический режим притоков
12. Ледовые поля

В настоящее время система включает блоки 1, 2, 4, 5, 6, 10. Материалы по каждому блоку представлены в виде отдельных *тематических таблиц*, содержащих информацию по всем районам Онежского озера (рис. 1). Система ориентирована на наполнение данными, полученными в гидрохимической и гидробиологической практике (Аналитические..., 2017; Методы..., 1976). Данные имеют традиционный формат записи наблюдений.

Для обеспечения наглядности формат исходных таблиц данных в ИСО сохранен, несмотря на завышенные расходы машинной памяти. В некоторых случаях для удобства использования эти таблицы разделены. Например, пробы по планктону всегда содержат данные по численности и биомассе основных групп, но далеко не всегда – по детальному таксономическому составу. Данные по видовому составу представлены

отдельной таблицей, связанной с первой таблицей по ключу «Номер пробы в данный год», YeNo.

Традиционные методы исследований

Федеральный исследовательский центр «Карельский научный центр Российской академии наук» обладает большим объемом данных о состоянии Онежского озера, представленных в многочисленных монографиях Института водных проблем Севера КарНЦ РАН (Экосистема..., 1990; Онежское озеро..., 1999; Биоресурсы..., 2008; Крупнейшие озера-водохранилища..., 2015 и др.). Кроме того, есть зарегистрированные базы данных по хлорофиллу «а» (Сабылина и др., 2018), планктону (Сярки, Куликова, 2012; Сярки и др., 2015) и бентосу озера (Полякова, 2012).

Вместе с тем до последнего времени не было единой информационной системы, позволяющей извлекать и активно использовать разнообразные характеристики озерной экосистемы. Отсутствовала технология извлечения и сопряжения разнородных данных для изучения их причинно-следственных взаимосвязей. Необходимость разработки экологической информационной системы «Онего» (ЭИСО) определяется тем, что она дает возможность экспрессного извлечения разнообразной информации и выявления экологических закономерностей функционирования водной экосистемы. База данных должна быть доступна для использования всеми заинтересованными исследователями. Атрибутивная информация должна соединяться с географической, т. е. между базой данных и геоинформационной системой должна быть организована простая и понятная связь. Информационная система должна нести ряд шаблонов для извлечения типичных выборок, включающих разнородные данные, предназначенные для анализа.

Оригинальные методы исследований

Методы, используемые при разработке информационной системы, определялись необходимостью получения нужной информации разного типа в форме выборок, количественная обработка которых поможет выявить зависимости между извлеченными переменными. При этом связь должна быть установлена как между географическими, так и между атрибутивными характеристиками объектов водного тела Онежского озера.

Для этих целей наиболее эффективными современными компьютерными средствами являются следующие программы. Для ввода вручную первичной информации ис-

пользуется пакет MS Excel; часть данных переведена в электронный формат Excel из сканограмм полевых журналов. Основная атрибутивная информация (включая координаты точек отбора проб) накапливается в среде пакета MS Access, который организует реляционную базу данных по Онежскому озеру (БДО), обеспечивая связь исходных таблиц по ключевым полям. Здесь с помощью процедуры «запрос» из одной или нескольких базовых таблиц создаются смешанные выборки, подлежащие дальнейшей обработке. Географическая информационная система для Онежского озера (ГИСО) включает топографические, гидрологические данные, точки отбора проб, коллекцию космических снимков. Она представлена в среде программы QGIS (URL: <https://qgis.org/ru/site/>) и служит как целям иллюстрации, так и основой для построения запросов по географическим критериям. Ключом для объединения данных из разных таблиц является локализация во времени и пространстве – срок и область акватории для отбора химических и гидробиологических проб. Связь между программами Access и QGIS осуществляется при подключении атрибутивных таблиц в среду ГИСО. Для обслуживания разных целей из базы данных строятся следующие типы выборок:

- ряды данных для построения иллюстраций,
- выборки данных для поиска статистических зависимостей,
- выборки данных для проведения многомерного анализа,
- группы данных для построения имитационных моделей.

Обработка получаемых выборок выполнялась в средах MS Excel и объектно-ориентированного языка R. Необходимо отметить, что средства пакета R позволяют обойтись без всех выше перечисленных программ. Однако от его исключительного использования мы отказались по следующим соображениям. На наш взгляд, созданная система должна сохранить прозрачность и быть доступной для использования базы данных всеми участниками исследовательского процесса.

Структура ЭИСО

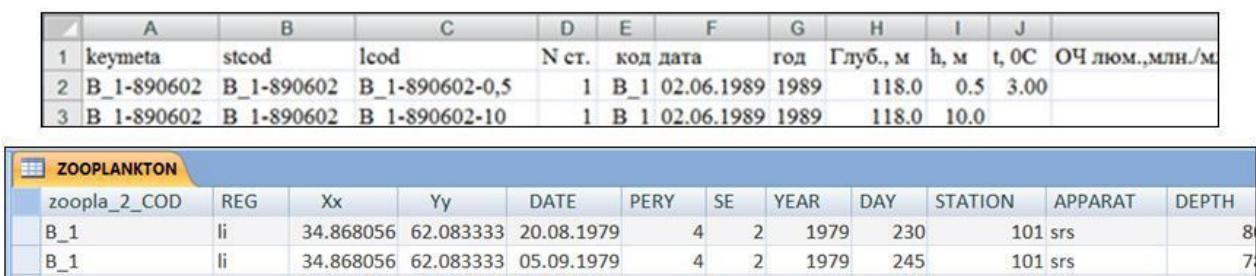
Исходные таблицы с данными оформлялись в среде Excel. Все таблицы были унифицированы, снабжены общими уникальными составными ключами, поля переименованы. В результате все записи стали равнозен-

но уникальны, т. е. каждая проба содержит описание всех характеристик, но имеет свой уникальный индекс (ключ).

Основные виды проблем при подготовке данных для базы

Все данные хранятся и модифицируются в среде Access, поэтому типы и структура данных должны соответствовать требованиям этой программы. Все поля (названия столбцов) были переименованы на латинице заглавными буквами (прежние и новые названия сохранены в файле легенды с расшифровкой обозначений). Многие значения в ячейках исходных таблиц включали кириллические символы, которые не читаются в использованных программных средах, и были заменены на латинские буквы.

Исправлены все ошибки, из-за которых поля имели неправильный текстовый формат, в том числе все даты, ранее записанные разными способами, текстовые примечания в числовых полях вроде «мало», «следы», «-», а также разные варианты написания дробных чисел (то с точкой, то с запятой). Формат всех таблиц для компонентов среды одного типа и за разные годы был унифицирован, и все они сведены в одну таблицу для одного типа данных. Разные способы написания координат проб заменены на одинаковый формат – градусы в виде десятичной дроби. Каждой пробе были приписаны собственные координаты, часть из которых определялась по описаниям района и места отбора пробы (наиболее старые данные) (см. рис. 1).



The screenshot shows two tables in Microsoft Access. The top table is a raw data dump with columns: A (keymeta), B (stcod), C (lcod), D (N ст.), E (код), F (дата), G (год), H (Глуб., м), I (h, м), J (t, 0C), and K (OЧ лям.,млн./м). The bottom table is a structured header for 'ZOOPLANKTON' with columns: zoopla_2_COD, REG, Xx, Yy, DATE, PERY, SE, YEAR, DAY, STATION, APPARAT, and DEPTH. Data rows are shown for 'B_1' in both tables.

