

СРЕДА ПОЛИВАРИАНТНОГО АНАЛИЗА ДИНАМИЧЕСКИХ МОДЕЛЕЙ АГРОЭКОСИСТЕМ

Топаж А.Г., Полуэктов Р.А., Медведев С.А.

Агрофизический НИИ Россельхозакадемии

E-mail: topaj@hotmail.ru

Аннотация: Традиционные прецеденты использования компьютерных динамических моделей продукционного процесса растений, такие как идентификация параметров, прогноз урожая и анализ агротехнологий порождают необходимость множественных расчетов одной и той же модели с разными наборами входных данных. С целью автоматизации соответствующих операций авторами разработана универсальная среда поливариантного анализа динамических моделей агроэкосистем. Она позволяет зарегистрировать в системе произвольную динамическую модель и обеспечить планирование и проведение с ней компьютерного многофакторного эксперимента. Приводятся примеры использования созданной системы для решения практических и исследовательских задач в агроэкологии.

Ключевые слова: Среда моделирования, информационная система, база данных, модель продукционного процесса

Введение

Компьютерные динамические модели продукционного процесса растений приобретают все большую роль в качестве интеллектуального ядра систем поддержки технологических решений в агроэкологии. С математической точки зрения, любая динамическая модель представляет собой сложный расчетный алгоритм, позволяющий получать интересующий пользователя вектор результатов (прогноз урожайности, даты наступления фенологических стадий развития) для заданного набора входных данных (погода и агротехника в течение сезона вегетации, параметры культуры и почвы и т.п.) Подобная разовая процедура прогона модели может быть названа единичным или одновариантным расчетом. Однако, типичный прецедент использования модели в любой системе поддержки решений чаще всего предполагает не единичный, а множественный расчет одной и той же модели с разными наборами входных параметров с целью анализа и сравнения получаемых результатов. Можно перечислить несколько проблем или практических задач, приводящих к необходимости многократного расчета с указанием источника вариантности входных данных для каждой из них (таблица 1).

Таблица 1. Задачи, требующие поливариантного анализа динамической модели агроэкосистемы.

№	Проблема	Источник вариантности данных
1	Параметрическая идентификация модели	Перебираемые значения параметров
2	Оперативный прогноз урожайности в течение сезона вегетации	Сценарии погоды
3	Поиск оптимальных агротехнологических решений	Разные варианты технологий (сроки и нормы технологических воздействий)

4	Исследование влияния климатических изменений на агроэкосистему	Синтетические сценарии погоды для будущего климата
5	Точное земледелие	Пространственная неоднородность посева и почвенного покрова.

С учетом сказанного, крайне перспективной выглядит идея создания некой специализированной среды или инфраструктуры, которая позволяла бы производить автоматизированную подготовку данных, многократный расчет модели и анализ получаемых результатов в пакетном режиме. Более того, принимая во внимание универсальный характер соответствующих операций, желательно, чтобы эта среда позволяла выполнять их не с некой единственной, а с произвольной динамической моделью. В сообщении представляется прототип подобной системы автоматизации компьютерного эксперимента, инкапсулирующей в себе две базовые функциональные составляющие — регистратор описаний внешних моделей продукционного процесса и универсальную инфраструктуру их поливариантного анализа.

Принципы расширения мета-информации в реляционной базе данных.

Требования, заявленные к системе, обуславливают необходимость поддержки ее широкой пользовательской настраиваемости. В частности, должна поддерживаться возможность регистрации в ней произвольных внешних моделей с описанием форматов их входных и выходных данных. В то же время сами эти данные тоже должны быть информационным наполнением системы. Таким образом, информационная модель приложения разбивается на два уровня разной степени абстракции. Нижний уровень — это «модель-ориентированные» данные, то есть данные, относящиеся к той или иной конкретной зарегистрированной в системе модели (наборы входных параметров и результатов расчетов). При этом способ и формат их представления должны быть такими, чтобы соответствующая модель могла естественным образом воспринимать их в процессе своего функционирования под управлением универсальной среды. Верхний уровень — это данные о каждой модели как таковой, включая описание форматов используемых ею данных на некоем общем для всех моделей языке. Подобный вид «информации об информации» в информатике называется метаданными. Задача соединения и непротиворечивого взаимодействия данных и метаданных в рамках единой информационной системы имеет долгую историю и много альтернативных подходов для возможного разрешения. По счастью, в рамках проектируемой системы нет необходимости решать проблему в полном масштабе, то есть отображать в схеме базы данных такие абстрактные понятия объектно-ориентированного подхода как «сущность» или «ассоциация». Ограниченная предметная область рассмотрения позволяет оперировать в информационной модели нашего приложения более конкретными терминами. Действительно, любая динамическая модель агроэкосистемы задействует данные, которые можно разбить на ограниченное число информационных доменов. Эти домены (называемые в нашей системе «факторами») формируют основу информационной модели среды поливариантного анализа.

