

ИСПОЛЬЗОВАНИЕ СУБД И ГИС ДЛЯ РАЗВИТИЯ И АНАЛИЗА РЕГИОНАЛЬНОЙ ОКЕАНОЛОГИЧЕСКОЙ БАЗЫ ДАННЫХ ПО ЧЁРНОМУ МОРЮ

А.М. Новикова

Институт природно-технических систем, РФ, г. Севастополь, ул. Ленина, 28
E-mail: anelkapi@bk.ru

Сформирован новый обобщенный массив базы данных по температуре вод Чёрного моря за период 1884–2017 гг. с использованием ГИС QGIS и СУБД SpatiaLite на основе баз данных ИПТС, БМЭ и JCOMMOPS, содержащей сведения о съёмках с буёв Argo. Специальные инструменты обработки данных в SpatiaLite значительно уменьшили время обращения запросов к БД при одновременном экономии объёма памяти, занимаемом базой. Привлечение сторонних баз данных позволило увеличить общий объём БД ИПТС в 1,6 раз по количеству станций и в 3 раза по измерениям. Также существенно пополнился объём данных контактных съёмок для последних 10 лет. Анализ пространственного распределения данных по основным горизонтам (0, 50, 75, 100, 200, 300, 500 и 1000 м) показал равномерность их расположения по всей акватории Чёрного моря. Создана карта глубин Чёрного моря на основе растровой батиметрической карты ETOPO.

Ключевые слова: ГИС, системы управления базами данных, Чёрное море, температура.

Поступила в редакцию: 03.05.2018.

Введение. Анализ и управление пространственно-распределёнными данными удобно осуществлять с помощью инструментария географических информационных систем (ГИС). Однако, когда объём данных превышает 1,5 млн записей в таблице, в ГИС значительное время уходит на их загрузку и обработку. В таком случае на помощь исследователю приходят системы управления базами данных (СУБД). Одним из последних нововведений в открытых ГИС, например QGIS, было подключение возможностей управления большими массивами с помощью пространственных СУБД PostGIS и SpatiaLite через модуль DbManager [1]. В данной работе формат хранения океанографических баз данных позволил обратиться к ним через язык SQL СУБД SpatiaLite [2].

К настоящему времени накопилось достаточное количество данных инструментальных наблюдений за параметрами морской среды. Для Чёрного моря в открытом доступе находятся несколько массивов за период с 1884 по 2017 гг. Автором проводилось сравнение и объединение трёх баз данных, содержащих сведения о температуре вод Чёрного моря. В состав океанографической базы

данных больших морских экосистем (БМЭ) по Чёрному морю включено 86836 станций за период с 1884 по 2012 годы [3]. База данных Института природно-технических систем содержит 106170 станций с 1890 по 2013 годы [4]. База данных JCOMMOPS (Центр наблюдений за программами совместной технической комиссии по океанографии и морской метеорологии) съёмок с буёв Argo и других приборов по Чёрному морю содержит 18224 станций с 1998 по 2017 годы [5].

Цель работы – получение единой обновленной базы данных, включающей дублирующие станции, для последующего её использования в региональной океанографической ГИС и анализа гидрологических полей, их пространственно-временной изменчивости. В ходе работы были выполнены следующие задачи: загрузка в SpatiaLite более 7 млн. значений данных инструментальных наблюдений через интерфейс QGIS; проверка баз данных на погрешности, а также на целостность и взаимную совместимость; поиск и исключение полных и частичных совпадений по станциям измерений температуры вод Чёрного моря; анализ структуры полученной БД и предварительный контроль качества данных.

Материалы и методы. Загрузка данных в СУБД SpatiaLite через QGIS предполагает предварительный импорт текстовых файлов с данными и конвертирование их в формат .shp (shape file), принимаемый за стандарт во многих ГИС. После такой процедуры их может считать и перевести в свой формат хранения сам SpatiaLite, который оперирует не только точечными, но и линейными и площадными объектами при взаимодействии между таблицами, извлечении статистики и прочих расчётов.

Одна база данных позволила объединить 8 таблиц разной наполненности и структуры. С помощью конструктора SQL-запросов составлена последовательность программируемых текстовых обращений к таблицам баз БМЭ, ИПТС и Argo для их взаимного сравнения.