Рис. 1. Пример исходного и итогового заголовка таблицы (фитопланктон)

Fig. 1. The example of the initial and final table header (phytoplankton)

Состав базы данных и геоинформационной системы

Все таблицы базы данных «Онего» (БДО) сведены в общий файл среды Access – BDO.accdb. Всего в настоящее время мы структурировали в среде Access следующие таблицы:

1. фитопланктон (phytoplankton),
2. зоопланктон (zooplankton),
3. зообентос (zoobenthos),
4. бактериопланктон (bacterioplankton),
5. химический состав вод озера и притоков (chemiwater),
6. гидрологический режим некоторых притоков (rivers).

Несмотря на рутинный характер и неопределенный тип многих полей, которые ранее служили индексами для отбора записей, мы сохранили их в базе. Это позволяет в сложных случаях восстановить принадлежность некоторых проб. Для текущей полуавтоматической работы с данными были созданы новые ключевые поля YEAR (год отбора пробы), NO (номер пробы в пределах текущего года). Для отбора данных служат поля: номер пробы, фенологическая дата отбора,

глубина отбора, координаты точки отбора. Средствами Access несложно рассчитать различные биологические индексы, как, например, «индекс качества воды» – отношение численности сапрофитных бактерий к общей численности всех бактерий. Индексы создаются с помощью простого запроса и могут быть добавлены как новые поля к таблице данных (запрос на обновление).

Для обслуживания типичных (традиционных) задач во все таблицы добавлено поле REG, в которое внесены индексы семи «главных» регионов акватории Онежского озера (ре – Петрозаводская губа, ко – Кондопожская губа, li – Лижемская губа, ро – Повенецкий залив, со – южное Онего, bo – Большое Онего). В других случаях отбор данных по географическим критериям (принадлежность точки к той или иной акватории) выполняется в среде QGIS с помощью ГИСО, включающей в себя картографическое описание водоема и взятых проб. Оформлены следующие слои с таблицами (проекция картографическая UTM / WGS84 Zone 36N):

1. контуры береговой линии Онежского озера и основные притоки,
2. рельеф дна.

При решении конкретных задач оперативно создаются слои точек отбора проб (для перечисленных выше компонент 1–5), в качестве таблицы данных выступает запрос из среды Access в формате *.csv. Уместная в таком случае технология OLE оказалась более громоздкой; для довольно небольших объемов используемой нами информации мы избрали другой путь связи между средами Access и QGIS.

Главная функция экологической информационной системы «Онего» (ЭИСО) состоит в быстром и безошибочном формировании выборок данных для исследования тех или иных зависимостей между переменными или их распределения по акватории Онежского озера. В системе ЭИСО возможен отбор проб в пространстве (область акватории, глубина), во времени (год, сезон) и по типу объекта (биологический, химический, метеорологический). Ниже рассмотрены процедуры формирования типичных запросов с примерами.

Выборка из одной таблицы для «стандартных» областей акватории озера

Поскольку индекс «стандартной» области уже присутствует в таблицах данных, нет необходимости обращаться к ГИСО (среды QGIS), и запрос для этих регионов создается только в среде Access. Здесь существенным этапом оказывается построение простого запроса, в котором отбираются те поля из тематической таблицы, которые содержат нужные данные. Далее таблица этого запроса экспортируется в простой формат (*.csv) (в пункте «Экспорт в текстовый файл» нужно вручную заменить расширение txt на csv). Обработка выполняется в среде Excel или R (рис. 2, 3). Очередность использования программы такая: (Excel ®) Access (® R). Собственно формирование выборки происходит только в среде Access, но данные поставляет среда Excel, а обработка выполняется в среде R – поэтому они упомянуты, но помещены в скобках.

Первым примером может служить подсчет числа проб зоопланктона, собранных в разных областях Онежского озера.

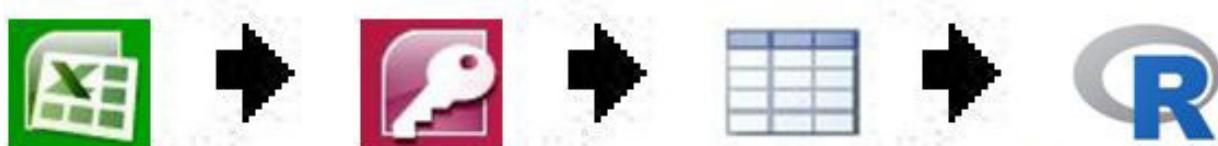


Рис. 2. Схема построения выборки для «стандартных» областей Онежского озера

Fig. 2. The scheme of constructing a sample for the "standard" areas of Onega Lake

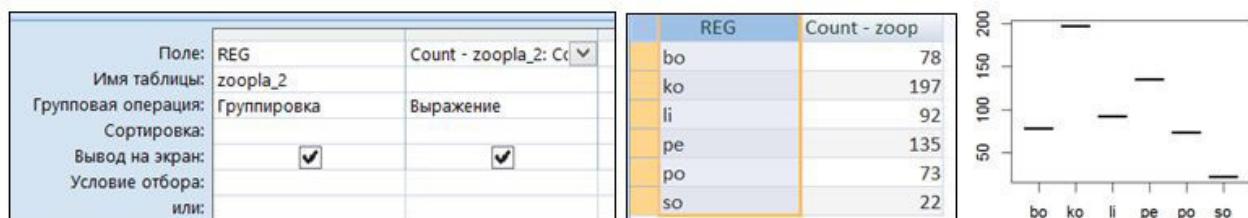


Рис. 3. Запрос на выборку числа проб зоопланктона в «стандартных» областях

Fig. 3. Request for selecting the number of zooplankton samples in "standard" areas

Второй пример – изучение сезонной динамики фитопланктона. В запросе заданы критерии для отбора объекта (численность фитопланктона в столбе воды, кл./л), области акватории (bo, Большое Онего), про-

должительности периодов наблюдений в течение года (ежемесячно). Запрос строится только в среде Access и обрабатывается в R (рис. 4).

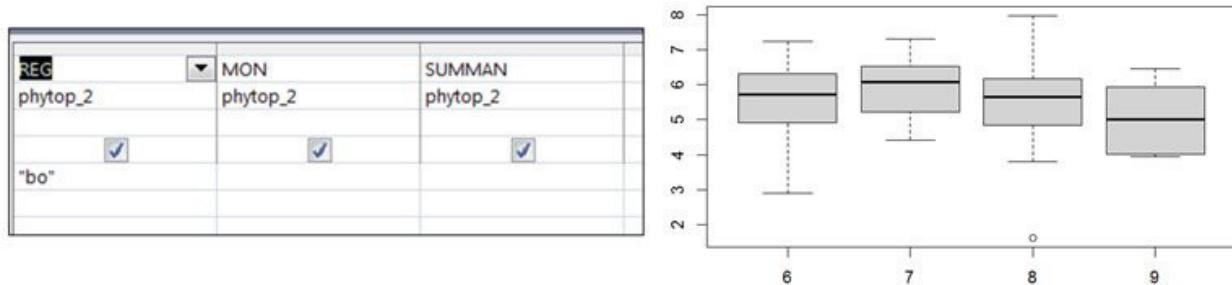


Рис. 4. Структура запроса на выборку показателей численности фитопланктона из центрального Онего за все годы по месяцам и «ящики с усами» для каждого месяца

Fig. 4. The structure of the request for selecting indicators of the phytoplankton abundance from the central Onega for all years by month and "boxes with a mustache" for each month

Выборка из одной таблицы для разных областей акватории озера

Если «стандартный» вариант подразделений акватории не достаточен для целей исследования, приходится обращаться к ГИСО. В среде QGIS создается региональный слой нового подразделения акватории, индексы новых районов посредством оверлея приписываются к внедренной из Access тематической таблице, которая вместе с индексами регионов экспортируется в среду обработки данных – Excel, или R, или же обрабатывается в среде Access (сводный запрос) (рис. 5). Очередность использования программы следующая: (Excel ®) Access ® QGIS ® Access (® R, Excel, Access).

