



Н.Н.Ташатов<sup>1</sup>, М.Е.Ашимова<sup>2</sup>

<sup>1,2</sup>Евразийский национальный университет им. Л.Н.Гумилева, Астана, Казахстан  
E-mail: <sup>2</sup>moldir\_kksu@mail.ru

## Использование Big Data и Data Analytics в интеллектуализации мониторинга и моделирования метеорологических систем

**Аннотация.** Основной целью данной работы является системный анализ современного состояния мониторинга (целей, возможностей, технологий) метеорологических данных, в частности, больших данных, аналитики (майнинга) и оперативно-аналитических витрин данных, их информационных срезов (профилей), онтологического представления интеллектуальной поддержки принятия метеорологических решений. Также рассмотрены задачи моделирования с использованием временных рядов в условиях их зашумленности (рассматривается лишь «белый», гауссов шум), консолидации данных и реализации моделей управления данными и процессом метеорологического мониторинга. Исследование использует методы анализа, декомпозиции, ситуационного моделирования, параметрической идентификации, математической статистики, классификации и риск-менеджмента. Основные результаты работы: 1) аналитика мониторинга как системной инфраструктуры «экосистемы» с учетом востребованности информационных объектов (трехуровневой для простоты); 2) модели вычислений, обеспечивающей хранение, анализ и адаптивность данных по предметным потребностям, «на лету»; 3) структурная онтологическая схема представления принятия решения на основе мониторинга, а также соответствующей ей формально-математическая модель. Полученные результаты в практическом мониторинге позволят решать как сложные задачи (на базе Big Data), так и задавать (оценивать) допустимые в мониторинге риски, строить алгоритмы идентификации и программы (планы) безопасного, управляемого мониторинга в условиях плохо определенных, нечетких ситуаций.

**Ключевые слова:** мониторинг; интеллектуальный; большие данные; аналитика; метеорология

DOI: [doi.org/10.32523/2616-7263-2023-145-4-9-22](https://doi.org/10.32523/2616-7263-2023-145-4-9-22)

### 1. Введение

Мониторинг климата, метеорологических ситуаций не только становится более динамичным, но и сложным, в системном понимании.

Термин «Big Data» (2008г., Nature) – интегральный, системный термин, объединяющий инновационные технологий представления и обработки, анализа сильно больших

массивов данных разной структуры, геолокации, хранения, которые ранее обработать было невозможно [1]. С привлечением смартфонов, компьютеров, приложений, в реальном режиме и с соблюдением требований «5V» (Volume, Variety, Velocity, Value, Veracity):

- 1) Volume – большие массивы, требующие совершенных технологий обработки;
- 2) Variety – параллельная обработка данных различных по форматам и способам управления (источникам, базам);
- 3) Velocity – рост скорости актуализации данных в реальном режиме;
- 4) Value – необходимо извлечь из данных все ценности, полезные свойства в них;
- 5) Veracity – соответствие данных и выводимых из них связей.

В метеорологии требуется мобилизация знаний [2], оптимизация метеоданных (рис.1).

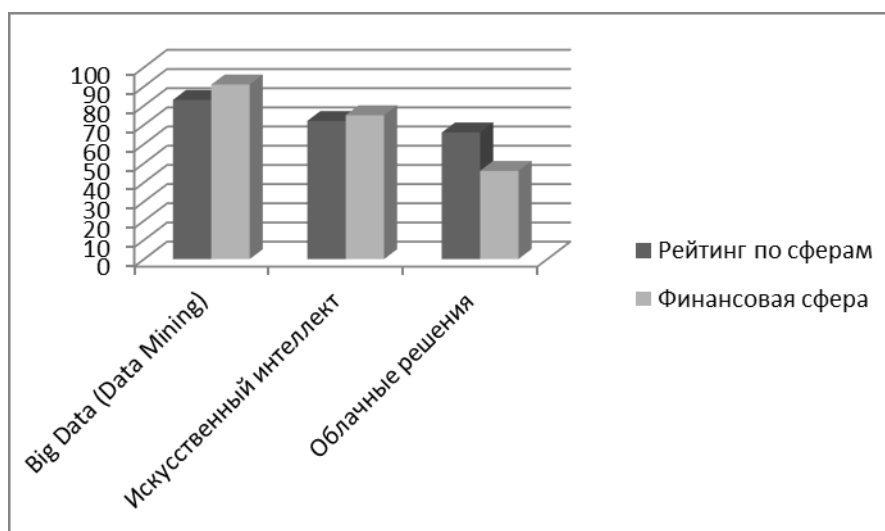


Рис.1. Сравнительный рейтинг Big Data по странам СНГ

Big Data в метеорологических системах позволяют исследователям сосредоточить внимание на данных, технологиях их обработки и анализа, в частности, Data Analytics, Data Mining. Например, Weather Company (частная компания от IBM) предлагает свыше 26 млн. ежедневных погодных сетевых прогнозов и инструментарий обоснования решений по погоде с личных метеостанций. Другой пример – Центр управления возобновляемой энергией (СЕКОЕР), предоставляющий решения в реальном удаленном режиме с помощью миллионов единиц климатических данных, в виде 3 тыс. графиков за год.

Обработка метеоданных потребует сравнительного анализа – с реальными или многолетними (сезонными) показателями погоды. Для этого используют часто несложные расчеты, например, температуры, силы ветра, осадков, их отклонений. Временным рядам данных свойственна зашумленность («белый, гауссов» шум). Он мешает консолидации данных – поиску корреляций, проверке статистических гипотез, исключению избыточности и др.

Временные ряды, отфильтрованные от «шумов» дают ценную информацию [3]. Но этого недостаточно, приходится применять API-сервисы [4] типа Метеосервис.ру, OpenWeatherMap, WorldWeatherOnline и более глубокий анализ с помощью фреймворка ApacheHadoop и на основе Big Data, Data Mining или нейронной сети [5] и глубокого машинного обучения. Например, сервис «Яндекс. Погода» научил нейросеть прогнозировать осадки на основе обучения с использованием Big Data Росгидромета на десятиминутный горизонт и строить карты осадков. Google применяет распределенную обработку на основе Big Data и компьютерных кластеров MapReduce.