Мы рассматриваем следующие факторы: «местность», «погода», «сорт/культура», «почва», «начальное состояние» и «технология» для входных данных и единственный информационный домен «результаты» как выход модельного расчета. Основное

предположение состоит в том, что единичный вариант расчета модели можно представить в виде кортежа ссылок на конкретные значения каждого из predetermined факторов, список которых един и неизменен для любой используемой модели. Таким образом, любые данные для расчета произвольной модели представляются в виде набора данных (DataSet) со специфичным и настраиваемым для каждой модели списком полей (колонок). Этот набор данных однозначно связывается по вторичному ключу с одним из «уровней фактора» - записью в корневой таблице, соответствующей одному из predetermined информационных доменов. При этом каждый элементарный набор данных может содержать единственную запись (например, общие параметры конкретной культуры) или множество записей (например, свойства почвы по слоям или параметры культуры, привязанные к фазам развития). Таким образом, общая проблема отображения («мэппирования») метаинформации в реляционной базе данных сводится для нашего случая к частной задаче представления в базе произвольных табличных данных с настраиваемым набором колонок.

Для решения данной проблемы в нашей системе мы воспользовались возможностями, прописанными в стандарте SQL92. В частности, в нем специфицированы средства доступа и управления информацией о структуре рассматриваемой базы данных, путем запросов не к конкретным таблицам, а к общей метамодельной информации, хранящейся в т.н. INFORMATION_SCHEMA. Оригинальный подход, примененный в нашей среде поливариантного анализа, состоит в комбинировании в рамках единой информационной модели пользовательских таблиц (список которых произвольным образом настраиваем и расширяем) и predetermined запросов к INFORMATION_SCHEMA. Иными словами, в базе данных системы определены и структурно неизменяемы несколько базовых таблиц и запросов, составляющих ее информационное ядро. А таблицы, хранящие данные, необходимые для расчета каждой новой регистрируемой в среде динамической модели, могут появляться и исчезать в базе данных системы по мере ее эволюции. Изоляция данных, составляющих информационное ядро приложения, от структур, создаваемых конечным пользователем в интерфейсе регистрации и/или редактирования описаний моделей, поддерживается с помощью нескольких специальных соглашений и ограничений.

Подобный подход содержит следующие преимущества по сравнению с известными авторам альтернативными методами расширения метаинформации в реляционных СУБД:

- При выборке информации, необходимой для формирования сценария расчета конкретной модели, система работает с реальными физическими таблицами, используя стандартный оптимизированный механизм SQL-запросов, что не приводит к потерям эффективности по объему памяти и быстродействию.
- Все необходимые взаимосвязи и ограничения (вторичные ключи и т.д.) в модель-ориентированных таблицах создаются автоматически
- Возможности работы с расширенными метаданными не лимитируются используемыми инструментами, специфичными для конкретной базы данных
- Система переносима на любые базы данных, поддерживающие спецификацию INFORMATION_SCHEMA.

Программный интерфейс взаимодействия среды поливариантного расчета и внешней динамической модели агроэкосистемы

Система поливариантного расчёта предоставляет общие алгоритмы обработки данных в произвольных форматах. В этом отношении она близка к классическому шлюзу обмена данными, поэтому её архитектура сходна с типичной архитектурой шлюза с теми же самыми фундаментальными понятиями: схема, адаптер и сценарий. В нашей системе эти понятия имеют следующую трактовку.

Схема – это структура таблиц, созданная пользователем для хранения исходных данных и результатов работы для каждой конкретной зарегистрированной в системе модели производственного процесса.

Адаптер – это стандартный программный интерфейс, объявленный в системе поливариантного расчёта, через который происходит обмен данными. Адаптер решает три задачи:

1. преобразование исходных данных, хранящихся в базе данных системы поливариантного расчёта, в формат данных, требуемый конкретной моделью производственного процесса;
2. запуск модели на выполнение со сформированным набором данных;
3. «обратное» преобразование результатов работы модели из того формата, в котором их генерирует модель, в промежуточный формат данных, предназначенный для сохранения в базе данных системы поливариантного расчёта.

