

Н.А. ЖУКОВА, О.В. СМЕРНОВА
**МЕТОДЫ АДАПТИВНОЙ ОБРАБОТКИ И АНАЛИЗА
МЕТЕОРОЛОГИЧЕСКИХ И ОКЕАНОГРАФИЧЕСКИХ
ДАННЫХ В АРКТИЧЕСКОМ РЕГИОНЕ**

Жукова Н.А., Смирнова О.В. Методы адаптивной обработки и анализа метеорологических и океанографических данных в Арктическом регионе.

Аннотация. В настоящее время в арктическом регионе отмечается резкое изменение климата. Это связано со значительными изменениями, происходящими в окружающей среде – в атмосфере, океане, ледовом покрове. Научные учреждения Росгидромета, территориальные управления и региональные центры приема и обработки данных, системы мониторинга различных видов обстановки, в том числе ледовой, регулярно осуществляют наблюдения за состоянием окружающей среды арктического региона. Полученные измерения нуждаются в эффективной обработке и анализе для изучения, контроля и прогнозирования состояний окружающей среды.

В статье рассматриваются методы адаптивной обработки метеорологических и океанографических данных в рамках двух этапов – верификации и регуляризации данных. Приводится описание алгоритмов обработки для этих этапов.

Ключевые слова: обработка данных, анализ данных, алгоритмы анализа обработки данных.

Zhukova N.A., Smirnova O.V. Methods of Meteorological and Oceanographic Data Adaptive Processing and Analysis in Arctic Region.

Abstract. At the present time harsh climate change is noted in the Arctic region. This is due to significant changes in the environment – the atmosphere, the ocean, the ice cover. The research institute of Rosgidromet, territorial administration and territorial centers of data receipt and processing, monitoring system of different type of situation, including ice situation, are regularly observed for state of environment of the Arctic region. Receiving data needs in effective processing and analysis to study, control and predict of states of the environment.

The paper discusses methods of meteorological and oceanographic data adaptive processing and analysis for proposes of two stages – data verification and regularization. The description of processing algorithms for these stages is provided.

Keywords: data processing, data analysis, algorithms of data processing and analysis.

1. Введение. Для изучения, контроля и прогнозирования состояний окружающей среды Арктического региона необходима разработка эффективной автоматизированной системы сбора, обработки и анализа данных. Арктический регион достаточно специфичен с точки зрения обработки и анализа информации в связи с тем, что характеризуется экстремальными погодно-климатическими условиями. Особенности Арктического региона обуславливают сложность разработанных и используемых в настоящее время моделей и методов обработки и анализа данных и систем, которые их реализуют.

Современные системы обработки данных для Арктического региона узкоспециализированы, ориентированы на участие экспертов предметной области, и не обеспечивают комплексного анализа всех данных, получаемых из различных источников. В связи с постоянно увеличивающимся объемом данных происходит существенное увеличение времени, затрачиваемого на их обработку и анализ, что обусловлено недостаточной автоматизацией процесса обработки таких данных.

В результате конечные пользователи метеорологических и океанографических данных, которые не являются экспертами в предметной области и используют для решения прикладных задач готовые продукты анализа, не имеют возможности работы с актуальными данными. Продукты анализа, применяемые в рыболовной отрасли, службах, эксплуатирующих порты, корабли и т.д. обновляются примерно два раза в год. Отсутствие учета оперативных данных приводит к снижению точности получаемых прогнозных оценок, что для многих отраслей недопустимо.

Широкое использование методов ручной обработки и анализа данных приводит к тому, что часть ошибок, в первую очередь связанных с ошибками в работе измерительных устройств, устраняется значительно позже. Общее время выявления и устранения ошибок может составлять порядка двух лет, что связано с необходимостью проведения совместного анализа данных, получаемых от различных источников по одинаковым регионам.

Наиболее значимые проблемы, связанные с использованием существующих моделей, заключаются в том, что в настоящее время они перестали в полной мере удовлетворять предъявляемым к ним требованиям с точки зрения качества формируемых результатов. Причиной этого являются, в первую очередь, резкие изменения в климате, наблюдаемые в Арктическом регионе в течение последних лет.