В последнее время мониторинг метеорологических ситуаций использует не только Big Data, но и формализации описаний, представлений данных на основе онтологий и предиктивную аналитику [6]. Это не просто дань современным технологиям, а требование современного риск-менеджмента, упреждающей защиты от аварий, катастроф [7]. Так, в РФ эффективно используют Big Data метеорологических данных корпорации Сбербанк, Газпром-нефть, Северсталь, Аэрофлот, Русагро и др.

Обращение к Big Data в метеорологии стало обыденностью. Поэтому в работе мы проводим актуальный системный анализ и исследуем некоторые модели, подходы к принятию решений в метеорологии, в частности, мониторинга.

## 2. Методы

В нашем исследовании были использованы методы, способствующие моделированию метеорологических данных, а именно: методы анализа, декомпозиции, ситуационного моделирования, параметрической идентификации, математической статистики, классификации и риск-менеджмента.

## 3. Результаты и обсуждение

Реализуя цифровые трансформации в метеорологии приходится работать с Big Data, библиотеками [8]. Без релевантного аналитического обеспечения принятия решений, моделирования (например, ситуационного, визуального) невозможно обойтись. Эксперты используют разнообразный инструментарий на всех этапах цикла актуальности бизнес-данных.

Актуализируемые при этом данные хранятся, как правило, в транзакционных OLTP-системах (On-Line Transaction Processing Systems) [9], а также в файлах и внешних системах. Они могут быть и слабо структурированы, недостоверны, противоречивы, без предобработки.

Big Data в метеорологии дает преимущества:

- 1) упрощение мониторинга и его оперативный запуск;
- 2) рост пользователей данных, привлечение целевых клиентов;
- 3) ускорение взаимодействий с потребителями, поставщиками данных;
- 4) повышение темпа взаимодействий и др.

Big Data позволяет создавать и широко использовать профили потребителей, снижать издержки по обработке данных, адаптировать мониторинг к сложным условиям (например, COVID-19), сложному оборудованию.

Модели больших, корпоративных хранилищ метеорологических данных – многомерные, сложные при распределенной обработке в узлах кластеров данных, в понимании (восприятии) контекста данных, а также в динамической настройке под задачу, особенно, параметрически. Необходимы витрины данных, вычисления в памяти (In-Memory Computing, IMC), «на лету», использование ассоциативных моделей данных, предикативной аналитики. Их используем, адаптируем в данной работе.

### *Big Data и оперативные витрины данных в метеорологии*

Подход с использованием оперативных витрин актуализируемых метеоданных является частью концепции оперативно-аналитических витрин [10], направленной на доступность больших метеорологических данных и многих их источников посредством снижения объема и сложности данных. Например, сведением их к выборкам, позволяющим получать близкий к целевому, результат в режиме, близком к реальному.

Если рассматривать некоторый информационный объект, отражающий метеорологическое явление (метеопроцесс), то для целей информационно-логического (инфологического) моделирования можно опираться на профили (срезы) объекта, которые

представимы взаимосвязанными таблицами (тематическими, ориентированными на некоторую проблему, группу пользователей). Такие таблицы должны обладать свойством минимальной достаточности.

Пусть  $I = \{i_1, i_2, \dots, i_n\}$ ,  $N = |I|$  – элемент объекта, к которому требуется оперативный доступ для определенной группы метеорологических задач, потребителей. Информационный каталог ресурсов  $K = \langle I_1, I_2, \dots, I_N \rangle$ ,  $\forall i_j \in I_i$ , где  $I_i$  –  $i$ -ый информационный ресурс, в котором каждый  $j$ -ый элемент  $i_j$  принадлежит одному из классов – нормативно-справочной информации  $S$  или предметно-ориентированных данных  $D$ .

Востребованность (запросы) элементов ресурсов в течение периода  $T = \langle t_1, t_2, \dots, t_k \rangle$  разных групп пользователей – различна. Частоту обращений (запросов) к элементам  $K$  обозначим  $f_K(i_j)$ , она вычисляема (максимально) как:

$$f_{max} = \max_T(\langle f_{t_1}(i_j), \dots, f_{t_k}(i_j) \rangle), t_i \in T, l = 1, 2, \dots, k,$$

где  $f_{max}$  – максимум значений функций частот востребованности (запросов) к информационным ресурсам,  $f_{t_l}$  – частоты запросов к ресурсу  $i_j$  в период  $t_l$ .

Ранжирование по востребованности можно производить по формализованному индексу:

$$w_k = \frac{f_k}{f_{max}}.$$

По этому индексу можно фильтровать данные  $K$ , загружаемые в оперативное хранилище метеоданных.

Если определить шкалу востребованности элементов информационного каталога  $W_K = \{w_{k_i}\}$ , то можно построить множества (классы) востребованности элементов информационных ресурсов. Например, «трехклассовую» – высокочастотной (В), среднечастотной (С) и низкочастотной (Н) востребованности. Элементы В-класса включаются в оперативное хранилище всегда. Элементы С-класса могут быть включены лишь при достаточно свободных ресурсах оперативного хранилища. Элементы Н-класса не включаются в оперативное хранилище метеоданных.

Модели бизнес-аналитики используют различный аппарат, но наиболее часто – алгебраический, теоретико-множественный [11]. Модель вычислений позволяет хранить обрабатываемые данные в самой оперативной памяти [12], что даёт ключевые преимущества:

- 1) упрощение анализа данных (из-за сокращения уровней структурирования данных);
- 2) адаптивность модели (из-за подключения новых источников данных согласно текущим потребностям бизнеса, «на лету»);
- 3) повышение точности (из-за уменьшения фрагментации данных);
- 4) хранение активных структур данных и запросов в памяти, фильтрация (из-за отсутствия необходимости охранять промежуточные результаты на диске).

Для построения витрин метеорологических данных для заданного множества метеорологических процессов  $B_1, B_2, \dots, B_n$ , вначале идентифицируется классы (иерархии) атрибутов сущностей, соответствующие потребностям системных аналитиков. Уровни детализации данных определяются предметным аналитиком, метеорологом, агротехником или сейсмологом [13-14].