В нашем примере из среды Access экспортируем запрос из таблицы zooplankton

в формате *.csv (в пункте «Экспорт в текстовый файл» нужно вручную заменить расширение txt на csv) и импортируем его в среду QGIS («Добавить слой CSV»). Здесь же строим слой с двумя регионами Петрозаводской губы (PE_12), который накрывает отдельные группы проб зоопланктона (zoopla) (рис. 6А). Задаем индекс областям (в поле id), для вершинной части – 2, для горловой – 1. Выполняя оверлейную процедуру Пересечение, указываем сначала слой точек (проб) (рис. 6Б), к которому будут добавлены индексы из слоя областей (рис. 6В). Экспорт этой расширенной таблицы в удобном формате *.csv позволяет импортировать ее в программы обработки данных, например, построить сводную таблицу в среде Excel (рис. 6Г) или Access, или в R.



Рис. 5. Схема построения выборки для разных областей Онежского озера
Fig. 5. The scheme of sampling construction for different areas of Onega Lake

Выборка из одной таблицы для разных областей акватории и лет

В предыдущих примерах отбор данных был выполнен сначала по атрибутивным критериям из таблиц среды Access, а затем по географическим критериям из таблиц среды QGIS. Однако возможны задачи, когда сначала должны работать географические критерии, а затем – атрибутивные.

В качестве примера организации выборки рассмотрим один из заливов Онежского озера. Весной Петрозаводская губа заполняется водами притока р. Шуя, которые «запираются» в заливе из-за весеннего термического бара (Петров, 1990). С началом летней

стратификации озерные воды начинают поступать в залив. Организуем выборку для построения камерной модели разбавления в Петрозаводской губе речных вод озерными. Для построения камерной модели эту акваторию разбили на 5 областей (рис. 7). Задача состоит в том, чтобы для каждой из этих областей получить из базы данных значение концентрации железа – регионального маркера речных вод для каждого периода исследований. В качестве периодов исследования назначены 2008 и 2009 гг. Таким образом, выборка должна содержать для каждой из 5 областей по 2 значения концентрации железа (в летний период) (рис. 7А).

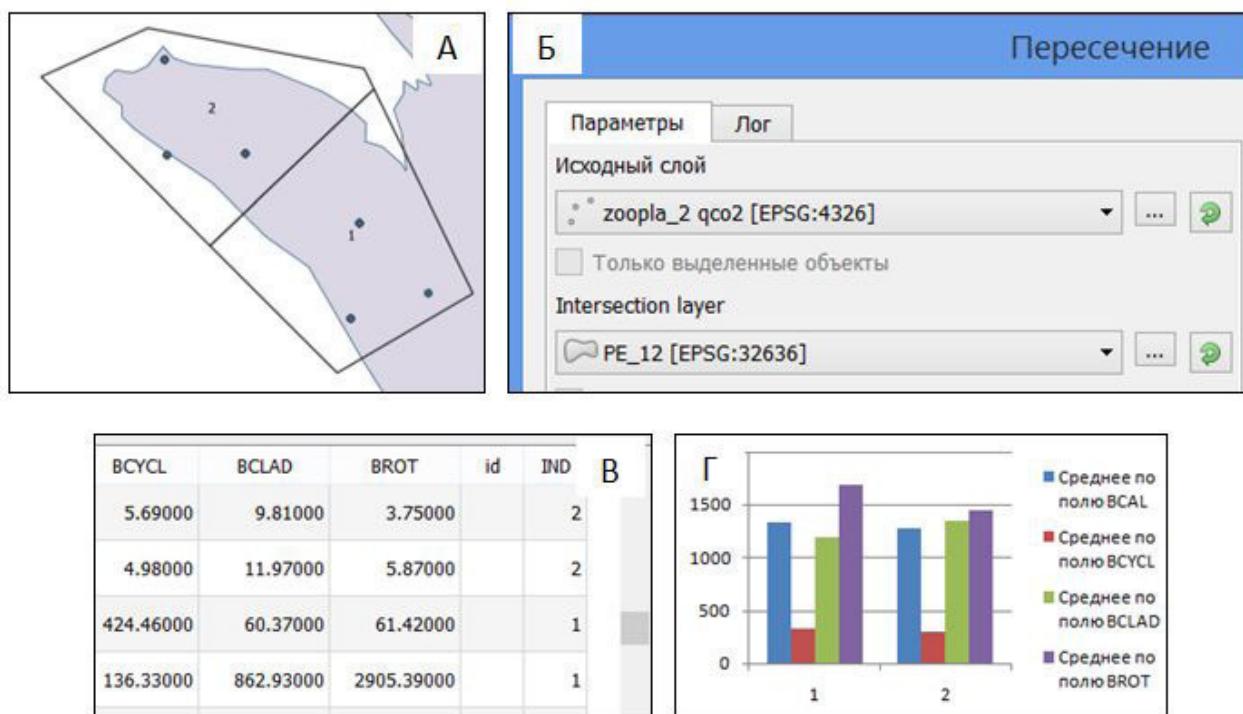


Рис. 6. Построение выборки для проб по численности зоопланктона Петрозаводской губы
Fig. 6. The sampling construction for zooplankton abundance samples in Petrozavodsk Bay