Совокупность рассмотренных проблем обуславливают наличие объективной необходимости разработки новых методов и моделей обработки и анализа данных в Арктическом регионе.

Новые методы и модели обработки данных предлагается строить, в первую очередь, опираясь на получаемые по региону оперативные данные, при обработке которых, должен осуществляться учет накопленных исторических данных. Для организации эффективных процессов обработки данных необходимо применение адаптивных методов, основанных на комбинации существующих методов и моделей, а также эмпирических методов, использующих алгоритмы статистической

обработки и интеллектуального анализа. Разработка новых методов адаптивной обработки и анализа данных позволит повысить точность данных, предоставляемых конечным пользователям, за счет повышения оперативности обработки новых данных, автоматизации процесса обработки, учета специфики источников данных и районов получения данных.

2. Описание метеорологических и океанографических данных.

Исходные метеорологические и океанографические данные предоставлены Арктическим и антарктическим научно-исследовательским институтом, а также получены из различных центров обработки и анализа данных.

Метеорологические данные по арктическому региону представляют собой результаты наблюдений на станциях гидрометеорологической сети в синоптических сроки. Эти данные можно разделить на два типа: измерения и гридированные метеорологические данные.

Метеорологические измерения получают со станций всемирной метеорологической организации, относящихся к акватории Баренцева моря. Данные поступают со станций по каналам Автоматической передачи данных с 6-часовым интервалом в сроки 0, 6, 12 и 18 часов. Измерения включают информацию об атмосферном давлении на уровне моря, температуре и влажности на уровне 2 метров от поверхности, скорости и направлении ветра на уровне 10 метров от поверхности в текстовом формате. Пример метеорологических измерений приведен в табл. 1.

Таблица 1. Пример метеорологических данных

date	time	id	wind-dir	wind-speed	temperature	humidity	pressure
2009-01-01	00	20107	330	6	-17.0	60	1021.1
2009-01-01	00	22113	320	7	-7.0	96	999.4
2009-01-01	00	22028	50	15	-4.8	93	998.7
2009-01-01	00	20744	360	7	-7.5	71	987.9
2009-01-01	00	23024	260	3	-9.2	79	989.8
2009-01-01	03	20107	360	6	-16.7	60	1020.0
2009-01-01	03	22113	340	10	-6.8	83	1000.7
2009-01-01	03	23024	170	2	-14.4	85	990.0
2009-01-01	03	22292	80	6	-8.8	86	986.3
2009-01-01	03	22193	360	4	-11.2	62	989.0

Гридированные метеорологические данные формируются как результат реанализа [4]. Гридированные данные остаются неизменными на протяжении всего периода реанализа. Сбор данных осуществляется со всех доступных источников: наземных, воздушных, надводных и спутниковых, также производится контроль качества и ассимиляция этих данных системой CDAS (система усвоения климатических данных).

Гридированные метеорологические данные включают информацию об атмосферном давлении на уровне моря, температуре, влажности, составляющих скорости ветра на уровне сигма 0.995 на географической сетке с пространственным разрешением 2.5x2.5 градуса.

По Арктическому региону в состав доступных океанографических данных входят измерения температур и солёности, полученные в период с 1870 года по настоящее время от различных источников. К основным источникам океанографических данных относятся: данные с океанографических станций OSD и CTD, данные со стационарных батитермографов IMB, данные с привязанных батитермографов ХВТ, данные с закорных буев MRB, данные с ныряющих буев PFL, данные с дрейфующих буев DRB, данные с регистратора океанографических данных UOR, данные с автономных подводных аппаратов GLD.

По каждому измерению предоставляется следующий набор параметров: геопространственные – широта, долгота и глубина, на которых измерение было выполнено; временные – год, месяц, день и час проведения измерений; значение температуры, солёности и скорости звука в водной среде. Общее число выполненных измерений составляет более двух миллионов. Фрагмент файла с исходными океанографическими данными, приведенными к единому формату, показан в табл. 2.

На рис. 1 показано распределение количества измерений по географическим координатам. На рис. 2 приведено распределение по глубинам и месяцам проведения измерений.

