

**БАНК ОКЕАНОЛОГИЧЕСКИХ
ДАННЫХ МГИ НАНУ:
СОДЕРЖАНИЕ И СТРУКТУРА БАЗ
ДАННЫХ, СИСТЕМА УПРАВЛЕНИЯ
БАЗАМИ ДАННЫХ**

*Суворов А.М., Андриющенко Е.Г.,
Годин Е.А., Ингеров А.В., Касьяненко Т.Е.,
Пластун Т.В., Халиулин А.Х.,*

Морской гидрофизический институт
НАН Украины
г. Севастополь, ул. Капитанская, 2
e-mail: suvorov@alpha.mhi.iuf.net

Рассматриваются структура современной версии БОД МГИ НАНУ и особенности программной системы управления, обеспечивающие способность к быстрым форматным преобразованиям данных и усвоению новых видов данных без перестройки системы управления, возможность проведения многопараметрических выборок и прямой доступ к сведениям о координатно-временном распределении данных, взаимодействие с международными системами сбора данных о морской среде.

Структура современной версии банка океанологических данных МГИ НАНУ формировалась на основе специализированных программных систем с учетом новых потребностей пользователей и необходимости осуществления международного и межведомственного обмена океанологическими массивами данных. Использование модельных и статистических подходов для диагностики состояния морской среды и составления прогнозов, необходимость анализа данных в различных пространственно-временных диапазонах привели к переходу от хранения данных в виде наборов файлов прямого доступа к разработке структуры связанных баз данных и программной системы управления ими. Участие в международном обмене данными показало необходимость унификации форматов хранения данных.

Данные измерений параметров морской среды, которыми располагает банк, получены не только в экспедициях судов МГИ НАНУ. Значительные массивы переданы научными организациями Украины, России, других государств (США, Болгарии, Турции, Румынии), Гидрографической службой Черноморского флота и МЦД (г. Обнинск).

Объемы океанологических данных и метаданных (служебная и справочная информация) в незначительной степени продолжают увеличиваться в результате проведения новых экспедиций. В настоящее время существенный вклад в пополнение банка вносит перенесение информации на магнитные носители из архивных источников и обмен по международным проектам. Примером успешного сотрудничества является реализация проекта MEDAR/MEDATLAS II, который осуществлялся под эгидой Комиссии Европейского Сообщества в течение 1999-2001г.г. с целью сохранения черноморских и средиземноморских данных по температуре, солености и биохимии. Восемь государств - участников проекта были представлены одиннадцатью мореведческими и двумя международными организациями. Значительный опыт по спасению и архивации данных, накопленный в МГИ НАНУ, определил роль института как координатора национального вклада Украины в реализацию проекта, причем отдел МИСТ являлся базовым по выполнению работ. В результате подготовлена к выпуску серия CD-ROM с данными, картами и документацией, охватывающая около 200 тысяч станций в Средиземном море и примерно 60 тысяч станций в Черном. Объем доступной пользователям банка информации существенно возрастает, особенно если учесть резко ограниченные финансовые возможности для организации экспедиционных работ.

Реорганизация структуры БОД и разработка соответствующей системы управления осуществляются одновременно с выполнением стандартных задач. Заполнение баз параметрами морской среды происходит с соблюдением определенных приоритетов. Естественный интерес к исследованию Черного, Азовского и Средиземного морей, а также исторически сложившееся внимание к сохранению результатов гидрологических и гидрохимических измерений определили первоочередное формирование соответствующих баз данных. В настоящее время гидрологическая база объединяет данные ~175 тысяч станций, ~140 тысяч из которых выполнено в Черном море. Гидрохимия представлена ~37 тысячами станций. Анализ наборов данных выявил большое разнообразие измеренных и вычисленных гидрооптических и гидробиологических параметров наряду с недостаточностью их

описания. С целью стандартизации форматов представления данных в базе оптических измерений проводятся совместные работы со специалистами отдела оптики, как исследователями и разработчиками измерительной аппаратуры. Основная часть хранящихся на магнитных носителях измерений скорости течений и метеорологических параметров уже переведена в форматы соответствующих баз. Метеорология представлена данными наблюдений МГИ за 1989-1995 г.г. и гидрометеорологической службы ЧФ (1929-1940, 1945-1982 г.г.). Базы течений охватывают данные, полученные в экспедициях НИС "Михаил Ломоносов" (600 буйковых постановок, 4357 горизонтов) и "Академик Вернадский" (418 буйковых постановок, 3524 горизонта). Однако значительные объемы измеренных параметров, которыми располагают научные фонды института, еще ждут своей очереди. Поэтому состав и объемы специализированных баз постоянно меняются (увеличиваются).