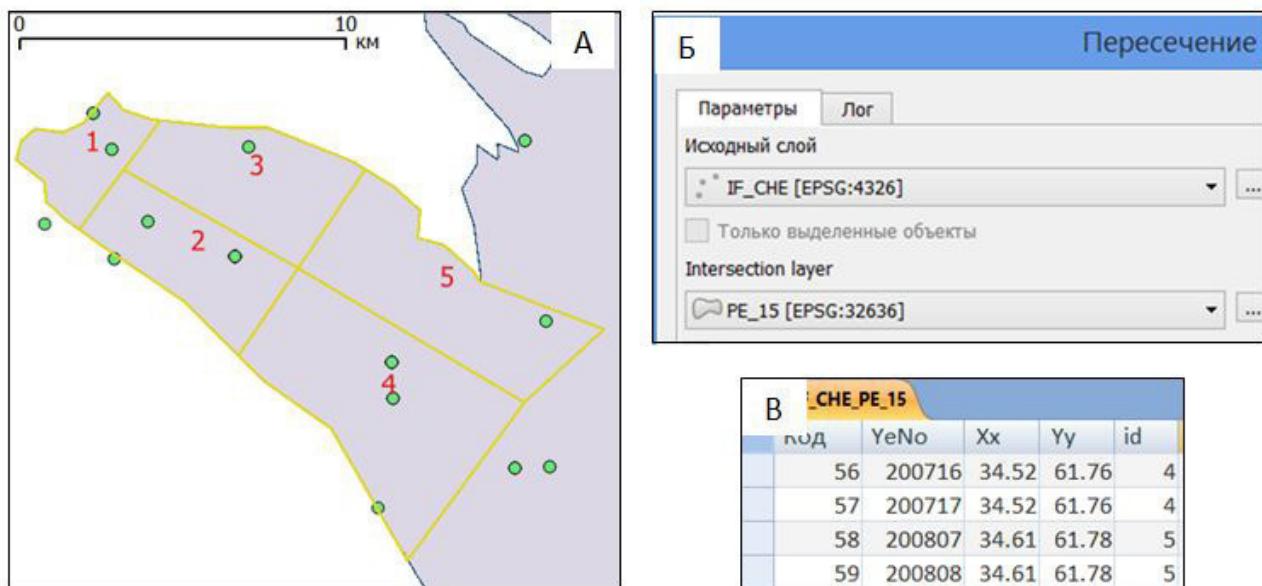


Рис. 7. Области Петрозаводской губы для модели разбавления шуйских вод (А), окно построения «Пересечения» слоя проб со слоем областей Петрозаводской губы (Б) и файл «Пересечения» в среде Access (В)
Fig. 7. The Petrozavodsk Bay areas for modeling the dilution of Shuiski waters (A), the "Intersection" window for the layer of samples with the layer of Petrozavodsk Bay areas (B) and the file "Intersection" in the Access (B)

Технологическая задача решается посредством создания географического индексного файла, на основании которого выполняется объединение нужных данных (рис. 8).



Рис. 8. Порядок построения выборки для разных областей Онежского озера
Fig. 8. The procedure of constructing a sample for different areas of Onega Lake

Сначала в среде Access из таблицы по химическому составу вод (chewater) формируется запрос, содержащий поля Xx, Yy (для позиционирования проб), YeNo (для связи данных), и экспортируется в формате *.csv (IF_CHE.csv).

Файл импортируется в среду QGIS с помощью процедуры Добавить слой CSV; координаты берутся из полей Xx и Yy. Затем создается слой пяти областей акватории (PE_17). Далее выполняется оверлейная процедура Пересечение (рис. 7Б), в результате которой каждая проба приобретает индекс заданной области (рис. 7В). Результат пересечения экс-

портируется в формат *.csv (IF_CHE_PE_15.csv), который затем импортируется в среду Access.

Этот географический индексный файл с помощью простого запроса присоединяется к таблице с химическими характеристиками онежских вод; в качестве ключевых используются поля YeNo (см. рис. 7). Затем таблица этого запроса используется для отбора нужных записей по критериям сроков (июнь, июль, август, сентябрь 2007 и 2008 гг.) и глубины (5–20 м) отбора проб. Вносим эти условия в структуру запроса и выполняем его (рис. 9).

Рис. 9. Структура и результат связи индексного файла с таблицей химических проб

Fig. 9. The structure and result of the connection of the index file with the table of chemical samples

Затем строим перекрестный запрос: в строках – годы, в столбцах – области акватории, в ячейках таблицы – средние значения (рис. 10). К сожалению, выставленные критерии для отбора данных оказались слишком жесткими – не для каждой зоны Петрозаводской губы они выполняются, и в результирующей таблице остаются пустоты. Очевидно, критерии следует смягчить, на-

пример, охватывать для усреднения более длительные периоды времени.

Полученный перекрестный запрос экспортируется в формат *.csv (Fe_Pe15.csv) и служит исходной таблицей данных для построения имитационной модели разбавления речных вод озерными в Петрозаводской губе.

[YEAR]	[id]	[FEOB]
chewater Запрос1	chewater Запрос1	chewater Запрос1
Группировка	Группировка	Avg
Заголовки строк	Заголовки столбцов	Значение

chewater Запрос1_перекрестный					
YEAR	1	2	3	4	5
2007	0.279	0.102		0.1435	
2008			0.14		0.147

Рис. 10. Структура и итог перекрестного запроса для данных по химии вод
Fig. 10. The structure and result of the cross-query for water chemistry data

Выборка из нескольких таблиц для разных областей акватории озера

Основная типично экологическая задача – поиск связи между некоторыми, количественно характеризующими зависимые экологические объекты. В отношении базы данных задача состоит в формировании выборки, включающей как минимум две переменных разного типа. Критерием для объединения данных из разных таблиц выступает принадлежность точек отбора этих проб из одной области акватории озера (географический критерий).

Выборка строится в три этапа: экспорт двух подготовленных запросов с точками отбора проб биоты и химических показателей качества воды (среда Access); построение зон акватории для выражения изучаемого градиента условий и отбор проб, попадающих в эти зоны (среда QGIS), объединение данных двух типов в таблицу и их обработка (среда R) (рис. 11). При подготовке выборок программы используются в следующем порядке: (Excel ®) Access ® QGIS ® R ® (R).



Рис. 11. Схема построения выборки из двух таблиц
Fig. 11. The scheme for constructing a sample from two tables

Для примера рассмотрим задачу «оценить состояние биоты по градиенту загрязнения в Кондопожской губе Онежского озера в летний период в последние годы». Предстоит объединение данных из двух таблиц среды Access (БДО) – по химическому составу (chem_3) и зоопланктону (ZOOPLANKTON).

Среда Access призвана отобрать данные, нужные для обработки.

Сначала необходимо выбрать список изучаемых переменных и сроки исследования. Создаем простой запрос, в котором из таблицы ZOOPLANKTON отбираем данные по биомассе четырех групп за 2000–2017 гг. (рис. 12).

Экспортируем запрос в формате *.csv из Главного меню: Внешние данные / Экспорт в текстовый файл. В диалоговом окне предлагается вариант имени файла «ZOOPLANKTON Запрос.txt», которое необходимо отредакти-

ровать. Во-первых, сократить и уточнить (в файле дана биомасса, В), во-вторых, задать расширение csv: «ZOO B.csv». В следующем диалоговом окне указать, что первая строка содержит имена полей, а разделитель полей – запятая, нажать «Готово».

Аналогично создается второй простой запрос, из таблицы CHEWATER, в котором отбираются данные по содержанию в воде кислорода за тот же период, и экспортируется в файл «CHE O2.csv».

Среда QGIS служит для внесения в отобранные данные географического индекса принадлежности каждой пробы к той или иной области акватории озера и отбора тех записей (проб), которые в эти области попадают.

Градиент загрязнения Кондопожской губы можно выразить косвенно разбиением акватории на серию участков, удаленных от

Xx	Yy	YEAR	BCAL	BCYCL	BCLAD	BROT
ZOOPLANKTON						
<input checked="" type="checkbox"/>						
[YEAR]>2000						

Xx	Yy	YEAR	BCAL	BCYCL	BCLAD	BROT
34.868056	62.083333	2001	2406.5692	404.7749	2008.483	181.2865
34.868056	62.083333	2005	2917	367	1832	66
34.868056	62.083333	2006	8574.137	977.2035	4515.1045	4038.641
34.868056	62.083333	2007	2795.0169	183.339	1556.0298	104.3133
34.868056	62.083333	2010	2834.01288	741.2238	522.99888	3046.67472

Рис. 12. Структура и содержание запроса по показателям зоопланктона

Fig. 12. The structure and content of the request for indicators of zooplankton

источника загрязнений на разное расстояние. В примере построен векторный слой с тремя зонами (КО 123), в единственное поле id внесены номера зон: 1, 2, 3 (рис. 13). Зонирование акватории залива для получения выборок по пространственному критерию осуществлялось экспертино.