3. Алгоритмы обработки и анализа метеорологических и океанографических данных. Анализ исходных метеорологических и океанографических данных показал, что измерения носят не регулярный характер, как по времени, так и по месту проведения измерений, за исключением данных, полученных со стационарных станций. Данные и их характеристики для различных регионов сильно отличаются и требуют особых решений для обработки. Кроме того, существенное влияние на полученные измерения оказывают внешние факторы, например, различные сезонные явления, состояния водной среды на границах смежных областей. Также данные содержат огромное коли-

чество ошибок, таких как шумы, выбросы, пропущенные значения, а также ошибки, связанные с погрешностями измерительных приборов и устройств. Все эти особенности необходимо учитывать при обработке и анализе метеорологических и океанографических данных.

Таблица 2. **Пример океанографических данных**

Latdeg	Longdeg	Year	Mnth	Dy	Hr	Sound	d/p	TEMP	PSAL
68.4	41.83333	1870	7	22	0	-9	0	9	-9
68.4	41.83333	1870	7	22	0	-9	36	6.4	-9
68.4	41.83333	1870	7	22	0	-9	45	5.7	-9
68.4	41.83333	1870	7	22	0	-9	54	5.6	-9
68.65	43.41667	1870	7	22	0	-9	0	7	-9
68.65	43.41667	1870	7	22	0	-9	45	4.9	-9
68.65	43.41667	1870	7	22	0	-9	55	4.7	-9
69.1	37.8	1870	7	31	0	-9	0	9.5	-9
69.1	37.8	1870	7	31	0	-9	36	6.9	-9
69.11667	38.18333	1870	7	31	0	-9	0	10.6	-9
69.11667	38.18333	1870	7	31	0	-9	36	7.4	-9
69.16667	38.91667	1870	7	31	0	-9	0	11.5	-9
69.16667	38.91667	1870	7	31	0	-9	36	6.4	-9
69.25	39.33333	1870	7	31	0	-9	0	11.1	-9

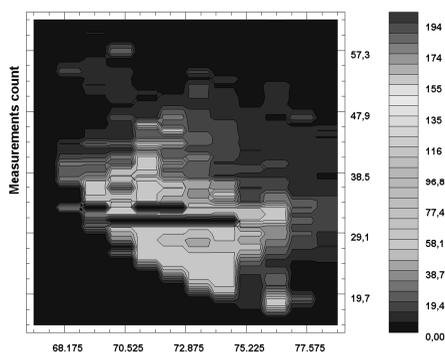


Рис. 1. Распределение количества измерений по географическим координатам.

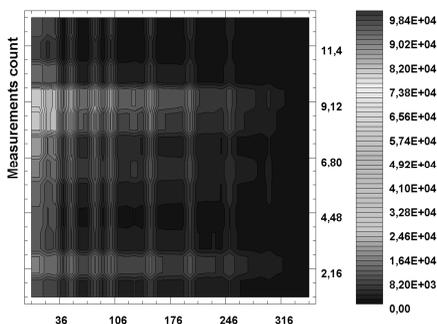


Рис. 2. Распределение количества измерений по глубинам и месяцам.

Обработка метеорологических и океанографических данных осуществляется в два этапа:

1. этап предварительной обработки (верификации) данных;
2. этап регуляризации данных.

Основной целью этапа верификации метеорологических и океанографических данных является систематизированное хранение, анализ и обработка полученных данных с целью их подготовки для решения задачи построения регулярных сеток данных.

На этапе верификации метеорологических и океанографических данных выполняются следующие шаги:

1. *гармонизация данных* обеспечивает преобразование данных из одного или нескольких форматов в формат единой модели предметной области [3];
2. *интеграция данных* предполагает решение задач оперативного контроля качества данных на основе выполнения набора специализированных тестов, а также выявление и исключение дублирующих значений;
3. *слияние данных* предполагает построение структурного представления наборов измерений, а также формирование образов измерений, как для различных источников данных, так и для различных районов, в которых получены данные [5]. При построении процессов обработки выбор алгоритмов осуществляется исходя из свойств обрабатываемых временных рядов, содержащих измерения.