В настоящее время банк содержит примерно полторы тысячи наборов данных. Так как наблюдения производились различными приборами с использованием различных методик, исходные наборы разнородных данных хранятся в несопоставимых форматах. Для эффективного использования таких массивов осуществляется перенос данных в форматы СУБД Paradox, а система управления базами данных реализуется в среде программирования Delphi. Соединяя визуальные и объектно-ориентированные принципы программирования, Delphi предоставляет возможности для операций с различными типами баз данных и работы с изображениями, что позволяет сохранить преемственность по отношению к функциональным возможностям предыдущих версий банка и дополнить новыми функциями современную структуру, упрощенная блок-схема которой представлена на рис. 1.

Вспомогательные базы (батиметрия, границы географических районов, справочники среднемесячных значений некоторых параметров морской среды по различным районам и др.) используются для контроля данных при загрузке в базы и в других целях.

Дополнительная информационная база объединяет рейсовую базу о НИС и

рейсах МГИ НАНУ (195 экспедиций) и каталог с его базами-справочниками.

Окончательному оформлению существующей структуры банка предшествовали работы по систематизации океанологических данных, результатом которых было создание регулярно пополняемого единого каталога океанологических данных Украины [1]. Его третья версия включает описания всех типов имеющихся в организациях Украины данных, относящихся к ресурсам и состоянию морской и прибрежной среды и хранящихся на любых типах носителей. В основу информационной структуры каталога положен расширенный набор сведений, содержащийся в известной форме каталога морских данных ROSCOP. Важнейшими из дополнительных полей основной базы каталога являются признак качества информации (результат экспертной оценки для каждого вида данных) и указание пути доступа к ней (поле связи). Каталог - это программная система, которая к основной базе подключает справочные базы кодов измерений, государств и организаций - источников данных; описаний приборов и методов океанографических наблюдений; сведений о носителях измерительной и регистрирующей аппаратуры (в том числе о 220 судах, выполнявших океанографические исследования). Самостоятельная ценность каталога как источника полного представления о проведенных научных исследованиях очевидна [1,2], а содержащаяся в нем информация может служить целью запроса или его частью. Однако не менее важно, что структура каталога, объединившая и стандартизовавшая общие описания разнородных наборов данных, определила его место в составе банка данных в качестве вспомогательной информационно - справочной базы, предназначенной для оперативного получения сведений о содержании банка и формирования запросов для поиска и выделения фрагментов основной базы данных. Особо следует отметить составную часть каталога - справочную базу кодов наблюдений, подробно освещенную в работе [2]. При ее разработке была проделана огромная работа по классификации имеющихся данных и присвоению новых кодов тем видам измерений, которые не были охвачены формой ROSCOP (введено около 300 новых кодов). База включает около 400 видов наблюдений, разделенных на группы