Сценарий – это набор уровней факторов, соответствующий единичному варианту прогона модели. Во время выполнения пакетного расчёта все данные, связанные с уровнями факторов сценария, извлекаются из базы данных, и передаются адаптеру модели в промежуточном формате. После расчёта модели для каждого сценария все результаты, полученные от адаптера в промежуточном формате, сохраняются в базу данных и связываются со сценарием.

Поскольку мы использовали .NET Framework в качестве основной платформы для разработки системы, промежуточный формат данных основан на стандартном компоненте ADO.NET DataSet. Адаптер представляет собой интерфейс, объявленный в специальной разделяемой сборке. В текущей версии системы поливариантного расчёта основной метод адаптера принимает два параметра: словарь, ключом которого является специальное перечисление FactorType, а значением – компонент ADO.NET DataSet, заполненный исходными данными для соответствующего фактора, и компонент BackgroundWorker, используемый для оповещения о ходе выполнения операции. Этот метод должен вернуть экземпляр ADO.Net DataTable, заполненный результатами расчёта модели. Структура всех таблиц, представленных компонентами ADO.NET, совпадает с той, которая была введена пользователем при регистрации модели. Второй метод интерфейса адаптера предназначен для импорта данных из существующего хранилища исходных данных модели. Он не возвращает значения и имеет два параметра: словарь из перечисления FactorType и наборов данных, которые требуется заполнить, и компонент BackgroundWorker.

В системе реализована плагиновая архитектура для работы с адаптерами. Все сборки, находящиеся в директории приложения, автоматически загружаются при старте и анализируются на предмет реализации интерфейса адаптера. После этого все найденные

типы доступны через меню на панели адаптеров. Кроме того, система содержит несколько готовых компонентов, позволяющих строить адаптер непосредственно через её пользовательский интерфейс, без написания какого-либо кода. Реализация адаптера по умолчанию содержит три вспомогательных интерфейса для реализации трёх фаз работы адаптера: загрузчик исходных данных, движок для запуска модели и загрузчик результатов. В настоящее время поддерживается формат Microsoft Excel для исходных данных, csv для результатов и вызов неуправляемой экспортируемой функции для запуска модели. В стадии разработки находятся универсальные адаптеры для моделей семейства DSSAT и моделей, совместимых с альтернативной универсальной оболочкой моделирования GUICS [2].

Автоматизация компьютерного эксперимента с моделями производственного процесса.

Приведем краткое описание общей (инвариантной по отношению подключаемым моделям) части информационной модели разработанной системы, обеспечивающей ее базовую функциональность — планирование и осуществление поливариантного (многофакторного) компьютерного эксперимента. Принципиальными концептами здесь выступают понятия «Проекта» и уже упомянутого выше «Сценария». Проект — это общее описание конкретного эксперимента, который мы хотим осуществить. В существующей версии присутствует ограничение, что в рамках проекта могут быть выполнены расчеты только по одной выбранной модели (в следующих версиях системы, направленных на создание адаптивных систем моделирования указанное ограничение будет снято). Проект состоит из нескольких сценариев, каждый из которых соответствует одному из будущих прогонов модели и представляется в виде набора контрольных директив, определяющих какие уровни факторов (то есть какие почва, культура, погода и т.д.) будут использованы в качестве входных данных для соответствующего варианта расчета.

Таким образом, планирование компьютерного эксперимента в рамках описываемой системы состоит из двух этапов — формирования всех интересующих нас в настоящий момент уровней факторов и объединения их в структуру «проект-сценарии». Опишем оба этих этапа более детально.

В системе предусмотрено несколько механизмов получения данных, формирующих «уровень фактора» для их сериализации и последующего использования в проектах. Прежде всего, пользователь может получить информацию из «родного» формата входных данных той или иной модели, используя «обратную» функцию преобразования в соответствующем адаптере. Обычно это будут «реальные данные», то есть действительно полученные результаты наблюдений в полевом опыте, параметры конкретных почв и культур и т.п. Кроме того, в систему встроены различные алгоритмы искусственного варьирования произвольных параметров или характеристик, служащие для получения «виртуальных» или «тестовых» уровней того или иного фактора. Простейшим таким алгоритмом является перебор и «размножение» величины того или иного параметра в заданном интервале значений с выбранным постоянным инкрементом. Как результат мы получаем набор уровней соответствующего фактора (например, набор искусственных «почв»), которые отличаются друг от друга только величиной варьируемого параметра (например, только значением влажностью завядания в слое 20-30 см). Дополнительно в системе реализован генератор погодных реализаций, базирующийся на алгоритме, предложенном Ричардсоном и Райтом

[2]. Он служит для стохастического моделирования методом Монте-Карло «синтетических» сценариев погоды, соответствующих климатическим условиям выбранной местности. Наконец, данные для вновь вводимого уровня фактора могут быть вручную введены пользователем в графическом интерфейсе или импортированы из текстового файла.