Первичная проверка баз данных на погрешности проводилась через серию запросов на наличие в базах пустых ячеек, где не указана глубина замера или нет записи в графах температуры и солёности, что дало возможность сократить объём памяти, занимаемой БД, а соответственно и ускорить в дальнейшем обращение к ней. Проверка на взаимную совместимость таблиц разных БД показала, что в БД ИПТС и БМЭ информация о названиях (или номерах) рейсов не совпадает, поэтому не может быть использована для сверки данных.

Поиск совпадений между базами проводился по таким параметрам, как географические координаты (точные и округленные значения), год, месяц, день и час съёмки. Также сравнивалось количество измерений в совпадающих станциях и значения температуры, снятые на них.

В структуре общей обновленной БД SpatiaLite были созданы дополнительные таблицы-представления, где хранятся метаданные и расчеты по исходным таблицам, например, сумма измерений по годам, сезонам, а также отдельным горизонтам и пр. Такие таблицы сохраняют только запрос к данным, но не сами данные. При этом изменения в содержании исходных таблиц находят отражение в представлениях, которые обновляются вместе со всей базой. Таблицы-представления при всём объёме инфор-

мации, извлекаемой ими из БД, позволяют поддерживать первоначальный объём базы, занимаемой ей на носителе. Это также свидетельствует об удобстве обработки данных через специализированную СУБД.

Предварительный контроль качества данных осуществлялся по гистограммам распределения значений температуры на контрольных глубинах с использованием модуля QGIS «Statist» [1]. Модуль позволяет не только визуально оценить гистограмму распределения частот оцениваемого параметра, но и получить значения таких статистических параметров, как минимальное и максимальное значения, ряд, среднее и медиана, стандартное отклонение и коэффициент вариации.

Дополнительно в ходе работы составлена новая векторная карта с линиями равных глубин на основе растровой батиметрической карты ЕТОРО [6] для Чёрного моря. Её разрешение составляет 2 минуты и позволяет получить достаточно подробную векторную карту изобат моря при условии контроля их качества. Отдельные изобаты затем использовались для контроля корректности координат точек измерений на соответствующих горизонтах, т.е. для пространственного контроля качества данных.

Результаты и их обсуждение. Сравнение баз показало следующие погрешности: пространственные при полном совпадении по дате (до точного времени) и округленным координатам; неполные записи графы «глубина моря на станции измерения» в разных базах; погрешности структуры БД – совпадающие по дате и координатам станции имеют разное количество горизонтов измерений (при совпадающих значениях температуры на одинаковых горизонтах).

В результате было удалено более 1 млн дублирующихся в трёх базах записей, а также пустые значения и точки, расположенные за пределами береговой черты (более 40 тыс. станций). В базу данных ИПТС добавлено 61 759 станций (4,6 млн измерений) с данными о температуре Чёрного моря на разных глубинах. Общий объём данных полученной базы составил 167 929 станций или 6,5 млн значений, что позволило увеличить объём БД Института в 3 раза. Структура

количественного распределения станций в трёх БД по годам представлена на рис. 1а. Там же присутствует итоговая обновленная БД, характеризующая распределение данных после сверки и объединения баз.

Как видно из рис. 1б, база данных Argo позволила существенно увеличить объём данных измерений за последние 12 лет, тогда как с конца 1990-х по 2017 гг. наблюдается значительный их спад по другим базам.