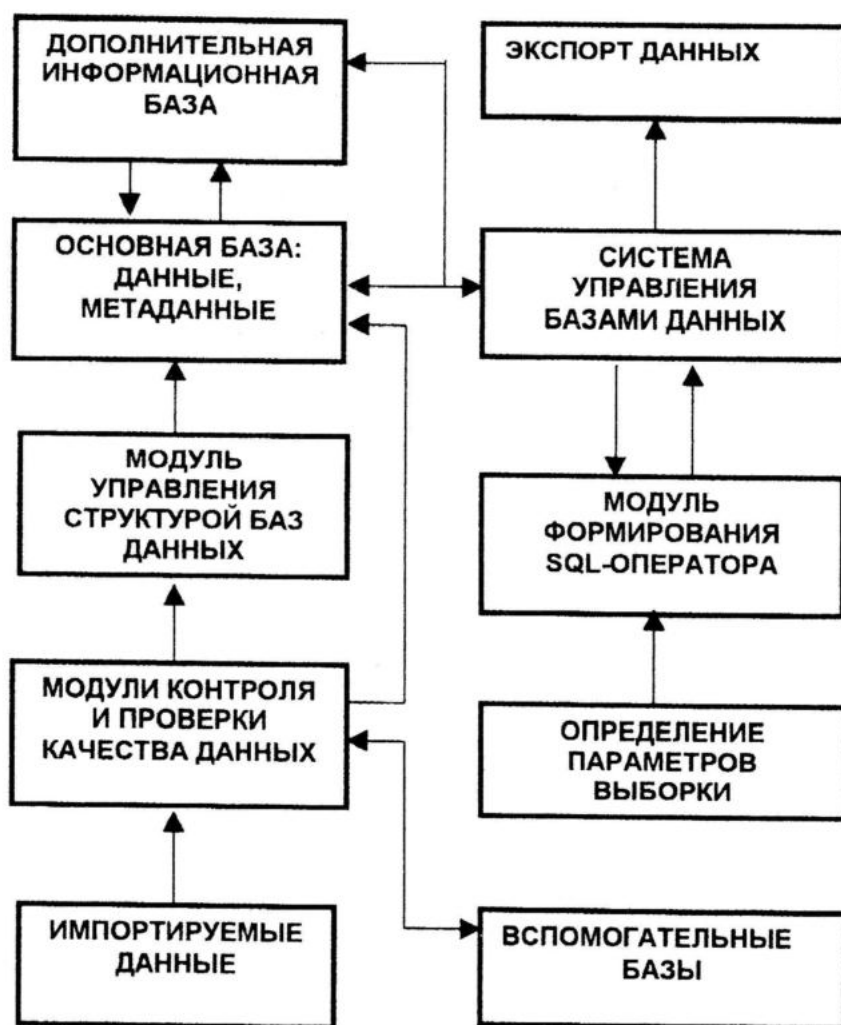


Рис.1 - Схема функционирования банка данных

по различным разделам океанологии. Наличие кода для каждого вида наблюдения дает возможность осуществлять не только поиск данных, но и выборку источников любого типа исследований. Пример их разнообразия - состав параметров метеорологической базы данных (см. таблицу 3).

Отдельным программным приложением является рейсовая база судов МГИ НАНУ, включающая подробные описания почти двухсот экспедиций, проведенных с 1957 по 2000г.г.[2]. Она содержит названия судов и номера рейсов, даты их начала и завершения, наименования научных проектов и имена исследователей, районы плавания, перечень проведенных исследований и время их выполнения, количество выполненных станций и их координаты.

Основная база данных объединяет специализированные базы данных и метаданных (гидрология, гидрохимия, загрязнения, оптико-биологические измерения, течения, метеорология).

Прежде чем перейти к описанию способа хранения данных, следует напомнить о стандартной схеме доступа к данным "станция - параметры", основанной на представлении наблюдений как совокупности произвольного количества измерений параметров в некоторых точках-станциях в пространственно-временной системе координат. Многие параметры, в свою очередь, являются функцией глубины (давления, условной плотности и т.д.). Базовым элементом, с которым обычно оперируют в таком случае при выполнении запроса, является станция. Очень часто это приводит к объе-

динению в одном файле данных измерений и метаданных, то есть к дублированию информации и существенному увеличению времени выборки. Таким образом, разработка единых форматов представления данных важна не только для обмена, но и обеспечивает следующие преимущества: унифицируется и сохраняется набор служебных и вспомогательных параметров, новые виды информации включаются в банк без изменения системы управления.

Для хранения числовых значений измеренных параметров и служебной информации используются таблицы Paradox, а размещение метаданных и данных измерений в разных базах (будем называть их информационными таблицами и таблицами данных) существенно сокращает объемы занимаемой памяти и качественно влияет на доступ к информации. Как правило, таблицы данных содержат основные первичные параметры, а вторичные вычисляются непосредственно во время выполнения выборок. Различным видам наблюдений (вертикальные профили, однократные измерения, временные ряды) соответствуют свои форматы хранения. Структура специализированных баз метаданных, содержащих набор обязательных служебно-информационных параметров, несколько различается между собой, однако полная идентичность полей пространственно – временной привязки является основой для объединения разнородной информации в одном запросе. В свою очередь, для всех видов измерений, проведенных в рейсах либо идентифицирующихся с неким номером станции, сформирована единая рейсовая база метаданных. Каждая запись этой информационной таблицы соответствует одной станции (либо зондированию, если их было несколько), то есть количество записей информационной таблицы соответствует количеству выполненных зондирований. Наличие поля - номера зондирования позволяет считать минимальной единицей запроса не станцию, а зондирование. Единая рейсовая база содержит время начала и завершения выполнения станции. Эта информация служит для временной привязки и идентификации вновь загружаемых данных, если точное время их получения не зафиксировано аппаратурой или записями в рейсовом журнале. Таким способом связываются измеренные параметры, относящиеся к данной станции. В

таблицах 1-5 представлена информационная структура некоторых тематических баз.