Для бизнес-процессов  $B_i, i = 1, 2, \dots, n$  строятся аналитические витрины:

- 1) оперативных метеоданных (из источников, оперативного хранилища), соответствующих метеопроцессов;
- 2) агрегированных данных, соответствующих обобщенным значениям атрибутов информационных объектов;
- 3) метаданных, соответствующих индексным связям содержимых оперативных метеорологических данных для связи данных между витринами.

Ядро системы оперативно, «на лету» строит ассоциативную модель, перестраивая ее адаптивно, интерактивно запросам пользователей без перезагрузки исходных данных. Если необходимо, с помощью частичной дозагрузки (инкрементирования) и подключения библиотеки класса Open Source (математика, статистика, нейро, визуализация и др.).

#### *Автоматизация и управление мониторинговой системой*

Автоматизация, а больше, интеллектуализация мониторинга, применение Big Data позволяют решать новые проблемы мониторинга, строить поддерживающие интеллектуальные системы.

Ключевые подсистемы такой системы определяем следующим образом:

- 1) БД (база данных) инфологических запросов, включающая единичные и повторяющиеся запросы пользователей, формирующая профили, планы, связи, задачи мониторинга и объектов, ситуаций (своеобразное «семантическое ядро» запросов);
- 2) БЗ (база знаний с экспертной системой) сценариев (планов) мониторинга;
- 3) библиотека (модули, процедуры, операторы), позволяющая реализовать план и программу мониторинга с отслеживанием автономности (модульности), параметризуемости, полноты (отсутствия дублирования) функций-задач, сценариев, адаптивной управляемости, устойчивости (надежности и корректности данных при наличии «шумов»), развиваемости и расширяемости решаемых задач;
- 4) интерфейсная подсистема (диалоговая реализация сценариев мониторинга, их визуализации, редактирования);
- 5) диспетчер – подсистема, динамически формирующая, согласно запросу и сценарию мониторинга цепочку и управляющая ее реализацией, как в последовательном, так и в параллельном режиме;
- 6) оценивания и интерпретации – для оценки успешности и интерпретируемости результатов мониторинга.

Мониторинговый процесс опирается (как в [15]), на базовые алгоритмические структуры и «примитивы» (аналоги макросов, функций пользователя).

Неопределенности, источники рисков являются атрибутами климатического мониторинга. Для оценки рисков привлекаются различные методы – экспертный, эвристический, статистический, нейросетевой и др.

Регулярный мониторинг – дорогостоящая процедура. На практике применяют мониторинг эпизодический, плановый. Потоки метеоданных и мониторинг позволяют охватить и все риски. Например, ключевые территориальные (региональные) риски:

- 1) обеспеченности ресурсами;
- 2) негативного воздействия («шумов»);
- 3) потери управляемости (планового развития, самоорганизации);
- 4) ИКТ-поддержки мониторинга и др.

Для каждого из рисков требуется апробированная процедура оценивания риск-напряженности (рис.2).

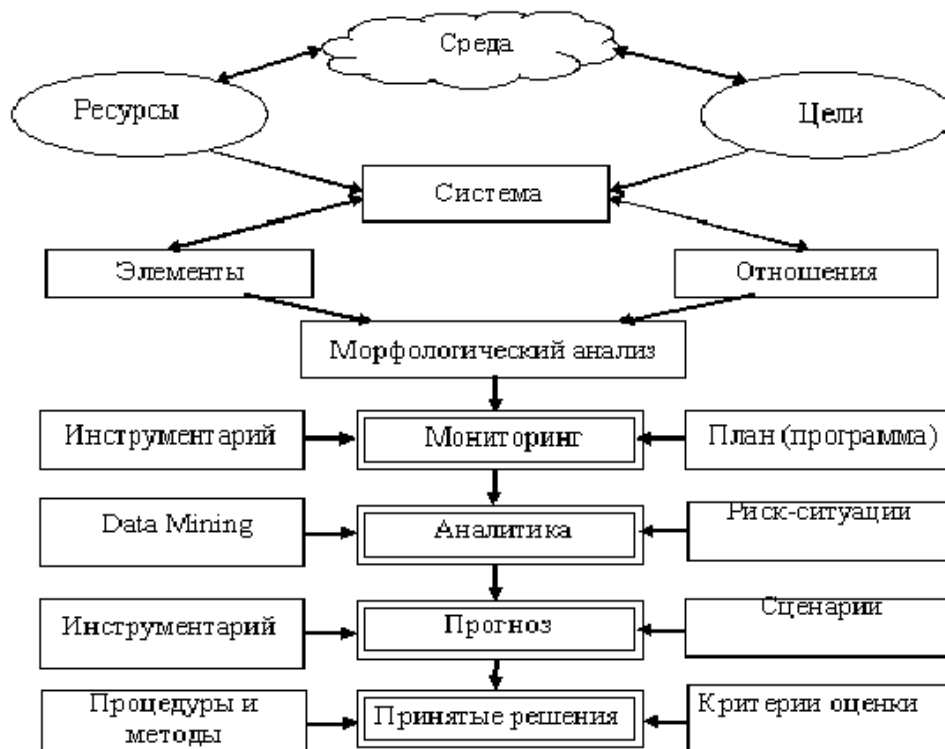


Рис. 2. Схема принятия решения на основе мониторинга

Некоторые параметры (показатели) ориентированы на экономику или технологии, не являются эффективными при интегральной метеорологической оценке системы, ее рисков. Необходим анализ взаимодействий для единой системы, методики идентификации параметров эволюции системы. К таким параметрам относятся, например, время, темп, мощность, скорость, емкость, затраты-доходы, полезность, эффективность, адаптивность и др.

Для оценки устойчивости метеорологического мониторинга можно ввести новые оценочные факторы, например, коэффициент качественного обслуживания:

$$K = S / \sum_{i=1}^n \alpha_i S_i,$$

где  $S$  – ресурсный, например, объемный или информационный показатель развития инфраструктуры,  $S_i$  – аналогичный показатель по инфраструктурному элементу (например, региональному),  $n$  – количество элементов (подсистем) системы,  $\alpha_i$  – весовой коэффициент, отражающий, в частности региональную обеспеченность (важность) фактора  $i$ .

В цифровом обществе, основанном на интеллектуальности и знаниях, для актуализации знаний применяется формальный когнитивный инструментарий – теория онтологии [16], фреймовая модель знаний [17].