Решение задач на этапе регуляризации метеорологических и океанографических данных обеспечивается за счет применения технологии слияния данных. Для слияния данных используются следующие алгоритмы: алгоритмы объективного анализа данных, обеспечивающие построение/уточнение регулярных сеток [2] (объективный анализ на

основе весовых функций, алгоритм оптимальной интерполяции, алгоритм на основе расчета коммулятивного взвешенного расстояния), алгоритмы реанализа данных, алгоритмы анализа полученных результатов, которые включают алгоритмы оценки продуктов анализа.

Для решения задачи автоматизации процесса обработки и анализа океанографических данных необходимо решить две основные задачи:

- на этапе верификации полученных данных обеспечить в кратчайшие сроки без привлечения экспертов решение задачи оценки их качества и возможности их дальнейшего использования для построения или уточнения узлов регулярной сетки;
- для каждого из этапов обработки в зависимости от состава обрабатываемых данных, источника данных и района их получения обеспечить выбор наиболее подходящего алгоритма обработки и анализа данных из набора реализованных алгоритмов обработки. Данная задача может решаться за счет описания правил применения алгоритмов и использования экспертной системы, обеспечивающей выполнение правил.

Для обеспечения оперативной автоматизированной верификации данных предлагается алгоритм оперативной обработки данных, основанный на применении алгоритмов интеллектуального анализа данных. Описание алгоритма приведено в листинге 1.

Входные данные:

1. массив исторических данных H , удовлетворяющих требованиям по качеству данных; каждый элемент данных представляет собой вектор, который описывает одно измерение в следующем признаковом пространстве: время (год, месяц, день), геопространственные координаты (широта, долгота, глубина) проведения измерений, значение измерения; в некоторых случаях целесообразно расширить признаковое пространство и рассматривать одновременно измерения нескольких параметров;
2. массив новых данных, для которых необходимо провести оценку качества $X = \{x_1, \dots, x_n\}$, где n - общее количество измерений.

Выходные данные:

1. вектор флагов $Q = \{q_1, \dots, q_n\}$, который определяет качество каждого из измерений;
2. список измерений D , которые являются дубликатами;
3. список измерений R , которые содержат ошибки.

Шаг 1 (обучение алгоритма)

1.1. (формирование компактных групп данных)

$f_{cl}(H) = C$, $C_T = \{c_1, \dots, c_k\}$, $H = \{h_{ij}\}$, $i \in (0, r)$, $j \in (0, m)$, где f_{cl} - алгоритм кластерного анализа; C - множество временных кластеров; k - общее количество кластеров; r - общее число накопленных измерений; m - количество признаков, описывающих измерения.

(анализ временных кластеров)

1.2. для $c_i \in C_T$

{если измерения в рамках кластера организованы компактно, кластер не требует дальнейшего разбиения и включается в итоговое множество кластеров}

если $|f_{cl}(c_i)| = 1$ **то** $C_R \leftarrow c_i$, где C_T - множество выходных кластеров
{кластер не является компактным и требует дальнейшей декомпозиции}

иначе $C_T \leftarrow f_{cl}(c_i)$

{расчет характеристик результирующих кластеров}

1.3. для $c_i \in C_R$

вычислить центры кластеров и стандартные отклонения значений измерений для каждого кластера

вычислить граничные значения B_i для измерений в сформированном кластере

Шаг 2 (применение алгоритма)

2.1. **определить** кластер $c_k \in C_R$, который соответствует региону, где проводились измерения $X = \{x_i\}_{i=0}^m$

{процедура определения граничных значений}

2.2. **выполнить** стандартную процедуру контроля качества данных с использованием тестов, определенных для примененного средства измерений, с использованием граничных значений B_k

заполнить вектор флагов Q с использованием результатов тестов контроля качества данных

{процедура выявления дублирующих значений}

2.3. **вычислить** $d = \min(\{dist(X, L_j)\})$, $j \in (0, \gamma)$, где γ - общее количество элементов в кластере c_k ; L_j - элемент из кластера c_k

если ($d < d_{гранч}$) **то** $D \leftarrow X$ дублирующее значение, где $d_{гранч}$ - априорно определенное граничное значение

{процедура удаления шума и выбросов}

2.4. $f_{cl}(X \cup \{L_j\}_{j=1}^{\gamma}) = C^E$, где L_j - элемент из кластера c_k ; C^E - множество полученных кластеров.