Выполняем импорт таблиц данных по зоопланктону и химическому составу воды с помощью процедуры «Добавить слой CSV». В диалоговом окне указываем поля с координатами и долготы (Xx), и широты (Yy), проекция картографическая, WGS84.

Далее всем пробам приписываем индекс тех зон, в которые они попали (см. рис. 13) – с помощью операции оверлея (Вектор / Пересечение). В том же диалоговом окне выбираем пункт Пересечение / Сохранить в файл и расширение *.csv. Получаем два экспортованных файла – ZOO В КО123.csv и СНЕ О2 КО123.csv с дополнительным полем id, в котором для каждой пробы указан номер зоны (рис. 14) (формат *.csv – текстовый, его можно посмотреть и в Блокноте).

Среда R дает возможность составить выборки из записей, совмещающих разные типы данных для оценки зависимости между ними. В нашем примере нужно получить записи, содержащие одновременно и оценки биомассы планктона, и значения концентраций веществ. Поскольку пробы разных компонентов озерной экосистемы исследователи берут в разное время и в разных точках, типичной оказывается ситуация, когда на одну запись одних данных приходит несколько записей данных другого типа.

Например, для третьей зоны за данный период обнаружилось 28 химических проб и 38 проб зоопланктона. Однако для любого вида оценки зависимостей необходимо формировать двумерную таблицу парных значений без пробелов.

Ликвидировать пропуски можно тремя общими способами. Для иллюстративных целей для одной области можно просто усреднить разные данные по столбцам (или взять другую статистику: медиану, дисперсию...); тогда каждая область получит одну-единственную запись, представленную двумя оценками состояния – биоты и среды (рис. 15). Такой вариант удобен также для построения картограмм пространственного распределения какого-либо показателя на всей акватории озера, разбитой на серию равновеликих областей (квадратов).

Второй способ предназначен для статистического анализа и призван сохранить разнообразие данных. Это означает, что изучаемые переменные должны каким-то образом образовывать пары (в пределах каждой зоны). Один из вариантов реализован в среде QGIS на вкладке Свойства / Связи, с помощью которого по ключевым полям (id) можно реализовать связь «многие к одному», когда всем показателям зоопланктона может быть приписано одно (последнее в списке) значение химической характеристики. Другой менее удачный вариант состоит в организации связи «все со всеми», когда каждое значение одной переменной будет образовывать пары со всеми значениями переменной другого типа. Таким ме-

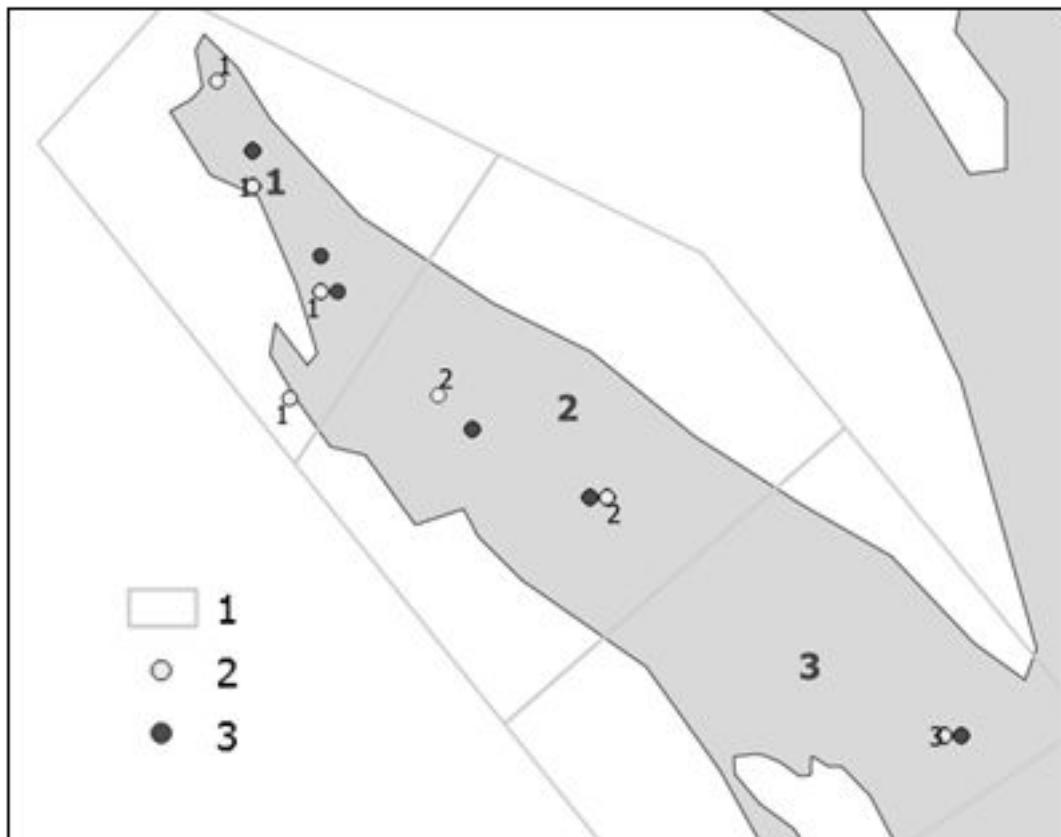


Рис. 13. Зоны Кондопожской губы (1), биологические (2) и химические (3) пробы
Fig. 13. Zones of Kondopoga Bay (1), biological (2) and chemical (3) samples

A screenshot of a Windows Notepad window titled 'ZO B KO 123.csv — Блокнот'. The window contains a table of data with the following columns: Xx, Yy, YEAR, BCAL, BCYCL, BCLAD, BROT, and id. The data is as follows:

Xx	Yy	YEAR	BCAL	BCYCL	BCLAD	BROT	id
34.25	62.19	1970	449.85	58.84	3262.17	2485.57	1
34.25	62.19	1999	645.12	250.17	7496.9	692.96	1

Рис. 14. Данные проб зоопланктона с номерами зон
Fig. 14. The data of zooplankton samples with zone numbers

тодом для третьей зоны можно получить 1064 сочетания (28×38) значений биомассы и содержания кислорода. К сожалению, так просто составленные выборки вызывают появление мнимых повторностей (Шитиков и др., 2008); с этой проблемой лучше не сталкиваться.

Построить адекватное сочетание неравного числа проб (n_1, n_2) можно с помощью процедуры ресamplingа (рандомизации или

бутстрепа) (Шитиков, Розенберг, 2014). Для подготовки рандомизированной выборки из таблиц одного и другого типа данных (в пределах одной зоны) случайным образом берут по n записей. Если при этом $n_1 > n > n_2$, то часть записей одного типа не будет избрана, а часть записей другого типа будет частично продублирована. Повторение процедуры в общем случае породит другую выборку. Они могут быть раздельно использованы в стати-

```
mch<-tapply(ch$O2, ch$id, mean)
mzo<-tapply(zo$BCLAD, zo$id, mean)
plot(mch, mzo, type='p', cex=3)
text(mch, mzo, cex=.8)
```

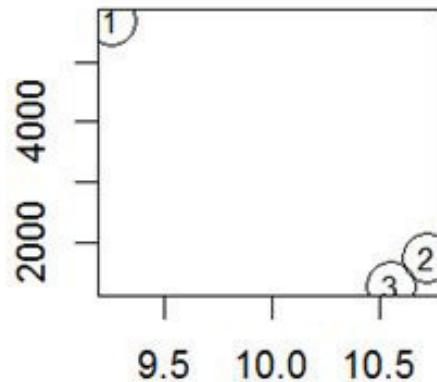


Рис. 15. Расчет соотношения средних уровней содержания кислорода в воде (ось абсцисс, ch\$O₂, содержание кислорода, мг/л) и биомассы Cladocera (ось ординат, zo\$BCLAD, биомасса ветвистоусых ракообразных, мг/л) в трех зонах Кондопожской губы

Fig. 15. Calculation of the ratio of average oxygen levels in water (abscissa axis, ch\$O₂, oxygen content, mg/l) and Cladocera biomass (ordinate axis, zo\$BCLAD, Cladocera biomass, mg/l) in three zones of Kondopoga Bay

стическом анализе для взаимной верификации полученных параметров. Программные средства R позволяют извлекать сотни и тысячи таких повторных выборок, что дает возможность строить распределения оценок зависимости (коэффициентов регрессии, корреляции и пр.) и на этой основе делать статистические выводы.