Важно полностью, релевантно формализовать и идентифицировать область изменений метеорологических параметров, особенно, области рисков. Здесь полезно применение онтологий, онтологического представления системных процессов, среды, особенно, риск-ситуаций.

На рис. 3 показан пример онтологического представления ситуационного мониторинга.



Рис.3. Пример онтологического представления ситуационного мониторинга

Автоматизация мониторинга, а лучше, – его интеллектуализация (в смысле системной поддержки искусственного интеллекта) потребует использования новых, высоких ИКТ и неклассических инфологических систем. Должны быть учтены параметры метеорологических данных, информационных потоков (типы источников, достоверность, значимость, фильтрацию структурированность и др.).

Онтологии помогают в этом – создают системное представление предметно-ориентированных знаний, помогают описывать системные модели, процессы и структуры, семантические связи для последующего отображения их на предметную область, например, сейсмологию [18] или иную предметную область с учетом пространственных изменений [19].

Есть еще и интегрирующая способность онтология-ориентированного подхода, интеллектуального мониторинга и интерактивного анализа метеорологических данных, заключающаяся в усилении междисциплинарных взаимосвязей (особенно, обратных) исследователей. На основе исследования таких связей реализуемы унифицированные интеллектуальные платформы для реализации технологии гибкого, интерактивного анализа данных метеорологического мониторинга. Они позволяют, «вопреки» классическим алгоритмическим принципам и свойствам алгоритмов, на процедурном уровне анализировать системно-информационные связи. Например, как в программировании «на лету», потоковом, экстремальном. Это открывает пути реализации новых классов мониторинга сложных задач современной метеорологии.

В классическую схему отношений при решении задачи поддержки распределенного мониторинга вида: «конечный пользователь – метеорологическая система – эксперт (метеоролог, аналитик данных)» современные технологии и требования вносят изменения.

Рассмотрим схему типа «конечный пользователь – экспертная (решающая) система».

Выделим два типа знаний – знания о метеорологическом объекте  $O$  и знания системные  $S$ :

$$O = \langle Q, R, P, F \rangle, \quad S = \langle A, K, M, D \rangle,$$

где:

1)  $Q = \{Q_1, Q_2, \dots, Q_n\}$  – совокупность специализаций рассматриваемого метеорологического объекта  $O$ ;

2)  $R = \{R_{11}, R_{12}, \dots, R_{nn}\}$  – совокупность бинарных отношений с другими метеорологическими объектами;

- 3)  $P = \{P_1, P_2, \dots, P_n\}$  – совокупность разнородных свойств  $O$ ;
- 4)  $Q_i = \{P_{1i}, P_{2i}, \dots, P_{ni}\} \in P$ ;
- 5)  $F$  – совокупность правил вывода (интерпретации) с которыми связаны методы (алгоритмы) идентификации и оценки признаков  $O$ ;
- 6)  $A = \{A_1, A_2, \dots, A_m\}$  – знания об алгоритмах обработки данных;
- 7)  $K = \{K_1, K_2, \dots, K_k\}$  – знания о средствах, средах взаимодействия (особенно, при выборе конкретного алгоритма) с конечным пользователем;
- 8)  $M = \{M_1, M_2, \dots, M_l\}$  – знания о методах (алгоритмах) преобразования данных согласно актуальной логике цепочки обработки метеорологических данных;
- 9)  $D = \{D_1, D_2, \dots, D_d\}$  – знания о преобразуемых данных (моделях данных), в частности спутниковых, подвергаемых соответствующему преобразованию (вейвлет, Фурье и др.).

Используются набор адекватных инструментов, например, онтологии, графовые структуры объектов, веб-интерфейс, экспорт данных в базы (MySQL) и другие средства ситуационного моделирования по запросам пользователей и заданным сценариям метеорологического мониторинга.

Пример онтологического представления ситуационного мониторинга представлен в статье, далее дадим краткое описание одному из вышеперечисленных наборов адекватных инструментов. Так, например, известно, что для комплексного использования больших наборов метеорологических данных необходимо создать распределенную программную инфраструктуру, основанную на инфраструктуре пространственных данных (ИПД) [20]. Геопортал инфраструктуры пространственных данных представляет собой, по мнению ученых, единую точку входа, предполагающую функциональности поиска географических информационных ресурсов, выборки данных, согласно заданным параметрам (функциональность доступа к данным), а также обработки и картографической визуализации в виде соответствующих сервисов и клиентских приложений [21].

На данный момент считается, что разработка клиентских приложений как элементов такой инфраструктуры должна выполняться с использованием современных веб- и ГИС-технологий [22]. В соответствии с требованиями директивы, INSPIRE к визуализации пространственных данных, приложение должно обеспечивать базовую функциональность стандартной ГИС, выражающуюся в просмотре данных, прокрутке, масштабировании и наложении графических слоёв, а также в отображении соответствующих метаданных.

Исходя из этого отметим, что на сегодня существует несколько информационных систем и сервисов, предоставляющих подобную функциональность. Система GeoBrain Online Analysis System (GeOnAS) предоставляет доступ к данным спутниковых наблюдений (NASA, USGS) через сервисы Open Geospatial Consortium (OGC, <http://www.opengeospatial.org>), построенные на базе ПО с открытым кодом GRASS GIS, и оснащена веб-интерфейсом, основанным на библиотеке DHTMLX (<http://dhtmlx.com>).

Сервис ncWMS [23] – это реализация сервиса OGC Web Map Service (WMS) для геопространственных наборов данных, представленных в формате netCDF. Он активно используется для визуализации данных в рамках геопорталов ИПД, но, к сожалению, слабо поддерживается стандартными ГИС. Портал Unidata THREDDS (<http://www.unidata.ucar.edu/software/thredds/current/tds/TDS.html>) предоставляет доступ к геопространственным данным и метаданным по OPeNDAP, OGC WMS и OGC Web Coverage Service (WCS). Этот продукт также поддерживает выборку данных с использованием ncWMS для визуализации результатов. Открытая распределённая архитектура Boundless / OpenGeo широко используется для разработки сложных геоинформационных приложений [24]. Она состоит из трёх уровней (данные, сервер приложений и графический интерфейс) и опирается на следующее открытое ПО: ПО Geoserver и Geowebcache (<http://geoserver.org>), реализующее сервисы OGC WMS, WFS, Web Processing Service (WPS); JavaScript-библиотеку OpenLayers (<http://openlayers.org>), которая обеспечивает базовую функциональность «тонкого» веб-



ГИС клиента; JavaScript-библиотеку GeoExt / ExtJS library [25] для разработки клиентских веб-приложений с интуитивно понятным графическим интерфейсом пользователя. Рассмотренные подходы нами были применены в разработке схемы хранения больших метеорологических данных, создании база метаданных, а также графического веб-ГИС клиент-пользователя.