если ($\exists c_k \in C^E: |c_k| = 1$) **то** $R \leftarrow X$ считается, что значение измерений содержит ошибку

иначе значение измерения считается корректным

Листинг 1. Алгоритм оперативной обработки данных.

Флаги качества $q_i \in Q$ представляются в бинарной форме (тест пройден / тест не пройден) и дополнительно могут содержать информацию о расстоянии от анализируемого значения до ближайшего граничного значения. Список дублирующихся значений D также как и список измерений, содержащих ошибку, R может быть расширен информацией о расстояниях между анализируемыми измерениями и измерениями, которые определены как дублирующие.

Для профилей параметров, которые представляют собой временные ряды измерений, необходимо обеспечить анализ не только значений измерений, но также выполнить оценку динамики изменения временных рядов. Алгоритм оценки качества данных и выявления дублирующих значений для профилей показан в листинге 2.

Входные данные:

1. массив исторических профилей, которые были обработаны с привлечением экспертов предметной области и удовлетворяют требованиям по качеству данных: $F = \{f_j\}_{j=0}^n$, $f_j = \{(t_i, v_i)\}_{i=0}^m$, где f - профиль параметра, t_i - время и координаты получения i -го измерения, n - количество исторических профилей, m - длина профиля;
2. массив вновь полученных данных, для которых необходимо провести оценку качества $O = \{o_1, \dots, o_s\}$, где s - общее количество профилей, которые необходимо проанализировать.

Выходные данные:

1. вектор флагов $Q = \{q_1, \dots, q_s\}$, который содержит оценки качества измерений в каждом профиле.

Описание алгоритма

Шаг 1 (обучение алгоритма)

{процедура построения сегментов}

1.1. для каждого f_k вычислить $f_{segm}(f_k) = S_k$, где f_{segm} - алгоритм сегментации, $S_k = \{t_j, v_j\}_{j=1}^r$, r - количество сегментов

для каждого S_k , $k = 0..n$

для каждого $s_j \in S_k$, $j \in 1, \dots, r$, r - количество сегментов

{процедура построения формализованного описания сегмента}

$f_{forml}(s_j) = \bar{g}_j$, где f_{forml} - алгоритм построения формализованного описания, \bar{g}_j - вектор характеристик сегмента, $j = 0, \dots, h$ где h - количество характеристик

{процедура построения формализованного описания профилей}

вычислить $\bar{g}_j = \cup_j \bar{g}_j$

Шаг 2 (этап применения)

{процедура построения кластеров профилей с одинаковым поведением}

$f_{cl}(\cup_k \bar{g}_k) = \{Z_i\}_{i=0}^p$, где f_{cl} - алгоритм кластерного анализа, Z - описания кластеров с профилями, p - общее количество кластеров

Шаг 3 (этап оценки результатов)

для каждого $o_i \in O$

вычислить расстояние $d = \min(\{dist(o_i, \{Z_i\}_{i=0}^p)\})$

если $d < d_{\text{гранич}}$ то o_i имеет типовое поведение

иначе o_i имеет нетиповое поведение

заполнить вектор флагов на основе рассчитанных значений расстояний

Листинг 2. Алгоритм оперативной оценки качества временных рядов измерений параметров.

Алгоритм предполагает выполнение следующих основных шагов: построение формализованного описания профилей; построение шаб-

лонов профилей, описывающих типовые варианты поведения временных рядов измерений; анализ соответствия вновь полученных профилей сформированным шаблонам.

Сегментацию профилей целесообразно осуществлять в соответствии с границами водных слоев. Для построения формализованного описания сегментов целесообразно использовать интерполяцию кубическими сплайнами, поскольку временные ряды измерений хорошо описываются гладкими функциями. В листинге 3 представлен предлагаемый алгоритм кластерного анализа для работы с профилями.

