

Разработка информационной системы о математических моделях окружающей среды, поддерживающих открытые стандарты моделирования

Александр Б. Столбов ¹

¹Институт динамики систем и теории управления имени В.М. Матросова Сибирского отделения Российской академии наук (ИДСТУ СО РАН), ул. Лермонтова 134, Иркутск, Россия

Аннотация

В статье предлагается информационная система, обеспечивающая формирование и обслуживание расширяемой базы характеристик математических моделей и соответствующего программного обеспечения, применяемого при комплексном исследовании окружающей среды (Integrated Environmental Modeling – IEM). IEM – это научная дисциплина, изучающая взаимосвязь основных характеристик общества и экономики с биосферой и атмосферой в рамках единой модельной структуры. Источником информации могут являться доступные в открытом доступе онлайн репозитории описаний (метаданных) математических моделей. Схема базы данных информационной системы основана на авторской концептуальной модели компонентно-ориентированного сценарного моделирования. Информация, хранящаяся в системе, будет использована для создания сервисов сценарного анализа Байкальской природной территории на основе открытых стандартов моделирования.

Ключевые слова

Комплексное исследование окружающей среды, открытые стандарты моделирования, репозиторий

1. Введение

Исследования, ориентированные на изучение компонентов окружающей среды, являются динамично развивающейся областью научных исследований. Благодаря обширной теоретической и практической работе специалистов в данной области накоплен значительный опыт, разработано множество математических моделей, созданы методические рекомендации, по некоторым природным объектам имеются данные многолетнего мониторинга. Одновременно с этим осуществляется процесс внедрения полученных результатов с целью регламентации деятельности экономики согласно научно-обоснованным выводам. В связи со сложностью изучаемых явлений научное сообщество в последние годы предпринимает целенаправленные усилия по организации междисциплинарных исследований окружающей среды. Одним из результатов этой деятельности является появление относительно новой дисциплины – комплексного моделирования окружающей среды (Integrated Environmental Modeling – IEM) или в другом переводе – интегрированное моделирование окружающей среды.

IEM опирается на объединение различных источников данных, включая геопространственные, метеорологические, гидрологические и социально-экономические. Цель состоит в том, чтобы создать целостное представление о взаимодействии между

разнообразными компонентами окружающей среды. Такой подход позволяет более точно и всесторонне понять сложные взаимосвязи между деятельностью человека, природными процессами и экологическими результатами. ИЕМ может быть применён к широкому спектру экологических вопросов, таких как изменение климата, управление водными ресурсами, загрязнение воздуха, сохранение биоразнообразия и планирование землепользования.

Одной из проблем применения ИЕМ является сложность интеграции разнообразных источников данных и моделей. Для решения этой проблемы с целью создания единой информационно-коммуникационной ИЕМ инфраструктуры предпринимаются коллективные усилия различных научных групп [1-5] для создания информационных баз и репозиториев для моделей и соответствующего программного обеспечения. Данная статья вносит вклад в это направление в контексте разрабатываемой в рамках гранта Министерства науки и высшего образования РФ цифровой платформы исследования Байкальской природной территории.

2. Описание информационной системы

2.1. Базовые возможности и архитектура

Под компонентом моделирования (далее, модель-компонент) понимается программная реализации алгоритма вычислений по математической или имитационной модели, выполненная на языке высокого уровня или в рамках некоторого каркаса моделирования. При этом авторы модели-компонента должны предоставить хотя бы минимальный набор поддержки своего инструментария для использования сторонними пользователями: поставляется хотя бы с минимальным описанием своих возможностей и доступно хотя бы в форме исполняемых файлов.

Задачей информационной система комплексного моделирования окружающей среды для Байкальской природной территории (ИС КМОС БПР) является формирование и обслуживание расширяемой базы характеристик (метаданных), а в перспективе и базы знаний, позволяющей разным группам пользователей (специалистам в области математического моделирования, сценарного анализа, прикладным программистам) получать как можно более полные описания моделей-компонентов с точки зрения интересующих их тем и компетенций. Фрагмент ИС КМОС БПР, описывающий процесс формирования БД моделей-компонентов представлен на Рисунке 1.