Завершающая стадия планирования компьютерного эксперимента состоит в формировании вариантов прогона модели в терминах проекта (набора сценариев). Эта процедура достаточно проста. Требуемый набор сценариев вычисляется как декартово перемножение всех выбранных для анализа уровней для каждого из факторов. Принципы формирования проекта в пользовательском интерфейсе системы поливариантного расчета демонстрируются на рис. 1.

Проект, созданный как набор сценариев, может быть «просчитан» в любой момент времени. Расчет проекта, то есть непосредственное проведение компьютерного эксперимента, сводится к последовательному запуску (путем вызова главной функции адаптера) соответствующей модели с набором входных данных, отвечающим текущему сценарию и записи полученных результатов в постоянное хранилище. После полного расчета проекта весь набор получившихся результатов может быть проанализирован с помощью встроенных в систему средств статистической обработки данных (в системе предусмотрены механизмы табличной и графической визуализации, а также специализированные плагины обработки данных для одно-, двух- и многофакторных экспериментов с поддержкой повторностей).

Примеры использования

Разработанная среда поливариантного расчета использовалась нами для решения целого ряда задач прикладного и теоретического характера. В частности, с ее помощью была исследована проблема оценки влияния глобальных климатических изменений на стабильность и продуктивность агроэкосистем Северо-Запада России. Основная идея заключалась в использовании стандартного генератора погоды для получения сценариев «погоды будущего» и использования их в качестве входной метеорологической информации для модели продукционного процесса AGROTOOL. Действительно алгоритм стохастического генератора погоды как многомерного случайного процесса использует в качестве исходных параметров некоторые климатические характеристики выбранной местности. Они, как правило, идентифицируются по данным фактических наблюдений за реальной погодой в течение нескольких лет. Изменив эти параметры в соответствии с теми или иными прогнозами климата будущего, и запустив процедуру стохастического моделирования, мы можем получить набор «синтетические» погод, характерных для этого измененного состояния климата. Таким образом, задача оценки влияния климатических изменений на продукционный процесс сельскохозяйственных культур сводится в терминах разработанной среды моделирования к планированию и осуществлению однофакторного компьютерного эксперимента с повторностями. При этом генератор погоды позволяет нам сгенерировать любое количество сценариев для последующего анализа результатов, то есть получить репрезентативную выборку для статистически обоснованных заключений.