#### 4. Выводы

Системный подход к мониторинговым исследованиям в метеорологии необходим не только для решения практических проблем, но и для создания методик, принципов эффективного и интеллектуального мониторинга, подающихся оцениванию (аудиту) заранее, априори, для планирования.

Идентификация метеорологических событий (особенно, рисков) должна основываться на достоверных данных, аналитических процедурах. Здесь без Big Data (Data Analytics) и разработки математической и аналитической поддержки эффективно решать метеорологические задачи в реальном режиме не обойтись. Понадобится Data Mining и распределенная обработка («облака», «туманы»). Возможность исследования «глубинных», не лежащих на поверхности метеорологических данных, сетей усиливают Data Mining, Social Mining, BI и др. Наиболее интересные процессы и отклики среды в исследованиях можно анализировать, зная лишь место, время и импульс отклика, управляя активностью в среде, эволюционным потенциалом системы. Для этого потребуются инфологическое и программно-технологическое обеспечение, соответствующие междисциплинарные модели, ситуационное моделирование, эффективный доступ к распределенным и разнородным данным от различных классов источников.

Необходимы инфраструктура и сервис, позволяющие не только актуализировать, но и обобщать, систематизировать данные, формируя онтологическую модель процессов, ситуаций и их отношений. Интегрированная интеллектуальная система мониторинга позволит актуализировать не только текущие метеоданные (например, спутниковые), но и Big Data.

Наш подход и предложенные модели используются для ситуационного (планового, программируемого) мониторинга с минимизацией рисков и выбором приемлемой риск-траектории, позволяющей оптимизировать, перераспределять ресурсы. Возможности Big Data позволяют производить глубокий статистико-аналитический мониторинг, поддерживаемый данными и алгоритмами.

Big Data инфраструктура – фактор преодоления «разрывов» в проводимых мониторинговых исследованиях в эпоху Big Data. Таков эволюционный процесс.

#### Список литературы:

1. Peng G., Downs R., Lacagnina C., Ramapriyan H., Ivanova I., et.al. Call to Action for Global Access to and Harmonization of Quality Information of Individual Earth Science Datasets // *Data Science Journal*, 2021, vol.20(1), p.19-29. DOI: <http://doi.org/10.5334/dsj-2021-019>
2. Baker K., Duerr R., Parsons M. Scientific Knowledge Mobilization: Co-evolution of Data Products and Designated Communities // *International Journal of Digital Curation*, 2016, Vol.10(2), pp.110–135. DOI: <https://doi.org/10.2218/ijdc.v10i2.346>
3. Колесников И.Н. Прогнозирование временных рядов посредством привязки событий // *Моделирование, оптимизация и информационные технологии*. 2019. –№7(4). –С.12–21. DOI:10.26102/2310-6018/2019.27.4.039
4. Переpletчиков В.И., Носков В.Ю. Разработка информационной системы сбора и анализа данных метеорологических сервисов на основе технологий Big Data // *Теплотехника и информатика в образовании, науке и производстве : сборник докладов ТИМ'2018 (VII Всероссийской научно-практической конференции студентов, аспирантов и молодых учёных с международным участием)*,

17–18 мая 2018 г. –Екатеринбург: ООО АМК «День РА», 2018. –С.487-493. <https://elar.urfu.ru/handle/10995/61614?mode=full>

5. Костромин Н.С., Сивова А.Н. Перспективы применения нейросетей для решения проблем ННН-рыболовства и пиратства в Арктической зоне России // Российская Арктика. –2020. –№11.- С.24-30 <https://cyberleninka.ru/article/n/perspektivy-primeneniya-neyrosetey-dlya-resheniya-problem-nnnrybolovstva-i-piratsstva-v-arkticheskoy-zone-rossii/viewer>

6. Колесников И.Н., Финогеев А.Г. Проактивный мониторинг событий на основе предиктивного анализа временных рядов // Модели, системы, сети в экономике, технике, природе и обществе, 2020, вып.1, с.111-125. DOI 10.21685/2227-8486-2020-1-9

7. Bakhmut A.D., Krylov A.V., Krylova M.A., Okhtilev M.Yu., Okhtilev P.A., Sokolov B.V. Proactive Management of Complex Objects Using Precedent Methodology // Artificial Intelligence and Algorithms in Intelligent Systems (R.Silhavy, ed.). Proceedings of 7th Computer Science On-line Conference 2018. –2018. –Vol.2. –pp.298-307. <https://www.semanticscholar.org/paper/Proactive-Management-of-Complex-Objects-Using-Bakhmut>

8. Эмиров Н.Д., Батталова С.С. Информационные услуги в современном информационном обществе: роль библиотек и их корпораций // Экономика и предпринимательство. 2017, №11(88), с.894-897 <https://www.elibrary.ru/item.asp?id=32407663>

9. Parekh A. Introduction on Data Warehouse with OLTP and OLAP. International Journal of Engineering and Computer Science, 2(08). – 2017. Retrieved from <http://www.ijecs.in/index.php/ijecs/article/view/1819>

10. Раевич А.П., Добронец Б.С. Разработка концептуальной модели оперативно-аналитических витрин данных // Моделирование, оптимизация и информационные технологии. 2019, №7(4). -с.1-13. [https://moit.vivt.ru/wp-content/uploads/2019/11/Raevich\\_4\\_19\\_1.pdf](https://moit.vivt.ru/wp-content/uploads/2019/11/Raevich_4_19_1.pdf). DOI:10.26102/2310-6018/2019.27.4.002.