Поле связи между базами метаданных и измеренных параметров является ключевое поле, содержащее уникальный номер, присвоенный каждой записи информационной таблицы и повторяющийся в каждой записи данных, относящихся к этой станции (зондированию). Неизменность номера кода обеспечивает однозначное соответствие данных и метаданных. В свою очередь, для связи рейсовой базы с информационными таблицами используется несколько полей - кодов (название судна, номер рейса, номер станции и др.). Это позволяет оперативно организовать классическую схему доступа к данным "рейс – станция – параметры" и создает возможности для осуществления более сложных условий выборки [3].

Управление базами. Главным требованием, предъявляемым к программной системе управления данными, было соблюдение принципа открытости (способности системы к развитию и обмену информацией). В нашем случае это означает не только обеспечение пополнения данных и их обработки без изменения системы управления, но и возможности изменения структуры уже существующих баз с соответствующей коррекцией работы системы [4]. Управление базами данных может быть достаточно сложным. В этом смысле среда Delphi универсальна, так как является эффективной интегрированной средой для разработки программных приложений. Важным фактором с точки зрения совместимости форматов и поддержки языка SQL является использование СУБД Paradox. Ключевые поля (поля связи), заранее предусмотренные в составе баз, обеспечивает структурную целостность основной базы данных.

В основу работы системы положено взаимодействие нескольких самостоятельных программных модулей, унифицированная структура данных обеспечивает их передачу из модуля в модуль. Система дает возможность одновременной работы с несколькими базами, организует оперативный доступ к данным и справочной информации, расширяет возможности выборки и предусматривает выдачу экспертных оценок.

Программное обеспечение включает типовые алгоритмы контроля и согласова

Таблица 1 - Структура базы гидрологических метаданных.

№ поля	Имя поля	Тип поля	Содержание поля
Field1	CAST_N	Long Integer	Уникальный номер станции
Field2	SHIP	Char	Код судна
Field3	CRUISE	Char	Номер рейса
Field4	YEAR	Number	Год
Field5	MONTH	Number	Месяц
Field6	DAY	Number	День
Field7	TIME	Float	Время по Гринвичу
Field8	LAT	Float	Широта
Field9	LONG	Float	Долгота
Field10	QUADR	Number	Квадрат Марсдена
Field11	REG	Char	Регион (аббревиатура)
Field12	KIND	Char	Код измерения
Field13	DEPTH	Float	Глубина места выполнения станции
Field14	NUM_STA	Char	Рейсовый номер станции
Field15	NUM_ZOND	Number	Номер зондирования
Field16	HOR_T	Number	Кол-во горизонтов изм.температуры
Field17	HOR_S	Number	Кол-во горизонтов изм.солености
Field18	LAST_Z	Number	Максимальный горизонт с измерениями

Таблица 2 – Структура базы гидрологических данных

№поля	Имя поля	Тип поля	Содержание поля
Field1	CAST_N	Long Integer	Уникальный номер
Field2	HOR_CH	Number	Горизонт
Field3	T	Number	Температура
Field4	S	Number	Соленость

Таблица 3 - Структура базы метеорологических данных.

№поля	Имя поля	Тип поля	Содержание поля
Field1	CAST_N	Long Int	Уникальный номер
Field2	P1	Number	Атмосферное давление
Field3	P2	Number	Температура воздуха
Field4	P3	Number	Относительная влажность
Field5	P4	Number	Абсолютная влажность
Field6	P5	Number	Упругость водяного пара
Field7	P6	Number	Дефицит упругости
Field8	P7	Number	Температура точки росы
Field9	P8	Number	Направление ветра
Field10	P9	Number	Скорость ветра
Field11	P10	Number	Температура моря
Field12	P11	Number	Дальность видимости
Field13	P12	Number	Общее количество облаков
Field14	P13	Number	Количество облаков нижнего и среднего яруса
Field15	P14	Number	Форма облаков нижнего яруса
Field16	P15	Number	Форма облаков среднего яруса
Field17	P16	Number	Форма облаков верхнего яруса
Field18	P17	Number	Высота нижней границы облаков
Field19	P18	Number	Количество осадков
Field20	P19	Number	Направление волн зыби первой системы
Field21	P20	Number	Направление волн зыби второй системы
Field22	P21	Number	Высота волн зыби первой системы
Field23	P22	Number	Высота волн зыби второй системы
Field24	P23	Number	Период волн зыби первой системы
Field25	P24	Number	Период волн зыби второй системы
Field26	P25	Number	Высота ветровых волн
Field27	P26	Number	Период ветровых волн