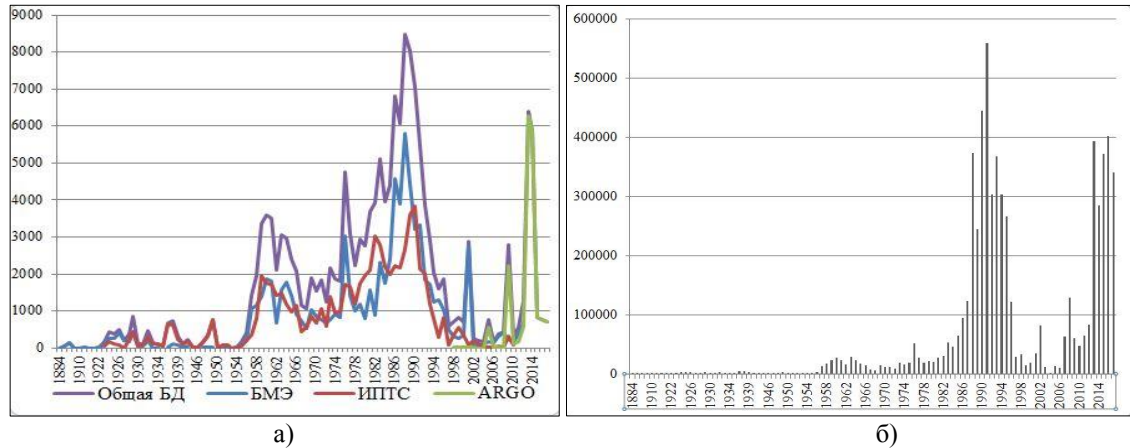


Рис. 1. Распределение количества станций (а) и измерений (б) по годам в базах данных (после сверки баз)

Также рассмотрено распределение количества измерений по различным горизонтам поверхностного (0, 50, 75 и 100 м) и глубоководного слоев (200, 300, 500 и 1000 м) в обновленной БД. В выборки по горизонтам попало следующее количество наблюдений: на поверхности – 134,6 тыс., на глубине 50 м – 57,3 тыс., 75 м – 44,9 тыс., 100 м – 43,7 тыс., 200 м – 39,9 тыс., 300 м – 37,3 тыс., 500 м – 23,9 тыс., 1000 м – 11,7 тыс. На рис. 2

представлено пространственное расположение точек измерений обновленной базы, где заметна значительная разница между поверхностным и глубоководным слоями, однако станции распределены равномерно по всей акватории на всех анализируемых горизонтах. Максимальная плотность данных – в прибрежной части, минимальная – в центре акватории.

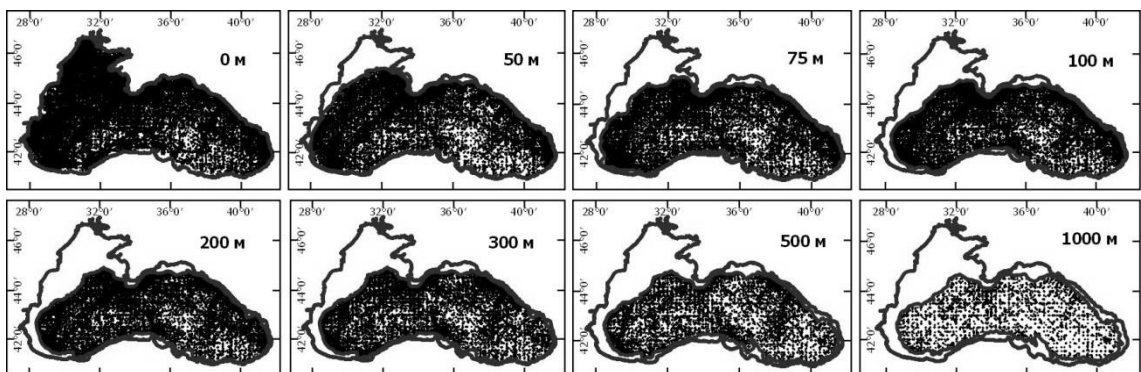


Рис. 2. Распределение точек измерений температуры вод Чёрного моря на горизонтах, м: 0, 50, 75, 100, 200, 300, 500 и 1000. Станции оконтурены изобатами соответствующей глубины

Распределение количества измерений обновленной базы по годам для каждого горизонта показано на рис. 3, где видно, что наибольший объём данных сосредоточен в пределах 1955–1995 гг. Отдельные пики приходятся на конец 80-х, а также середину 70-х гг. На горизонтах от 50 м значительный вклад в общий мас-

сив вносят данные с буев Argo, начиная с 2005 г. (в среднем около 800–1000 точек наблюдений за каждый год).