11. Ахрем А.А., Рахманкулов В.З., Южанин К.В. О сложности редукции моделей многомерных данных // Искусственный интеллект и принятие решений, 2016, №4. -с.79-85. [http://www.isa.ru/aidt/images/documents/2016-04/79\\_85.pdf](http://www.isa.ru/aidt/images/documents/2016-04/79_85.pdf)

12. Moving Towards Real-Time Analytics: All About In-Memory Computing and Self-service BI // Financial and credit activity: problems of the theory and practice, 2019, №1(28). -pp.272-278. [https://medium.com/@infopulseglobal\\_9037/moving-towards-real-time-analytics-all-about-in-memory-computing-and-self-service-bi-bac2e1936efd](https://medium.com/@infopulseglobal_9037/moving-towards-real-time-analytics-all-about-in-memory-computing-and-self-service-bi-bac2e1936efd)

13. Brezhnev R.V., Maltsev E.A. An Ontological Spatial Monitoring System for Agricultural Land Monitoring // Pattern Recognition and Image Analysis, 2015, vol.25, No.2, pp.201-208. <https://doi.org/10.1134/S1054661815020054>

14. Брагинская Л.П., Григорюк А.П., Ковалевский В.В. Онтологический подход к организации информационной поддержки исследований в активной сейсмологии. // Тезисы межд. конф. «Марчуковские научные чтения-2020» (МНЧ-2020, 19-23 октября 2020). –Новосибирск. -С.68-69. DOI:10.24411/9999-017A-2020-10369

15. Fomells M., Blasco D.M.J., Densnos Y-L., et al. ESA SNAP-Stamps Integrated Processing for Sentinel-1 Persistent Scatterer Interferometry // IGARSS, 2018, -pp.1364-1367. DOI:10.13140/RG.2.2.25803.90405

16. Раевич К.В., Маглинец Ю.А., Брежнев Р.В. Рассмотрение подходов к представлению знаний об оперировании пространственными данными в задачах мониторинга // Обработка пространственных данных в задачах мониторинга природных и антропогенных процессов (SDM-2019), Институт вычислительных технологий СО РАН, 2019. - С.172-177. <https://scholar.sfu-kras.ru/publication/41376555>

17. Бурцев М.А., Успенский С.А., Крамарева Л.С., Антонов В.Н., Калашников А.В., Балашов И.В., Кашницкий А.В., Лупян Е.А., Матвеев А.М., Прошин А.А. Современные возможности и перспективы развития Объединённой системы распределённой работы с данными НИЦ «Планета» // Современные проблемы дистанционного зондирования Земли из космоса. 2019, т.16, №3. -С.198-212. DOI:10.21046/2070-7401-2019-16-3-198-212.

18. Gruber T.R. The role of common ontology in achieving sharable, reusable knowledge bases // Principles of Knowledge Representation and Reasoning. Proceedings of the Second International Conference (J.A. Allen, R. Fikes, E. Sandewell, eds.). -Morgan Kaufmann, 1991, pp. 601-602. <https://dl.acm.org/doi/10.5555/3087158.3087222>

19. Chen J., Cohn A.G., Liu D., Wang S., Ouyang J., Yu Q. A survey of qualitative spatial representations. Knowledge Engineering Review, 2013, 30(1). –pp.106-136. <https://doi.org/10.1017/S0269888913000350>
20. Steiniger S., Hunter A.J.S. Free and open source GIS software for building a spatial data infrastructure. / In: Bocher E., Neteler M., (eds.), Geospatial Free and Open Source Software in the 21st Century, LNCG, Heidelberg, Springer, 2012. P. 247-261
21. Koshkarev A.V. Geoportal as a tool to control spatial data and services. // Spatial data. 2008. № 2. P. 6-14.
22. Yakubailik O.E. Geoformation geoportal // Computational Technologies. Special Issue 3. 2007. V. 12. P. 116-125.
23. J.D. Blower, A.L. Gemmell, G.H. Griffiths, K. Haines, A. Santokhee, X. Yang. A Web Map Service implementation for the visualization of multidimensional gridded environmental data // Environmental Modelling & Software. 2013. V. 47. P. 218-224.
24. Becirspahic and A. Karabegovic. Web portals for visualizing and searching spatial data // Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2015. Opatija, Croatia. P. 305-311.
25. Титов А.Г., Гордов Е.П., Окладников И.Г. Разработка Веб-ГИС на основе сервисов обработки и визуализации пространственных данных для анализа и прогнозирования региональных климатических изменений // Информационные и математические технологии в науке и управлении. 2016. № 4-2. С. 96-109.

## Метеорологиялық жүйелердің мониторингі мен модельдеуін зияткерлік етуде Big Data және Data Analytics қолдану

Н.Н.Ташатов<sup>1</sup>, М.Е.Әшімова<sup>2</sup>

<sup>1,2</sup> Д.Н.Гумилев атындағы Еуразия ұлттық университеті, Нұр-Сұлтан, Қазақстан  
E-mail: <sup>1</sup>tash.nur@mail.ru, <sup>2</sup>moldir\_kksu@mail.ru

**Аңдатпа.** Бұл жұмыстың негізгі мақсаты метеорологиялық деректердің, атап айтқанда, үлкен деректердің, аналитиканың (майнингті) және оперативті-аналитикалық деректердің, олардың ақпараттық кесінділерінің (профильдерінің) мониторингінің ағымдағы жағдайын (мақсаттары, мүмкіндіктері, технологиялары) жүйелі және метеорологиялық шешімдерді қабылдау үшін интеллектуалды қолдаудың онтологиялық көрінісін талдау болып табылады. Уақыт қатарларын олардың шулылығы жағдайында (тек «ақ», гаусс шуылдары қарастырылады), деректерді біріктіру және деректерді басқару модельдерін енгізу және метеорологиялық мониторинг үдерісінде модельдеу мәселелері қарастырылады. Зерттеуде талдау, декомпозиция, ситуациялық модельдеу, параметрлік сәйкестендіру, математикалық статистика, классификация және тәуекелдерді басқару әдістері қолданылады. Жұмыстың негізгі нәтижелері: 1) ақпараттық объектілерге сұранысты ескере отырып, «экожүйенің» жүйелік инфрақұрылымы ретінде аналитика мониторингі (қарапайымдылығы үшін үш деңгейлі); 2) деректердің сақталуын, талдауын және субъектінің қажеттіліктеріне сәйкес бейімделуін қамтамасыз ететін есептеу моделі, «жұмыс үстінде»; 3) мониторинг негізінде шешім қабылдауды ұсынудың құрылымдық онтологиялық схемасы, сондай-ақ оған сәйкес формалды математикалық модель. Тәжірибелік мониторингте алынған нәтижелер күрделі мәселелерді де (үлкен деректер Big Data негізінде) шешуге де, мониторингте қолайлы тәуекелдерді орнатуға (бағалауға), нашар жағдайда қауіпсіз, бақыланатын мониторинг үшін сәйкестендіру алгоритмдері мен бағдарламаларын (жоспарларын) құруға анықталған, анық емес жағдайларда мүмкіндік береді.