Итак, задача состоит в том, чтобы в среде R сформировать общую таблицу, каждая

запись которой имеет поля id (идентификатор региона), O₂ (содержание кислорода) и BCLAD (биомасса ветвистоусых ракообразных). Ориентируясь на количество проб, собранных в каждой зоне (таблица), можно принять объем пробы для ресамплинга равным 30, что близко к минимальному числу проб в любой из зон.

Число реальных проб и объемы выборок при ресамплинге

Зона	1	2	3
Число значений O ₂	29	41	28
Число значений DCLAD	57	52	38
Число значений в выборке	30	30	30

Алгоритм расчетов состоит в том, чтобы из таблиц СНЕ O2 KO123.csv (ch) и ZOO В KO123.csv (zo) извлекать по 30 строк, взятых в случайном порядке (sample) в пределах каждой зоны (ind), и сводить их в объединенную таблицу zoch (рис. 16). Всего получаем таблицу с 90 записями – с 90 парами значений BCLAD и O₂. Теперь можно рассчитать уравнения регрессии или коэффициенты корреляции.

Поскольку в полученной объединенной таблице использованы не все данные, имеет смысл выполнить описанную процедуру еще несколько раз, например 100 или 1000. В результате получаем распределение значений коэффициентов корреляции (см. рис. 16).

Характер этого распределения позволяет

судить о значимости отличия коэффициентов от нуля: если ноль входит в доверительный интервал для среднего значения коэффициента, значит, связь между переменными не доказана. В нашем примере практически все коэффициенты укладываются в интервал $-0.12 \dots -0.09$, в который ноль не входит. Доказана отрицательная связь между переменными: чем меньше содержание кислорода (т. е. выше загрязнение), тем выше биомасса ветвистоусых раков. Феномен объясняется тем, что органическое загрязнение вод в вершинной части Кондопожской губы не слишком токсично, но вызывает интенсивное развитие бактерий, снижающих уровень кислорода и служащих пищей для раков.

```
zoch<-data.frame(ind[i],  
+sample(ch[ch$id==ind[i],1],30,replace = TRUE),  
+sample(zo[zo$id==ind[i],1],30,replace = TRUE))
```

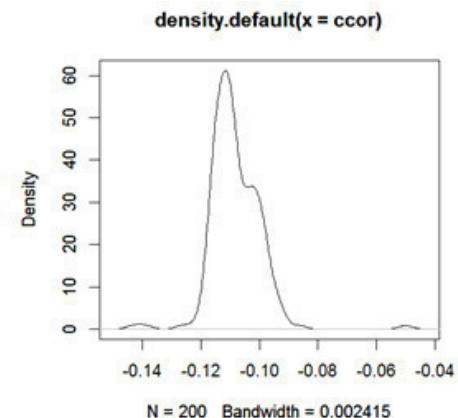


Рис. 16. Фрагмент кода для составления сочетаний показателей химического состава вод (ch) и биомассы зоопланктона (zo) в одной зоне (i) и распределение 200 коэффициентов корреляции между переменными BCLAD и O_2 для выборок объемом 30 проб

Fig. 16. The code fragment for the construction of combinations of water chemical composition (ch) and zooplankton biomass (zo) in one zone (i) and distribution of 200 correlation coefficients between variables BCLAD and O_2 for groups of 30 samples

Обсуждение

Основой любой информационной системы является качественная исходная база данных, от которой и зависит практическая значимость и точность получаемых прогнозов. Для создания ЭИСО были использованы многолетние (с 1964 г.) разноплановые данные по морфометрии, термическому, химическому и биологическому режимам Онежского озера. Важнейшей частью настоящей работы явилась организация данных в полноценную базу, наполненную всеми необходимыми характеристиками единого формата, что потребовало большого количества времени авторов.

Разработка экологической информационной системы «Онего» позволила систематизировать большой объем накопленных данных и выполнить анализ результатов гидрохимических и гидробиологических наблюдений по различным районам Онежского озера. Опыт использования ЭИСО показал ее высокую эффективность при формировании выборок данных, сократив в разы время анализа. Особенную важность это свойство системы приобретает при выполнении перекрестных запросов при работе одновременно с несколькими выборками. Немаловажным для эксперта представляется и возможность мгновенной визуализации данных анализа путем их географической привязки к карте Онежского озера.

ЭИСО позволяет перенаправить усилия с технической работы по формированию выборок на осмысление полученных зависимо-

стей, понимание происходящих изменений в состоянии экосистемы Онежского озера.

Вместе с тем представленная система не является завершенной. Один из вопросов, требующих решения при сведении данных разных таблиц, это соотношение между реальными объемами исходных данных и выборок, подготовленных в процессе ресамplingа для составления сводной таблицы. К сожалению, практически невозможен идеальный вариант полного совпадения двух исходных таблиц по числу проб и их локализации. Следовательно, случайность в композиции проб неизбежна, остается ее статистически описать. Вопрос требует дополнительной проработки, однако здравый смысл подсказывает направление рассуждений. Во-первых, если уменьшить объемы рандомизированных выборок ниже реального числа проб, это приведет к потере их репрезентативности и росту изменчивости расчетных коэффициентов. Во-вторых, если n будет существенно превышать n_1 и n_2 , благодаря изъятию с возвратом, в выборках появятся избыточно дублированные данные, что означает появление мнимых повторностей и ложный рост репрезентативности результатов. Вывод ясен: объем сводной выборки n (в пределах отдельного региона) не должен существенно отличаться от объемов выборок исходных переменных (n_1 и n_2); лучше ориентироваться на меньший объем ($\min(n_1, n_2)$), чтобы избежать проблемы мнимых повторностей (Шитиков и др., 2008). С другой стороны, чем сильнее отличаются n_1 и n_2 , тем больше операций ресамplingа

нужно выполнить, тем больше коэффициентов зависимости нужно рассчитать, чтобы апробировать все возможные сочетания значений исходных переменных. Поскольку расчет автоматизирован программой R, можно задать 100, 1000 или 10000 циклов – результирующее распределение будет все

более гладким и определенным. В нашем примере наиболее плавным оказалось распределение коэффициентов корреляции при объеме рандомизированной выборки, примерно равной минимальному числу проб в отдельной зоне и при 200 прогонах рандомизации (рис. 17).