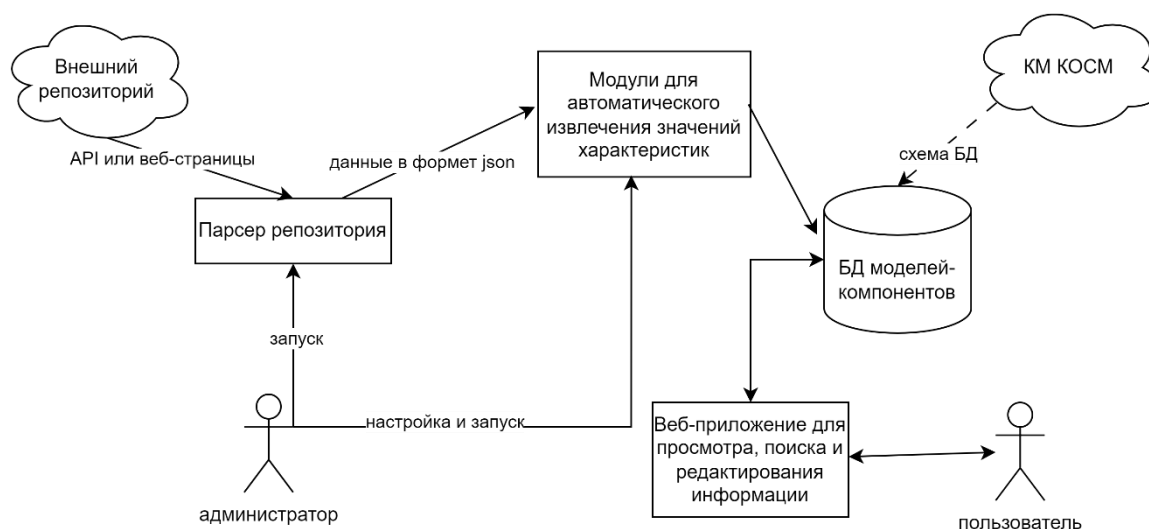


Рисунок 1: Фрагмент архитектуры информационной системы комплексного моделирования окружающей среды для Байкальской природной территории

Предполагается, что существует некий внешний репозиторий, который позволяет получить информацию о модели в структурированном виде. Так как в зависимости от репозитория способ доступа к информации; набор характеристик моделей-компонентов, их идентификаторы и

значения могут отличаться, то для каждого репозитория разрабатывается собственный специализированный парсер. Модули автоматического извлечения характеристик из полученных от репозитория описаний (в формате json) специально отделены от парсеров, чтобы иметь возможность применить разные стратегии выделения значений. База данных моделей компонентов основана на концептуальной модели компонентно-ориентированного сценарного моделирования (КМ КОСМ), кратко описанную в следующем разделе. Пользователь взаимодействует с ИС КМОС БПР через графический интерфейс веб-приложения и может добавлять собственные описания моделей-компонентов, но добавление данных посредством парсеров на текущий момент осуществляет только администратором (разработчиком).

ИС КМОС БПР реализуется с использованием Python, Flask, PostgreSQL, Selenium, WebDriver, React.

2.2. Информационная модель

Информационная модель ИС КМОС БПР основана на понятиях и отношениях концептуальной модели компонентно-ориентированного сценарного моделирования (КМ КОСМ) [6]. На данный момент КМ КОСМ имеет 6 проблемно-ориентированных уровней, в рамках которых рассматриваются различные аспекты процесса создания, развёртывания и использования моделей и соответствующих программных модулей. Понятия и отношения, содержащиеся в КМ КОСМ, описывают информацию: о требованиях к вычислительной архитектуре и поддерживаемым типам данных и файлов; об особенностях используемых каркасов моделирования; об уровне готовности компонента; о математической классификации уравнений и численных методов, включая ограничения на переменные и параметры моделей; типы вариаций вариации экзогенных и эндогенных переменных. В дополнение к шести уровням КМ КОСМ в зависимости от специфики исследуемых комплексных объектов могут добавляться уровни, описывающие предметную область. Для формирования структуры КМ КОСМ были использованы результаты, посвященные специфичным проблемным хранения информации и соответствующие информационные модели [7-10].