**Түйін сөздер:** мониторинг; интеллектуалды; үлкен деректер; аналитика; метеорология

## Using Big Data and Data Analytics to intelligently monitor and model meteorological systems

N.N.Tashatov<sup>1</sup>, M.E.Ashimova<sup>2</sup>

<sup>1,2</sup> L.N.Gumilyov Eurasian National university, Nur-Sultan, Kazakhstan

E-mail: <sup>1</sup>tash.nur@mail.ru, <sup>2</sup>moldir\_kksu@mail.ru

**Annotation.** The main purpose of this work is a systematic analysis of the modern state of monitoring (goals, capabilities, technologies) of meteorological data, in particular, big data, analytics (mining) and operational-analytical data storefronts, their information sections (profiles), ontological representation of intellectual support for meteorological decision making. The tasks of modeling using time series in conditions of their noisiness (only “white,” Gaussian noise is considered), data consolidation and implementation of data management models and meteorological monitoring process are also considered. The study uses methods of analysis, decomposition, situational modeling, parametric identification, mathematical statistics, classification and risk management. The main results of the work: 1) monitoring analytics as a system infrastructure of the “ecosystem,” taking into account the demand for information objects (three-level for simplicity); 2) a model of calculations that ensures the storage, analysis and adaptability of data according to subject needs, “on the fly”; 3) a structural ontological diagram of the presentation of decision making based on monitoring, as well as a formal mathematical model corresponding to it. The obtained results in practical monitoring will allow you to solve both complex tasks (based on Big Data), and set (evaluate) the risks allowed in monitoring, build identification algorithms and programs (plans) for safe, controlled monitoring in conditions of poorly defined, fuzzy situations.

**Keywords:** monitoring; intelligent; big data; analytics; meteorology

### References:

1. Peng G., Downs R., Lacagnina C., Ramapriyan H., Ivanova I., et.al. Call to Action for Global Access to and Harmonization of Quality Information of Individual Earth Science Datasets // *Data Science Journal*, 2021, vol.20(1), p.19-29. DOI: <http://doi.org/10.5334/dsj-2021-019>
2. Baker K., Duerr R., Parsons M. Scientific Knowledge Mobilization: Co-evolution of Data Products and Designated Communities // *International Journal of Digital Curation*, 2016, Vol.10(2), pp.110–135. DOI: <https://doi.org/10.2218/ijdc.v10i2.346>
3. Kolesnikov I.N. Prognozirovanie vremennyh ryadov posredstvom privyazki sobytij [Forecasting time series by linking events] // *Modeling, ipsum et notitia technology [Modeling, optimization and information technology]* 2019. –№7(4). –S.12–21. DOI:10.26102/2310-6018/2019.27.4.039.
4. Perepletchikov V.I., Noskov V.YU. Razrabotka informacionnoj sistemy sbora i analiza dannyh meteorologicheskikh servisov na osnove tekhnologij Big Data [Development of an information system for collecting and analyzing data of meteorological services based on Big Data technologies] // *Teplotekhnika i informatika v obrazovanii, nauke i proizvodstve : sbornik dokladov TIM'2018 (VII Vserossijskoj nauchno-prakticheskoy konferencii studentov, aspirantov i molodyh uchyonyh s mezhdunarodnym uchastiem), 17–18 maya 2018 g. [Heat engineering and informatics in education, science and production : collection of reports TIM'2018 (VII All-Russian Scientific and Practical Conference of students, postgraduates and young scientists with international participation)]*–Ekaterinburg: OOO AMK «Den' RA», 2018. –S.487-493. <https://elar.urfu.ru/handle/10995/61614?mode=full>
5. Kostromin N.S., Sivova A.N. Perspektivy primeneniya nejrosetej dlya resheniya problem NNN-rybolovstva i piratstva v Arkticheskoy zone Rossii [Prospects of using neural networks to solve the problems of IUU fishing and piracy in the Arctic zone of Russia] // *Rossijskaya Arktika*. –2020. –№11. –p.24-30 <https://cyberleninka.ru/article/n/perspektivy-primeneniya-neyrosetey-dlya-resheniya-problem-nnrybolovstva-i-piratstva-v-arkticheskoy-zone-rossii/viewer>
6. Kolesnikov I.N., Finogeev A.G. Proaktivnyj monitoring sobytij na osnove prediktivnogo analiza vremennyh ryadov [Proactive monitoring of events based on predictive analysis of time series] *Modeli, sistemy, seti v ekonomike, tekhnike, prirode i obshchestve [Models, systems, networks in economics, technology, nature and society]*, 2020, vyp.1, s.111-125. DOI 10.21685/2227-8486-2020-1-9