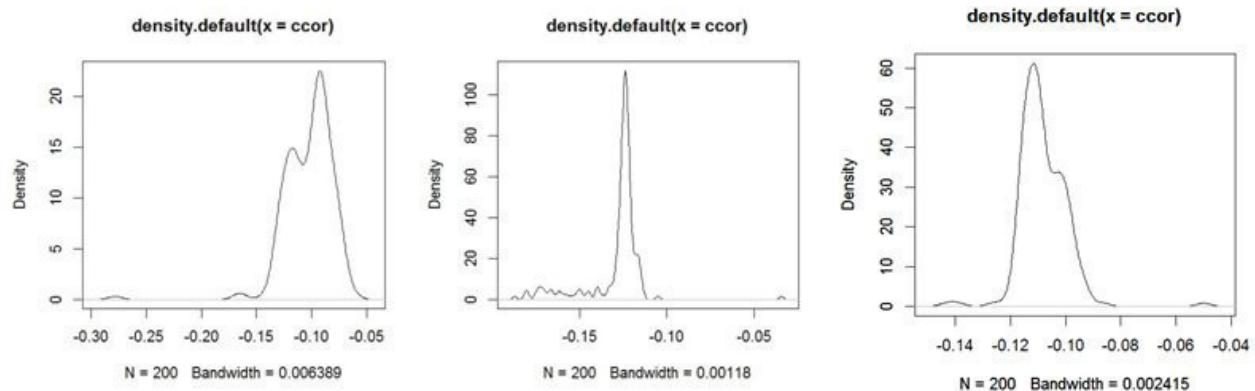


Рис. 17. Распределение k коэффициентов корреляции между переменными BCLAD и O_2 , полученных для разных случайных объединенных выборок разного объема (по n в каждой зоне): А: $k = 300$, $n = 10$, Б: $k = 200$, $n = 20$, В: $k = 200$, $n = 30$ (ось абсцисс – значения коэффициентов корреляции)

Fig. 17. Distribution of k correlation coefficients between variables BCLAD and O_2 obtained for different random combined samples of different volume (in n in each zone): A: $k = 300$, $n = 10$, B: $k = 200$, $n = 20$, В: $k = 200$, $n = 30$ (the abscissa axis – the values of the correlation coefficients)

Заключение или выводы

Онежское озеро – это водный объект, имеющий стратегическое значение как источник питьевой воды высокого качества, что определяет необходимость постоянного контроля за его состоянием. Проблемы экологического мониторинга Онежского озера связаны с высокой лимнической гетерогенностью его экосистемы. Экспрессная оценка современной экологической ситуации на водоеме возможна только на основе информационной системы, связывающей в единый комплекс большое количество накопленных данных и методы их анализа. Такой структурой является разработанная к настоящему времени экологическая информационная система «Онего» (ЭИСО). Использование предлагаемой информационной системы уже позволило

рассмотреть закономерности распределения общего железа (маркера речных вод) в воде Петрозаводской губы Онежского озера, в результате чего получена оценка влияния речных вод на водоем в новых климатических условиях. Важным результатом для понимания процессов, протекающих в загрязняемой сточными водами Кондопожской губе, является установление связей между химическим составом воды и показателями зоопланктона с учетом разных объемов используемых выборок. Другими важнейшими задачами, которые предстоит решить на основе использования ЭИСО, являются разработка экспертной системы для оценки экологической ситуации на водоеме и разработка модели формирования биопродуктивности экосистемы Онежского озера.

Библиография

- Аналитические, кинетические и расчетные методы в гидрохимической практике / Под ред. П. А. Лозовика, Н. А. Ефременко. СПб.: Нестор-История, 2017. 272 с.
- Биоресурсы Онежского озера / Ред. В. И. Кухарев, А. А. Лукин. Петрозаводск: Карельский НЦ РАН, 2008. 272 с.
- Бояринов П. М., Руднев С. Ф. Инструментальные исследования течения // Экосистема Онежского озера и тенденции ее изменения. Л.: Наука, 1990. С. 53–71.
- Калинкина Н. М., Белкина Н. А., Сидорова А. И., Галибина Н. А., Никерова К. М. Биотестирование

донных отложений Онежского озера с учетом их химического состава и показателей состояния глубоководного макрообентоса // Принципы экологии. 2017а. Т. 6. № 1. С. 25–55. DOI: [10.15393/j1.art.2017.6022](https://doi.org/10.15393/j1.art.2017.6022).

Калинкина Н. М., Коросов А. В., Сярки М. Т. К созданию экспертной системы Онежского озера: оптимизация мониторинга состояния экосистемы по показателям зоопланктона // Принципы экологии. 2017б. № 1. С. 117–132. DOI: [10.15393/j1.art.2017.5864](https://doi.org/10.15393/j1.art.2017.5864).

Крупнейшие озера-водохранилища Северо-Запада ЕТР: современное состояние и изменения экосистем при климатических и антропогенных воздействиях / Отв. ред. Н. Н. Филатов. Петрозаводск: КарНЦ РАН, 2015. 375 с.

Методы биологического анализа пресных вод . Л.: ЗИН АН СССР, 1976. 168 с.

Назарова Л. Е. Атмосферные осадки в Карелии // Труды Карельского научного центра РАН. 2015. № 9. С. 114–120.

Онежское озеро: Атлас / Отв. ред. Н. Н. Филатов. Петрозаводск: Карельский НЦ РАН, 2010. 151 с.

Онежское озеро. Экологические проблемы / Под ред. Н. Н. Филатова. Петрозаводск: Карельский НЦ РАН, 1999. 293 с.

Петров М. П. Термический режим // Экосистема Онежского озера и тенденции ее изменения. Л.: Наука, 1990. С. 32–37.

Полякова Т. Н. Макрообентос Онежского озера . Свидетельство о государственной регистрации базы данных № 2012620882. 31.08.2012.

Сабылина А. В., Теканова Е. В., Калинкина Н. М. Хлорофилл «а» в воде Онежского озера . Свидетельство о государственной регистрации базы данных № 2018621068. 13.07.2018.

Сярки М. Т., Куликова Т. П. Зоопланктон Онежского озера . Свидетельство о государственной регистрации базы данных № 2012621150. 9.11.2012.

Сярки М. Т., Теканова Е. В., Чекрыжева Т. А. Планктон пелагиали Онежского озера . Свидетельство о государственной регистрации базы данных № 2015620274. 13.02.2015.

Теканова Е. В., Сярки М. Т. Особенности фенологии первично-продукционного процесса в пелагиали Онежского озера // Известия РАН. Серия биологическая. 2015. № 6. С. 645–652.

Филатов Н. Н., Руховец Л. А., Назарова Л. Е., Георгиев А. П., Ефремова Т. В., Пальшин Н. И. Влияние изменений климата на экосистемы озер севера Европейской территории России // Ученые записки Российского государственного гидрометеорологического института. № 34. СПб.: РГГМУ, 2014. С. 49–55.

Фомина Ю. Ю., Сярки М. Т. Определение динамических характеристик зоопланктона Онежского озера // Принципы экологии. 2016. № 4 (20). С. 49–56. DOI: [10.15393/j1.art.2016.5223](https://doi.org/10.15393/j1.art.2016.5223)

Шитиков В. К., Цейтлин Н. А., Якимов В. Н. Мифы и реальность мнимых повторностей С. Хелберта // Проблемы экологического эксперимента (Планирование и анализ наблюдений) / Ред. Г. С. Розенберг, Д. Б. Гелашвили. Тольятти: СамНЦ РАН: Кассандра, 2008. С. 89–112.

Шитиков В. К., Розенберг Г. С. Рандомизация и бутстреп: статистический анализ данных по биологии и экологии с использованием R . Тольятти: Кассандра, 2014. 314 с.