Автоматизация извлечения данных возможно для таких характеристик моделей-компонентов, которые имеют значения, принадлежащие некоторому заранее известному множеству: {«да», «нет», «планируется»}, {«1D», «2D», «3D»} и т.п. В некоторых случаях даже, если по смыслу предполагается конечность возможных вариантов, значение характеристики даётся авторами моделей-компонентов в свободной форме. Например, для указания времени выполнения стандартного вычислительного сценария может быть задан как конкретный диапазон значений (от 1 до 6 часов), так и описание на естественном языке («время выполнения может варьироваться от минут до недель, в зависимости от длины симуляции, размера ландшафта и разрешения области»). При автоматическом извлечении ИС КМОС БПР помечает «подозрительные» значения для их последующей ручной обработки.

Существуют характеристики, по своему содержанию, требующие более детального представления на естественном языке. Среди них, отдельно отметим те, характеристики, которые определяют некоторую инструкцию на естественном языке. Одной из задач на будущее, может быть, привлечение современных методом ИИ для явного извлечения описываемых последовательностей действий, ограничений модели, численных методов.

3. Анализ моделей-компонентов на основе открытых репозиториях метаданных

На начальном этапе разработки ИС КМОС БПР для формирования БД были использованы онлайн репозитории CSDMS и zenodo. Кратно были рассмотрены и другие репозитории (Open Science Framework, Figshare), но моделей, посвященных тематике ИЕМ найдено не было. Также не были рассмотрены хранилища кодов типа github, так как извлечение структурированной информации из таких источников представляется более сложной задачей и предполагается как задача для будущих исследований. В качестве иллюстративного примера рассмотрим основные

характеристики компонентов математических моделей окружающей среды, хранимые на текущий момент в ИС КМОС БПР (400 единиц). Для некоторых вариантов характеристик представлены их наиболее популярные или граничные значения, а также статистика. Звёздочками помечены характеристики, автоматизировать извлечение которых сложно (по разным причинам). Минусом помечены характеристики, наличие которых наблюдалось у незначительного количества компонентов моделей (менее 20%).

- Поддерживаемые платформы: Linux, Unix, Windows, Mac OS.
- Язык программирования: Fortran77, Fortran90, IDL, Python, Matlab, C, C++, Java.
- Дата начала разработки: наиболее старая – 1965 год.
- Поддерживаемая размерность пространства: 1D (22%), 1.5D (2%), 2D (45%), 3D (20%).
- Страна разработчик: США (40%), Великобритания (4%), ЕС (10%), Китай (1%), не задано (10%). Здесь важно отметить, что оценка характеристик, связанных со страной происхождения моделей-компонентов и аналогичных, представляется недостоверной из-за особенностей доступных репозиторий, так как многие страны, включая Россию активно не публикуют свои результаты в репозиториях.
- Поддержка стандарта OpenMI (*): да (3%), нет (8%), планируется (15%), теоретически возможно (70%), неизвестно (4%).
- Поддержка стандарта BMI (*): да (20%), нет (4%), планируется (12%), теоретически возможно (60%), неизвестно (4%).
- Поддержка стандарта WMT (*): да (8 %), нет (4%), планируется (14%), теоретически возможно (70%), неизвестно (4%).
- Доступен исходный код: да (80%), нет (10%).
- Ссылки URL на проект (с кодами, описанием, примерами и т.п.).
- Тип лицензии на использование: GPL v3 (20%), BSD/ MIT (20%), GPL v2 (16%), Apache public license (6%).
- Время выполнения стандартного вычислительного сценария (*-): как правило нет точного диапазона, в некоторых случаях указываются «недели».
- Требования к памяти (*): от 10 Mb до 100Gb.
- Формат входных данных: строковый (ASCII), двоичный.
- Формат выходных данных: строковый (ASCII), двоичный.
- Необходима предобработка данных: да (80%), нет (20%).
- Необходима постобработка данных: да (80%), нет (20%).
- Есть средства визуализации: данных: да (40%), нет (60%).
- Есть подробное руководство: да (30%), нет (70%).
- Является компонентом для работы с данными (-): да (2%).
- Заданы данные для тестирования: в 10% моделей.
- Есть веб-сайт: в 40% случаев.
- Тип реализации: модуль (16%), автономное однозадачное (55%), другое (30%).
- Предметная область: моделирование наземных экосистем (60%), геодинамика (55%), климат (8%), гидрология (42%), моделирования морских процессов (16%), моделирование прибрежных зон (22%), криосфера (7%), экосистемы в целом (5%), моделирование форм рельефа (1%), биогенные вещества (0,7%), гуманитарные аспекты (1%). Здесь некоторые модели могут быть отнесены к нескольким предметным областям.