- 7 Bakhmut A.D., Krylov A.V., Krylova M.A., Okhtilev M.Yu., Okhtilev P.A., Sokolov B.V. Proactive Management of Complex Objects Using Precedent Methodology // Artificial Intelligence and Algorithms in Intelligent Systems (R.Silhavy, ed.). Proceedings of 7th Computer Science On-line Conference 2018. –2018. –Vol.2. –pp.298-307. <https://www.semanticscholar.org/paper/Proactive-Management-of-Complex-Objects-Using-Bakhmut>
8. Emirov N.D., Battalova S.S. Informacionnye usluzhi v sovremennom informacionnom obshchestve: rol' bibliotek i ih korporacij [Information services in the modern information society: the role of libraries and their corporations] // *Ekonomika i predprinimatel'stvo* [Economics and entrepreneurship] 2017, №11(88), s.894-897 <https://www.elibrary.ru/item.asp?id=32407663>
9. Parekh A. Introduction on Data Warehouse with OLTP and OLAP. *International Journal of Engineering and Computer Science*, 2(08). – 2017. Retrieved from <http://www.ijecs.in/index.php/ijecs/article/view/1819>
10. Raevich A.P., Dobronec B.S. Razrabotka konceptual'noj modeli operativno-analiticheskikh vitrin dannyh [Development of a conceptual model of operational and analytical data showcases]. // *Modelirovanie, optimizaciya i informacionnye tekhnologii* [Modeling, optimization and information technologies] 2019, №7(4). -s.1-13. [https://moit.vivt.ru/wp-content/uploads/2019/11/Raevich\\_4\\_19\\_1.pdf](https://moit.vivt.ru/wp-content/uploads/2019/11/Raevich_4_19_1.pdf). DOI: 10.26102/2310-6018/2019.27.4.002.
11. Ahrem A.A., Rahmankulov V.Z., YUzhanin K.V. O slozhnosti redukcii modelej mnogomernyh dannyh [On the complexity of the reduction of multidimensional data models // *Iskusstvennyj intellekt i prinyatie reshenij* [Artificial intelligence and decision-making], 2016, №4. - p.79-85. [http://www.isa.ru/aidt/images/documents/2016-04/79\\_85.pdf](http://www.isa.ru/aidt/images/documents/2016-04/79_85.pdf)
12. Moving Towards Real-Time Analytics: All About In-Memory Computing and Self-service BI // *Financial and credit activity: problems of the theory and practice*, 2019, №1(28). -pp.272-278. [https://medium.com/@infopulseglobal\\_9037/moving-towards-real-time-analytics-all-about-in-memory-computing-and-self-service-bi-bac2e1936efd](https://medium.com/@infopulseglobal_9037/moving-towards-real-time-analytics-all-about-in-memory-computing-and-self-service-bi-bac2e1936efd)
13. Brezhnev R.V., Maltsev E.A. An Ontological Spatial Monitoring System for Agricultural Land Monitoring // *Pattern Recognition and Image Analysis*, 2015, vol.25, No.2, pp.201-208. . <https://doi.org/10.1134/S1054661815020054>
14. Braginskaya L.P., Grigoryuk A.P., Kovalevskij V.V. Ontologicheskij podhod k organizacii informacionnoj podderzhki issledovanij v aktivnoj seismologii [An ontological approach to the organization of information support for research in active seismology] // *Tezisy mezhd. konf. «Marchukovskie nauchnye chteniya-2020»* [Abstracts of the international conference “Marchuk scientific readings-2020”] (MNCH-2020, 19-23 oktyabrya 2020). –Novosibirsk. -C.68-69. DOI:10.24411/9999-017A-2020-10369
15. Fomells M., Blasco D.M.J., Densnos Y-L., et al. ESA SNAP-Stamps Integrated Processing for Sentinel-1 Persistent Scatterer Interferometry // *IGARSS*, 2018, -pp.1364-1367. DOI:10.13140/RG.2.2.25803.90405
16. Raevich K.V., Maglinec YU.A., Brezhnev R.V. Rassmotrenie podhodov k predstavleniyu znaniy ob operirovanii prostranstvennymi dannymi v zadachah monitoringa [Consideration of approaches to the representation of knowledge about the operation of spatial data in monitoring ] // *Obrabotka prostranstvennyh dannyh v zadachah monitoringa prirodnyh i antropogennyh processov* [Processing of spatial data in monitoring tasks of natural and anthropogenic processes] (SDM-2019), Institut vychislitel'nyh tekhnologij SO RAN, 2019, s.172-177. <https://scholar.sfu-kras.ru/publication/41376555>
17. Burcev M.A., Uspenskij S.A., Kramareva L.S., Antonov V.N., Kalashnikov A.V., Balashov I.V., Kashnickij A.V., Lupyan E.A., Matveev A.M., Proshin A.A. Sovremennye vozmozhnosti i perspektivy razvitiya Ob'edinyonnoj sistemy raspredelyonnoj raboty s dannymi NIC «Planeta» [. Modern opportunities and prospects for the development of a unified system of distributed work with the data of SIC “Planet” ]// *Sovremennye problemy distancionnogo zondirovaniya Zemli iz kosmosa* [Modern problems of remote sensing of the Earth from space], 2019, t.16, №3. -S.198-212. DOI:10.21046/2070-7401-2019-16-3-198-212.
18. Gruber T.R. The role of common ontology in achieving sharable, reusable knowledge bases // *Principles of Knowledge Representation and Reasoning. Proceedings of the Second International Conference* (J.A. Allen, R. Fikes, E. Sandewell, eds.). -Morgan Kaufmann, 1991, pp. 601-602. <https://dl.acm.org/doi/10.5555/3087158.3087222>
19. Chen J., Cohn A.G., Liu D., Wang S., Ouyang J., Yu Q. A survey of qualitative spatial representations. *Knowledge Engineering Review*, 2013, 30(1). –pp.106-136. <https://doi.org/10.1017/S0269888913000350>

20. Steiniger S., Hunter A.J.S. Free and open source GIS software for building a spatial data infrastructure. / In: Bocher E., Neteler M., (eds.), Geospatial Free and Open Source Software in the 21st Century, LNC, Heidelberg, Springer, 2012. P. 247-261
21. Koshkarev A.V. Geoportal as a tool to control spatial data and services. // Spatial data. 2008. № 2. P. 6-14.
22. Yakubailik O.E. Geoformation geoportal // Computational Technologies. Special Issue 3. 2007. V. 12. P. 116-125.
23. J.D. Blower, A.L. Gemell, G.H. Griffiths, K. Haines, A. Santokhee, X. Yang. A Web Map Service implementation for the visualization of multidimensional gridded environmental data // Environmental Modelling & Software. 2013. V. 47. P. 218-224.
24. Becirspahic and A. Karabegovic. Web portals for visualizing and searching spatial data // Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2015. Opatija, Croatia. P. 305-311.
25. Titov A.G., Gordov E.P., Okladnikov I.G. Development of Web GIS based on services for processing and visualizing spatial data for the analysis and forecasting of regional climate changes // Information and mathematical technologies in science and management. 2016. No. 4-2. pp. 96-109.