Экосистема Онежского озера и тенденции ее изменения / Под ред. З С. Кауфмана. Л.: Наука, 1990. 264 с.

Ladoga and Onega – great European Lakes. Observations and Modeling / L. Rukhovets, N. Filatov (eds.). Chichester, UK: Springer, 2010. 302 p.

Благодарности

Работа выполнена в рамках Государственного задания Института водных проблем Севера КарНЦ РАН.

ECOLOGICAL INFORMATION SYSTEM «ONEGA»

KALINKINA
Nataliya Michailovna *Karelian Research Centre Russian Academy of Sciences Northern Water Problems Institute, cerioda@mail.ru*

KOROSOV
Andrey Viktorovich *Petrozavodsk State University, korosov@mail.ru*

TEKANOVA
Elena Valentinovna *Karelian Research Centre Russian Academy of Sciences Northern Water Problems Institute, etekanova@mail.ru*

Key words:
Lake Onega
ecology
information system
GIS
database

Summary: Lake Onega, the second largest freshwater reservoir in Europe, is a strategic source of fresh water of high quality. The value of the water body determines the need for constant monitoring of its condition under the impact of anthropogenic factors and climate change. The complexity of the limnetic structure of the reservoir has determined the need for the development of a special system for collecting and processing different information about the lake – the ecological information system “Onega” (EISO). The first block of the system was created in the MS Access. This block is a description of the composition, structure and functioning of the database on the different characteristics of Lake Onega, including hydrochemical, hydrological, hydrobiological, meteorological and geographical characteristics of the lake ecosystem, as well as the sources and levels of anthropogenic impact. In the MS Access, the relationships between different blocks of information are arranged. The types of queries and methods of their construction for the extraction of data samples to be further processed are discussed. In the second block of the EISO information system, all data are linked on the basis of cartographic information about the location of sampling points (geoinformation system) formatted in the QGIS. The information system allows solving scientific problems (describing the dynamics of the ecosystem), it also helps to solve practical problems, for example, to assess water quality. With the help of the EISO the distribution of total iron (a marker of river water) in the water of the Petrozavodsk Bay of Lake Onega was considered, which made it possible to assess the impact of river water on the reservoir in new climatic conditions. The processes of eutrophication in the Kondopoga bay of Lake Onega contaminated by waste water from pulp and paper mill were studied. The prospects of the development of the EISO for the development of the model of the formation of the bioproductivity of Lake Onega are discussed.

Reviewer: E. YU. Kosenko

Received on: 16 August 2018

Published on: 24 March 2019

References

- Analytical, kinetic and calculation methods in hydrochemical practice, Pod red. P. A. Lozovika, N. A. Efremenko. SPb.: Nestor-Istoriya, 2017. 272 p.
- Bioresources of Onega Lake, Red. V. I. Kuharev, A. A. Lukin. Petrozavodsk: Karel'skiy NC RAN, 2008. 272 p.
- Boyarinov P. M. Rudnev S. F. Instrumental studies of current, Ekosistema Onezhskogo ozera i tendencii ee izmeneniya. L.: Nauka, 1990. P. 53–71.
- Ecosystem of Lake Onega and trends of its change, Pod red. Z. P. Kaufmana. L.: Nauka, 1990. 264 p.
- Filatov N. N. Ruhovec L. A. Nazarova L. E. Georgiev A. P. Efremova T. V. Pal'shin N. I. Climate change impact on the ecosystems of lakes of the north of European Russia, Uchenye zapiski Rossiyskogo gosudarstvennogo gidrometeorologicheskogo instituta. No. 34. SPb.: RGGMU, 2014. P. 49–55.
- Fomina Yu. Yu. Syarki M. T. Determination of dynamic zooplankton characteristics of Lake Onega, Principy èkologii. 2016. No. 4 (20). P. 49–56. DOI: [10.15393/j1.art.2016.5223](https://doi.org/10.15393/j1.art.2016.5223)
- Kalinkina N. M. Belkina N. A. Sidorova A. I. Galibina N. A. Nikerova K. M. Bioassay of lake Onega bottom

- sediments toxicity based on their chemical composition and deepwater macrozoobenthos state, Principy ekologii. 2017a. T. 6. No. 1. P. 25–55. DOI: [10.15393/j1.art.2017.6022](https://doi.org/10.15393/j1.art.2017.6022).
- Kalinkina N. M. Korosov A. V. Syarki M. T. For creating an expert system of Lake Onega: optimization of monitoring the state of the ecosystem on zooplankton indicators, Principy ekologii. 2017b. No. 1. P. 117–132. DOI: [10.15393/j1.art.2017.5864](https://doi.org/10.15393/j1.art.2017.5864).
- Ladoga and Onega – great European Lakes. Observations and Modeling, L. Rukhovets, N. Filatov (eds.). Chichester, UK: Springer, 2010. 302 p.
- Methods of biological analysis of fresh water. L.: ZIN AN SSSR, 1976. 168 p.
- Nazarova L. E. Precipitation over the territory of Karelia, Trudy Karel'skogo nauchnogo centra RAN. 2015. No. 9. P. 114–120.
- Onega Lake. Ecological problems, Pod red. N. N. Filatova. Petrozavodsk: Karel'skiy NC RAN, 1999. 293 p.
- Onega Lake: Atlas, Otv. red. N. N. Filatov. Petrozavodsk: Karel'skiy NC RAN, 2010. 151 p.
- Petrov M. P. Thermal regime, Ekosistema Onezhskogo ozera i tendencii ee izmeneniya. L.: Nauka, 1990. P. 32–37.
- Polyakova T. N. Macrozoobenthos of Lake Onega. Svidetel'stvo o gosudarstvennoy registracii bazy dannyh No. 2012620882. 31.08.2012.
- Sabylina A. V. Tekanova E. V. Kalinkina N. M. Chlorophyll "a" in Onega Lake water. Svidetel'stvo o gosudarstvennoy registracii bazy dannyh No. 2018621068. 13.07.2018.
- Shitikov V. K. Ceytlin N. A. Yakimov V. N. Myths and reality of S. Hurlburt's pseudoreplication, Problemy ekologicheskogo eksperimenta (Planirovanie i analiz nablyudeniy), Red. G. P. Rozenberg, D. B. Gelashvili. Tol'yatti: SamNC RAN: Kassandra, 2008. P. 89–112.
- Shitikov V. K. Rozenberg G. S. Randomization and bootstrap: statistical analysis of data on biology and ecology using R. Tol'yatti: Kassandra, 2014. 314 p.
- Syarki M. T. Kulikova T. P. Zooplankton of Lake Onega. Svidetel'stvo o gosudarstvennoy registracii bazy dannyh No. 2012621150. 9.11.2012.
- Syarki M. T. Tekanova E. V. Chekryzheva T. A. Plankton of Lake Onega pelagic zone. Svidetel'stvo o gosudarstvennoy registracii bazy dannyh No. 2015620274. 13.02.2015.
- Tekanova E. V. Syarki M. T. Peculiarities of phenology of the primary production process in the pelagic zone of lake Onega, Izvestiya RAN. Seriya biologicheskaya. 2015. No. 6. P. 645–652.
- The largest lakes-reservoirs of the North-west European part of Russia: current state changes of ecosystems under climate variability and anthropogenic impact, Otv. red. N. N. Filatov. Petrozavodsk: KarNC RAN, 2015. 375 p.