Другие важные характеристик: продолжается ли разработка модели и кода; есть ли поддержка многопроцессорности. Заданные в свободной форме описания: процесса предобработки и постобработки данных; входных и выходных параметров; ключевых физических параметров; уравнений и численных методов; ограничений по времени и пространству, на переменные и параметры и т.п.

4. Заключение

В статье представлены результаты, посвященные первому этапу разработки информационной системы о математических моделях, использующихся при комплексном

исследовании окружающей среды (Integrated Environmental Modeling). Рассмотрен процесс извлечения информации из доступных открытых репозиториях в информационную систему. На основе множества понятий концептуальной модели компонентно-ориентированного сценарного моделирования сформированы характеристики моделей-компонентов и проведён сравнительный анализ. Одним из направлений по совершенствованию ИС КМОС БПР является анализ более сложных хранилищ (типа github) для явного извлечения данных и кодов, а не только описаний.

Информация, хранящаяся в ИС КМОС БПР, позволит при условии дальнейшего совершенствования информационной системы сформировать множество моделей и соответствующего программного обеспечения, пригодного для создания сервисов сценарного анализа Байкальской природной территории на основе открытых стандартов моделирования.

5. Благодарности

Работа выполнена в рамках гранта № 075-15-2024-533 Министерства науки и высшего образования РФ на выполнение крупного научного проекта по приоритетным направлениям научно-технологического развития (проект «Фундаментальные исследования Байкальской природной территории на основе системы взаимосвязанных базовых методов, моделей, нейронных сетей и цифровой платформы экологического мониторинга окружающей среды»).

6. Список литературы

- [1] G.E. Tucker, E.W.H. Hutton, M.D. Piper, B.Campforts, T. Gan, K.R. Barnhart, A.J. Kettner, I. Overeem, S.D. Peckham, L. McCready, J. Syvitski, CSDMS: a community platform for numerical modeling of Earth surface processes, 2022. doi:10.5194/gmd-15-1413-2022.
- [2] Q. Harpham, A. Hughes, R.V.Moore, Introductory Overview: The OpenMI 2.0 Standard for Integrating Numerical Models, *Environmental Modelling & Software* 122 (2019) 1–13. doi:10.1016/j.envsoft.2019.104549.
- [3] C.A. Buahin, J.S. Horsburgh, Advancing the Open Modeling Interface (OpenMI) for integrated water resources modeling, *Environmental Modelling & Software* 108 (2018) 133–153. doi:10.1016/j.envsoft.2017.02.028.
- [4] M.M. Morsy, J.L. Goodall, A.M. Castronova, P. Dash, V. Merwade, J.M. Sadler, M.A. Rajib, J.S. Horsburgh, D.G. Tarboton, Design of a metadata framework for environmental models with an example hydrologic application in HydroShare, *Environmental Modelling & Software* 93 (2017) 13–28. doi:10.1016/j.envsoft.2017.02.028.
- [5] A.G. Lodhi, A.N. Godrej, D. Sen, A.A. Barana, URUNME: A generic software for integrated environmental modeling, *Environmental Modelling & Software* 134 (2020). doi:10.1016/j.envsoft.2020.104737.
- [6] А.А. Лемперт, А.Б. Столбов, В.В. Козлов, Использование онтологического подхода для интеллектуальной поддержки сценарного моделирования водных объектов Приангарья, *Информационные и математические технологии в науке и управлении* 19 (2020) 14–24. doi: 10.38028/ESI.2020.19.3.002.
- [7] M. Elag, J.L. Goodall, An ontology for component-based models of water resource systems, *Water Resour. Res.* 49 (2013) 5077–5091. doi: 10.1002/wrcr.20401.
- [8] R. Raskin, M. Pan, Knowledge representation in the Semantic Web for Earth and Environmental Terminology (SWEET), *Computers & Geosciences* 31 (2005) 1119–1125. doi: 10.1016/j.cageo.2004.12.004.
- [9] R. Argent, An overview of model integration for environmental applications - Components, frameworks and semantics, *Environmental Modelling & Software* 19 (2004) 219–234. doi: 10.1016/S1364-8152(03)00150-6.
- [10] J. Horsburgh, D. Tarboton, D. Maidment, I. Zaslavsky, A relational model for environmental and water resources data, *Water Resources Research* 44 (2008) 1–12. doi: 10.1029/2007WR006392.