Алгоритм кластеризации для применения его в алгоритме построения шаблонов должен обеспечить адаптивный выбор количества кластеров и при разбиении на кластеры учитывать априорные данные о возможных вариантах типового поведения. В качестве базового алгоритма используется алгоритм адаптивного выбора подклассов (АВП) [1]. Основное преимущество выбранного алгоритма состоит в том, что количество классов не фиксировано. На вход алгоритма подаются описания профилей параметров. В результате работы алгоритма кластеры формируются таким образом, чтобы они соответствовали вариантам типового поведения параметров.

Входные данные:

1. формализованные описания профилей $\{\bar{g}_k\}_{k=0}^n$, n - количество профилей, $MaxNCl$ - максимальное количество кластеров.

Выходные данные:

1. множество кластеров Z_j ; каждый кластер определяет одно из возможных типовых поведений параметра.

Описание алгоритма

Шаг 1 (инициализация алгоритма)

- 1.1. если (существуют профили, для которых известен тип)

то берется первый вектор \bar{g}_k из них

- 1.2. сформировать первый кластер Z_1 , включающий вектор описания профиля $\bar{g}_k = \langle g_k^1, g_k^2, \dots, g_k^n \rangle$ с центром $\bar{e}^{(1)} = \langle e_k^1, e_k^2, \dots, e_k^n \rangle$, где верхний индекс ⁽¹⁾ обозначает номер кластера

определить $N^{(1)} = 1$ - количество профилей в формируемом кластере

определить $N_{clust} = 1$ - количество кластеров

Шаг 2 (обработка векторов характеристик профилей)

для

найти кластер m , где расстояние w минимальное:

$$w: \|\bar{g}_j - \bar{e}^{(m)}\| \rightarrow \min, \text{ где } p_{jm} = \begin{cases} \infty, & \text{если профили не сравнимы} \\ \sum_{k=1}^h (g_{kj} - g_{km})^2, & \text{в других случаях} \end{cases}$$

h - общее количество характеристик, включенное в формализованное описание профиля. Профили считаются не сравнимыми, если в соответствии с априорными данными они принадлежат к различным типам

если $(w < r^{(u)})$ где $r^{(u)} = \left(\frac{1}{2}\sqrt{r^{(j)} + \sqrt{r^{(m)}}} + \sqrt{w(r^{(j)}, r^{(m)})}\right)^2$ – граничное значение для расстояния
то добавить вектор \vec{g}_j в кластер m
вычислить $\bar{e}^{(m)} = \frac{\bar{e}^{(m)}N^{(m)} + \vec{g}_j}{N^{(m)+1}}$
 $N^{(m)} = N^{(m)} + 1$
иначе
если $(N_{clust} < MaxNCl)$
то найти кластеры k_1 и k_2 с минимальным расстоянием $dist$
если $(dist < w)$
то слить кластеры k_1 и k_2
вычислить $r^{(k1)} = \left(\frac{1}{2}\sqrt{r^{(k1)} + \sqrt{r^{(k2)}}} + \sqrt{w(k_1, k_2)}\right)^2$
 $\bar{e}^{(k1)} = \frac{\bar{e}^{(k1)}N^{(k1)} + \bar{e}^{(k2)}N^{(k2)}}{N^{(k1)} + N^{(k2)}}$
 $N^{(k1)} = N^{(k1)} + N^{(k2)}$
где $N^{(k2)}$ – количество векторов в кластерах k_1 и k_2
создать новый кластер Z_j для профиля j
вычислить $N_{clust} = N_{clust} + 1,$
 $Z \leftarrow Z_j$
создать новый кластер Z_j для профиля j
вычислить $N_{clust} = N_{clust} + 1$
 $Z \leftarrow Z_j$

Листинг 3. Алгоритм сегментации временных рядов измерений параметров.

Для решения задачи кластерного анализа может использоваться любой другой алгоритм кластерного анализа, основанный на вычислении расстояний или на смеси распределений, не требующий априорной информации о количестве кластеров.