Небольшие отличия в сезонном ходе температуры между исходной БД Института и обновленной полной базой заметны в табл. 1, при одновременном уменьшении ошибки стандартного

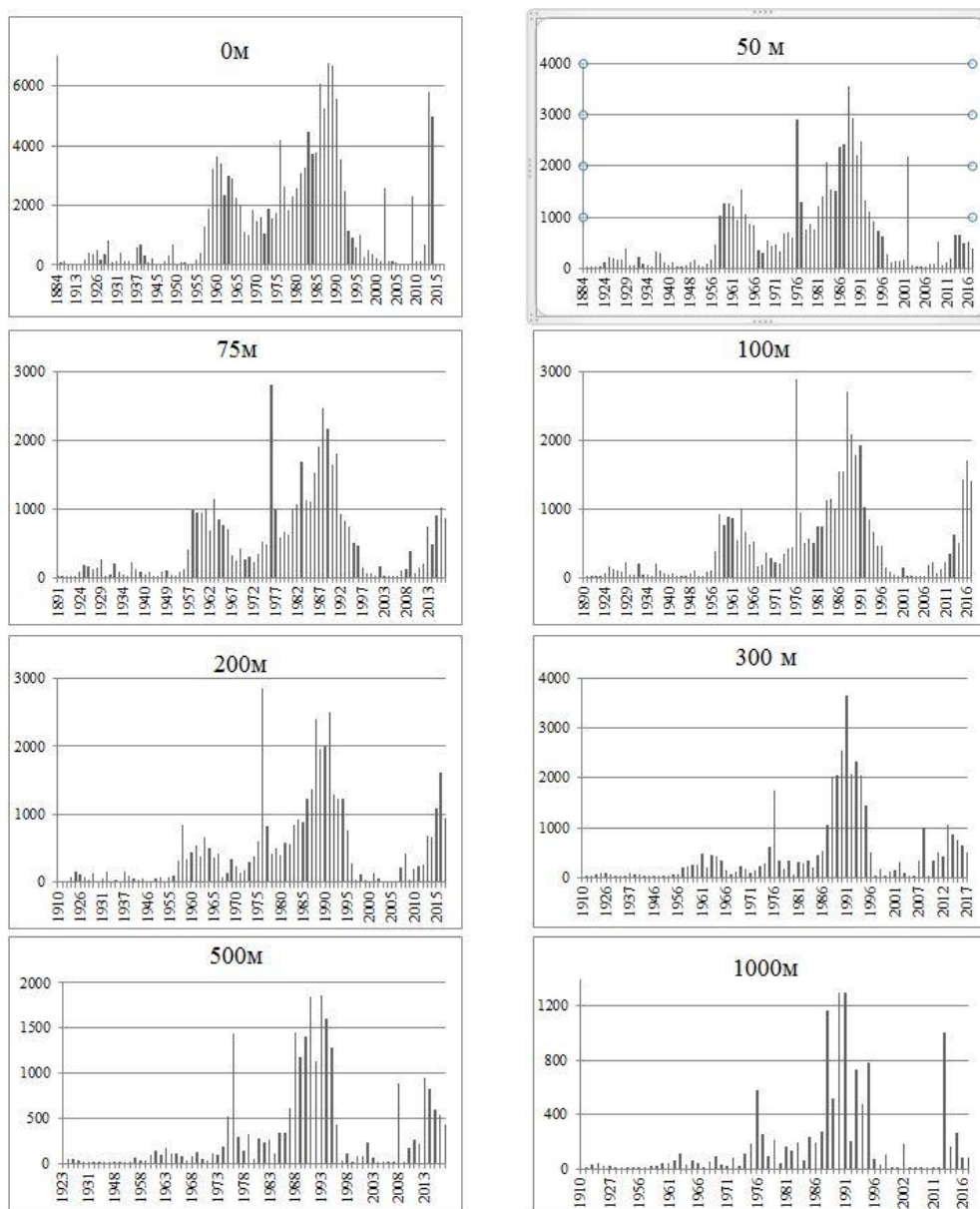


Рис. 3. Распределение количества измерений по годам на горизонтах, м: 0, 50, 75, 100, 200, 300, 500 и 1000

Таблица 1. Сезонный ход температуры на поверхности моря, его стандартные отклонения и количество измерений по месяцам в исходной БД Института и обновленной полной базе