Таблица 4 - Структура базы гидрохимических данных

№поля	Имя поля	Тип поля	Содержание поля
Field 1	CAST_N	Long Integer	Уникальный номер
Field 2	Depth	Number	Глубина
Field 3	O2	Number	Растворенный кислород
Field 4	PH	Number	pH
Field 5	Alk	Number	Щелочность
Field 6	PO4	Number	Фосфаты
Field 7	Total_P	Number	Суммарный фосфор
Field 8	SiO3	Number	Силикаты
Field 9	Si	Number	Кремний
Field 10	NO2	Number	Нитраты
Field 11	NO3	Number	Нитриты
Field 12	NO2_NO3	Number	Нитраты + нитриты
Field 13	Total_N	Number	Суммарный азот
Field 14	NH4	Number	Аммиак
Field 15	H2S	Number	Сероводород
Field 16	Chl_A	Number	Хлорофилл "А"
Field 17	PHEO	Number	Феофитин
Field 18	Okisl	Number	Окисляемость

Таблица 5 - Структура базы измерений течений

№поля	Имя поля	Тип поля	Содержание поля
Field	H	Integer	Горизонт
Field	V	Number	Скорость течения
Field	F	Integer	Направление течения
Field	T	Number	Температура

ния данных с последующим помещением в тематические базы; алгоритмы получения вторичных параметров и ряда расчетных характеристик морской среды; алгоритмы статистической обработки и спектрального анализа. В дальнейшем планируется включение в систему блока знаний, реализующего подключение моделей диагноза и прогноза состояния морской среды и передачу в модель выбранных данных.

Таким образом, можно говорить о современной версии банка океанологических данных как о программно управляемой структуре связанных баз, непрерывно пополняемой данными и наращиваемой по возможностям их анализа. Остановимся подробнее на некоторых функциях.

Загрузка данных является важным этапом формирования базы, обеспечивающим репрезентативность включенной информации. Измеренные параметры должны быть подвергнуты контролю не только для исключения ошибок, но и с целью предотвращения дублирования сведений, полученных из разных источников (в значитель-

ной части исходных массивов отсутствует привязка к судну или рейсу). На предварительной стадии осуществляется удаление явных сбоев, замена отсутствующих значений специальным кодом и другие рутинные операции, затем данные подвергаются другим видам анализа.

Модуль контроля осуществляет перевод данных в единые форматы и контроль их пространственно - временной привязки. Реализуются алгоритмы редактирования координат (градусы и минуты переводятся в десятичную систему счисления), проверки координат станций на попадание в контуры бассейна, оценки скорости судна на переходе между станциями, сравнения глубины места с батиметрическими данными (проверка глубины зондирования по изобатам). Визуальное определение ошибки сопровождается предоставлением нужной записи для исправления. После программного преобразования данных в соответствующие форматы заполняются поля тематических баз (либо эта операция выполняется вруч-

ную для впервые переносимых параметров).

Модуль проверки качества предусматривает более тщательный анализ данных, для чего используется программное обеспечение, реализующее алгоритмы проверки с использованием информации вспомогательных баз (справочники среднемесячных значений некоторых параметров морской среды и др.). Значения величин физических параметров проверяются на выход за границы возможного интервала их изменения (контроль границ допустимых значений); присваиваются признаки качества, характеризующие попадание параметров в интервалы $\pm 2\sigma$ и $\pm 3\sigma$; производится контроль статической стабильности гидрологических зондирований (отсутствие инверсий плотности).

Наличие типовых процедур не исключает применения к конкретным видам измерений специальных методов контроля и редактирования. В процессе формирования базы данных течений, содержащей в настоящее время около 600 тыс. записей метаданных и несколько миллионов записей данных (направление течения, скорость течения и температура воды на каждом горизонте), все исходные массивы прошли первичную проверку на соответствие форматов и попадание в интервал изменения. Однако основная проблема заключалась в установлении соответствия между результатами измерений и их пространственно-временной привязкой. Для этого по известной информации о времени начала и завершения работы измерителей, интервалах дискретности датчиков, с учетом остановок в работе измерителей и удаления сбойных записей, были рассчитаны даты и время измерений на каждом горизонте, а затем произведено программное редактирование данных и пересчет метаданных.

После формирования файлов производится их экспертная оценка с целью присвоения каждой записи признака качества (кода достоверности).