6. Заключение. В статье рассмотрены алгоритмы адаптивной обработки и анализа метеорологических и океанографических данных, основанные на применении алгоритмов интеллектуального анализа данных. Предложенные алгоритмы позволяют значительно повысить эффективность и оперативность обработки и анализа данных, что особенно важно для различных систем изучения, контроля и прогнозирования состояний окружающей среды.

Дальнейшие исследования связаны с применением методов бикластеризации для обработки и анализа метеорологических и океанографических данных. Данные методы позволяют учитывать различные факторы, влияющие на точность обработки данных, например, время и место полученных измерений. Кроме того, методы бикластеризации позволяют увеличить скорость обработки данных.

Литература

1. *Геппнер В.В.* Об одном подходе к задачам классификации, подчиняющихся мультимодальным распределениям / В.В. Геппнер, Г.М. Емельянов // Известия ЛЭТИ (Известия Ленинградского электротехнического института). СПб., 1969. № 85. С. 29–33.
2. *Кораблев А.А., Пнюшков А.В., Смирнов А.В.* Создание океанографической базы данных для мониторинга климата в Северо-Европейском бассейне Арктики // Труды ААНИИ. 2007. т. 447. С. 85-108.
3. *Попович В.В., Воронин М.Н.* Гармонизация, интеграция и слияние данных – три источника и три составляющих части геоинформационных технологий // Труды международного семинара «Интеграция информации и геоинформационные системы». СПб: СПИИРАН. 2005. С. 67–74.
4. NCEP/NCAR Reanalysis at PSD. URL: <http://www.esrl.noaa.gov/psd/data/reanalysis/reanalysis.shtml> (дата обращения: 23.04.2013).
5. *Zhukova N., Ignatov D., Smirnova O.* Dynamic information model for oceanographic data representation. Proceedings of workshop associated with the 20th International Conference on Conceptual Structures (ICCS 2013). 2013. Mumbai, India. P. 82–97.

Жукова Наталия Александровна — к.т.н., старший научный сотрудник научно-исследовательской лаборатории объектно-ориентированных геоинформационных систем СПИИРАН, ассистент кафедры математического обеспечения ЭВМ факультета компьютерных технологий и информатики С.-Петербургского государственного электротехнического университета (СПбЭТУ). Область научных интересов: методы статистической и интеллектуальной обработки сложных нестационарных процессов, проектирование и разработка распределенных систем обработки измерительных данных. Число научных публикаций — 40. gna@oogis.ru; СПИИРАН, 14-я линия В.О., д. 39, г. Санкт-Петербург, 199178, РФ; р.т. +7(812)328-3337, факс +7(812)328-4450.

Zhukova Nataly Alexandrovna — Ph.D., senior researcher, Research laboratory of object-oriented geo-information systems, SPIIRAS, lecturer, Software Department, Faculty of Computing Technologies and Informatics, St. Petersburg Electrotechnical University (SPETU). Research interests: methods of statistical and intellectual processing of non-stationary processes, development of distributed systems for measuring data processing. The number of publications — 40. gna@oogis.ru; SPIIRAS, 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone +7(812) 328-01-79, fax +7 (812) 329-08-63.

Смирнова Оксана Вячеславовна — к.т.н., старший научный сотрудник С.-Петербургский электротехнический университет «ЛЭТИ». Область научных интересов: математическое моделирование сложных процессов, геоинформационные технологии. Число научных публикаций — 40. sov@oogis.ru, www.oogis.ru; СПбЭТУ «ЛЭТИ», ул. Профессора Попова, д.5., РФ; р.т. +7(812)355-9682, факс +7(812)355-9674.

Smirnova Oksana Vyacheslavovna — Ph.D., senior researcher, St.Petersburg electrotechnical university «LETI». Research interests: mathematical modeling of complex process, geoinformation technology. The number of publications — 40. sov@oogis.ru, www.oogis.ru; SPb ETU "LETI", 5, str. prof. Popova, St.Petersburg, Russia;; office phone +7(812)355-9682, fax +7(812)355-9674.