Месяц	Общая БД			БД ИПТС		
	среднее	станд. откл.	кол-во измер.	среднее	станд. откл.	кол-во измер.
1	8.18	1.72	7794	8.17	2.09	3228
2	7.26	2.07	8797	6.86	2.15	5085
3	7.56	1.73	9967	7.28	1.81	6766
4	9.89	2.12	10089	9.68	2.01	7449
5	15.22	2.48	13797	15.16	2.47	11510
6	19.87	2.5	11869	19.80	2.54	10354
7	22.74	2.59	14738	22.72	2.7	11982
8	23.73	2.31	16509	23.56	2.36	12222
9	20.81	2.75	11028	20.77	2.84	9276
10	16.82	2.65	11555	16.85	2.8	9348
11	12.99	2.59	10326	12.70	2.68	7919
12	10.71	1.91	8068	10.74	2.11	4784

отклонения в данных и более равномерном распределении количества данных по месяцам в обновленной БД.

Заключение. В ходе работы разработана методика сравнительного анализа баз данных с использованием пространственной СУБД SpatiaLite и экспертной оценки, объединяющая широкий инструментарий ГИС по визуализации и пространственной обработке данных, а также быстрый доступ и управление ими в виде таблиц с помощью языка запросов SQL.

Сформирован новый обобщенный массив базы данных по температуре вод Чёрного моря за период 1884–2017 гг. (по данным гидрологических съёмок из различных источников) в структуре ПО ГИС QGIS. Заложены основы для дальнейшего развития региональной океанологической базы данных по Чёрному морю и комплексной геоинформационной системы анализа данных.

Общий объём данных обновленной базы по температуре вод моря составил 167 929 станций или 6,5 млн значений, что позволило увеличить объём БД Института в 3 раза.

СПИСОК ЛИТЕРАТУРЫ

1. https://docs.qgis.org/2.14/en/docs/user_manual/plugins/plugins_db_manager.html (дата обращения: 18.04.2018).
2. <https://www.gaia-gis.it/gaia-sins/> (дата обращения: 18.04.2018).
3. *Атлас* климатических изменений в больших морских экосистемах Северного полушария (1878–2013). Регион 1. Моря Восточной Арктики. Регион 2. Чёрное, Азовское и Каспийское моря / Г.Г. Матишов, С.В. Бердников, А.П. Жичкин [и др.]. Ростов н/Д.: Изд-во ЮНЦ РАН, 2014. 256 с.
4. *GIS* Института природно-технических систем / В.В. Мельников, А.Б. Полонский, А.А. Котолупова [и др.] // Системы контроля окружающей среды. Севастополь: ИПТС, 2016. Вып. 4 (24). С. 49–55.
5. <http://www.jcommops.org> (дата обращения: 18.04.2018).
6. <https://www.ngdc.noaa.gov/mgg/global/etopo2.html> (дата обращения: 18.04.2018).

THE USE OF DATABASE MANAGEMENT SYSTEMS AND GIS FOR DEVELOPMENT AND ANALYSIS OF THE REGIONAL OCEANOLOGICAL DATABASE FOR THE BLACK SEA

A.M. Novikova

Institute of Natural and Technical Systems,
Russian Federation, Sevastopol, Lenin St., 28

A new generalized array of the database on the temperature of the Black Sea waters for the period 1884–2017 was formed using GIS QGIS and spatial database management system SpatiaLite. There were used three databases: Institute DB, DB of Big Marine Ecosystems, as well as data from JCOMMOPS database containing information from the Argo buoys. Special data processing tools in SpatiaLite have significantly reduced the time required to access database requests while saving the amount of memory that the database uses. Attraction of additional databases allowed to increase the total volume of the Institute database 1.6 times by the number of stations and 3 times by measurements. Also, the volume of data of contact surveys for the last 10 years has significantly increased. Analysis of the spatial distribution of data on the main horizons (0, 50, 75, 100, 200, 300, 500 and 1000 m) showed the uniformity of their location throughout the Black Sea. A map of the depths of the Black Sea was created on the basis of the ETOPO raster bathymetric map.

Keywords: GIS, database management systems, the Black Sea, temperature.