Модуль управления структурой обеспечивает включение новых данных в основную базу. Если для очередного вида измерений предусмотрены поля в структуре тематической базы, то записи попадают в соответствующую таблицу сразу после проверки качества. Если новые параметры или наборы данных не охвачены существ-

ующими таблицами, модуль либо изменяет структуру выбранных таблиц (вводит дополнительные поля для помещения новых параметров), либо создает новую таблицу с полями для размещения данных и полями связи.

Запрос - часть управления базами данных, осуществляемая в среде Delphi с помощью структурированного языка запросов SQL (Structured Query Language). Организация возможностей для оптимального формирования запроса является одной из главных задач, так как в большинстве случаев выборка - это подготовка информации для решения задач многопараметрического междисциплинарного анализа данных наблюдений. Разработанная программная система и выбранная структура баз данных обеспечивают разнообразные способы доступа к информации. Кодовые поля связывают различные базы посредством цепочки быстрых логических операций и формируют путь доступа к нужной информации, существенно снижая время выполнения запроса и делая его практически одноуровневым.

Модуль выборки включает визуальный построитель запросов, формирующий и исполняющий динамические SQL-операторы, и SQL - редактор, позволяющий накладывать условия на значения любых полей основной или справочной баз в различных комбинациях. Классическая схема доступа к данным "рейс - станция - параметры" легко реализуется, а данные из любых источников (как рейсовые, так и данные с неполным описанием) попадают в выборку по другим параметрам запроса. Программная система управления поддерживает одновременную работу с несколькими базами данных, позволяя выполнить комплексный проблемно-ориентированный запрос по указанной группе параметров. Выделение данных в заданном диапазоне глубин и по пространственно-временным критериям обеспечивает реализацию выборок типа "разрез" или "сечение" (нужный параметр на заданной глубине, снабженный координатами, дает файл для построения карт). Можно выбирать только реально измеренные данные либо использовать возможности интерполяции данных по глубине, то есть получать данные на тех глубинах, где не проводились измерения.

Модуль обработки данных предусмотрен для временного и пространственного осреднения и сглаживания, статистического и спектрального анализа. Соответствующие методы реализуются в виде динамических библиотек

Модуль визуализации обеспечивает построение карт, графиков, табличных зависимостей, масштабирование, определение координат и расстояний, наложение выбранных станций на географические контуры, построение полей в виде изолиний.

Экспорт данных, как следует из сказанного выше, возможен в различных типах форматов. В простейшем случае результатом выборки являются два выходных файла: справочно-информационный файл (записи метаданных с уникальным номером) и файл данных (каждая запись содержит набор параметров на одном горизонте, снабженный соответствующим уникальным номером).

В заключение следует отметить: опыт научно-производственной деятельности в области использования информации об окружающей среде показывает, что наибольшую отдачу можно получить, если деятельность по сбору, хранению, обработке и доведению до пользователей океанологических данных осуществляется в рамках единой информационной системы. Как ее часть развивается современная версия БОД МГИ НАНУ, реализующая программно управляемую информационную структуру, непрерывно пополняемую данными и наращиваемую по функциональным

возможностям. Порядок доступа к океанологической информации определен рядом документов, которые обеспечивают приоритет организаций - поставщиков данных на их использование в течение оговоренного начального срока и сохранение авторских прав за специалистами, выполнявшими измерения.

ЛИТЕРАТУРА

1. Андрущенко Е.Г., Халиулин А.Х. Единый каталог океанологических данных Украины. - Сб. конф. "Системы контроля окружающей среды", МГИ, Севастополь, 1998 г., С.131-133.
2. Суворов А.М., Андрущенко Е.Г., Островская И.Г., Пластун Т.В., Халиулин А.Х. Справочная база океанологических наблюдений, выполненных в научных рейсах МГИ НАНУ с 1957 по 1998 годы. - Сб. конф. "Системы контроля окружающей среды", МГИ НАНУ, Севастополь, 1999 г., С.177-180.
3. Суворов А.М., Халиулин А.Х., Пластун Т.В., Островская И.Г., Андрущенко Е.Г. Структура банка МГИ НАНУ. Сб. конф. "Системы контроля окружающей среды", МГИ НАНУ, Севастополь, 2001 г., С.194-198.
4. Суворов А.М., Халиулин А.Х., Ингеров А.В. Программное обеспечение базы океанологических данных МГИ НАНУ. Сб. конф. "Системы контроля окружающей среды", МГИ НАНУ, Севастополь, 2002 г., С.182-184.