Поддержка исследований. Работа выполнена в СПбГЭТУ при финансовой поддержке работ по проекту Министерством образования и науки Российской Федерации в рамках договора № 02.G25.31.0058 от 12.02.2013 г

Рекомендовано лабораторией объектно-ориентированных геоинформационных систем, заведующий лабораторией Попович В.В., д.т.н., проф.

Статья поступила в редакцию 31.07.2013.

РЕФЕРАТ

Жукова Н.А., Смирнова О.В. **Методы адаптивной обработки и анализа метеорологических и океанографических данных в Арктическом регионе.**

В настоящее время в арктическом регионе отмечается резкое изменение климата. Это связано со значительными изменениями, происходящими в окружающей среде – в атмосфере, океане, ледовом покрове. Научные учреждения Росгидромета, территориальные управления и региональные центры приема и обработки данных, системы мониторинга различных видов обстановки, в том числе ледовой, регулярно осуществляют наблюдения за состоянием окружающей среды арктического региона. Ежедневно в различные хранилища данных поступает более двух миллионов измерений, которые нуждаются в эффективной обработке для мониторинга и прогнозирования состояний окружающей среды.

Необходима разработка новых методов адаптивной обработки и анализа данных, которые позволят повысить точность данных, предоставляемых конечным пользователям за счет повышения оперативности обработки новых данных, автоматизации процесса обработки, учета специфики источников данных и районов получения данных.

Исходные метеорологические и океанографические данные предоставлены Арктическим и антарктическим научно-исследовательским институтом, а также получены из центров обработки и анализа океанографических данных.

Обработка метеорологических и океанографических данных осуществляется в два этапа:

1. этап предварительной обработки (верификации) данных;
2. этап регуляризации данных.

Основной целью этапа верификации метеорологических и океанографических данных является систематизированное хранение, анализ и обработка полученных данных с целью их подготовки для решения задачи построения регулярных сеток данных. Технология реализации этапа верификации получаемых от различных источников данных предполагает выполнение следующих элементов обработки: гармонизации данных, интеграции данных, слияния данных с использованием алгоритмов интерполяции данных, построение статистических моделей данных по различным районам.

Выполнение этапа регуляризации данных предполагает реализацию двух технологий: технологии построения общей статистической модели по накопленным измерениям, технологии построения регулярной сетки по накопленным статистическим данным, пересчета регулярной сетки с использованием процедур реанализа.

Необходимо дальнейшее развитие методов адаптивной обработки и анализа данных для увеличения оперативности и точности результатов.

SUMMARY

Zhukova N.A., Smirnova O.V. **Methods of Meteorological and Oceanographic Data Adaptive Processing and Analysis in Arctic Region.**

At the present time harsh climate change is noted in the Arctic region. This is due to significant changes in the environment – the atmosphere, the ocean, the ice cover. The research institute of the Rosgidromet, territorial administration and territorial centers of data receipt and processing, monitoring system of different type of situation, including ice situation, are regularly observed for state of environment of the Arctic region. Receiving data needs in effective processing and analysis to study, control and predict of states of the environment.

It is necessary development of new methods of the data adaptive processing and analysis. These methods will allow increasing the data accuracy, providing to end-users, by means of the rising of operational of new data processing, the automatization of processing, the accounting of specifics of data sources and areas of data acquisition.

Initial meteorological and oceanographic data was provided by Arctic and Antarctic Research Institute and also received from several other data acquiring and analysis centers. Data that is described and analyzed refers to the area of the Barents Sea.

Meteorological and oceanographic data processing is performed in two stages:

1. pre-processing (verification) stage;
2. regularization stage.

Main objectives of data verification are gathering, storage, processing and analysis received data so that it can be used for solving problem of building regular grids. On the verification stage from various data sources assumes performance of the following steps of processing: harmonization data, integration data, fusion data with using algorithms of data interpolation, creating of the statistical data models on various regions.

Performance of the regularization stage assumes realization of two technologies: technologies of creation of the general statistical model on the cumulative measurements, technologies of creation of a regular grid according to the saved-up statistical data, recalculation of a regular grid with use of reanalysis procedures.

It is necessary further development of methods of adaptive data processing and analysis for increase in operational and accuracy of results.