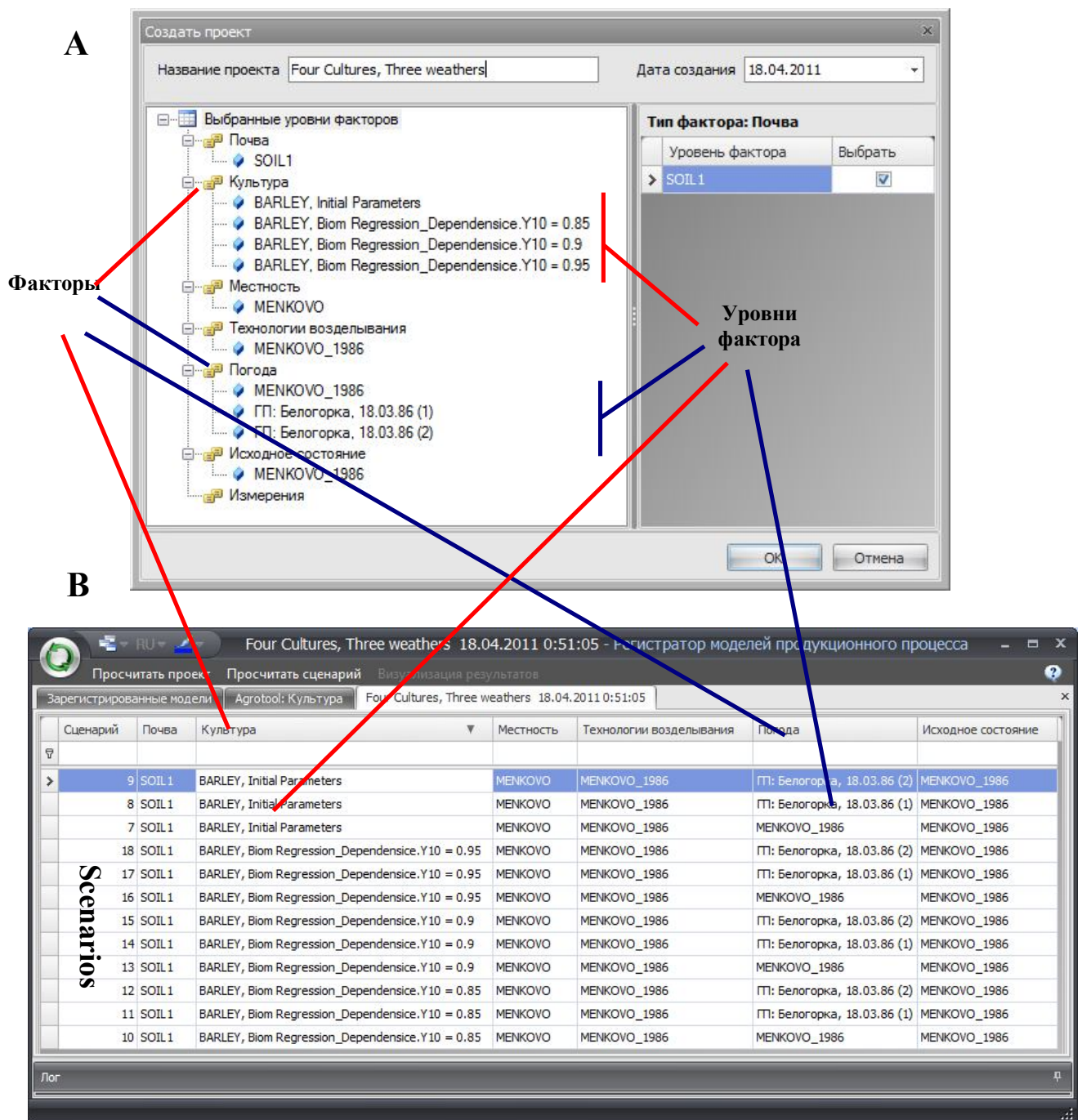


Рис. 1. Планирование многофакторного эксперимента в среде поливариантного расчета (А – диалог создания проекта, В – визуализатор структуры проекта).

Другая проблема, решенная нами с помощью описанной системы, заключалась в поиске оптимальной стратегии орошения для засушливых регионов Среднего Поволжья. Задачей было найти наиболее экономически эффективные параметры системы автоматического полива (критическая величина почвенного влагозапаса, достижение которой в процессе иссушения должно «включать» оросительный механизм и норма стандартного полива). В качестве критерия оптимизации выступала оценка интегральной экономической выгоды хозяйствования (стоимость урожая минус суммарные затраты на полив). Ясно, что пожалуй единственным средством оценки потенциального урожая в

зависимости от конкретной стратегии орошения в конкретных условиях среды может выступать динамическая компьютерная модель. Таким образом, задача сводилась к постановке двухфакторного компьютерного эксперимента, где в качестве первого фактора выступало параметризованное описание стратегии автоматического полива, а в качестве второго (интерпретируемого для целей статистического анализа как повторность) – погода. Обработка результатов проведенного эксперимента позволила выработать обоснованные рекомендации по выбору эффективной политики сельскохозяйственного водопользования в региональном масштабе.

Перспективы

Основной вектор дальнейшего развития системы нам видится в применении в ней средств и методов параллельных и облачных вычислений. Действительно, расчет проекта в сегодняшней версии – это независимый, но последовательный прогон используемой модели с различными наборами входных параметров. Естественным представляется использовать как аппаратные возможности современной техники, так и новейшие подходы в области информационных технологий для того, чтобы организовать параллельный расчет сценариев в рамках одного проекта. Масштаб решения при этом может быть совершенно различным – от распараллеливания обработки сценариев по процессорам в пределах одного компьютера, до развертывания полноценного кластера распределенных грид-вычислений в глобальной сети. Более того, представляется, что внесение параллелизма в среду поливариантного расчета – это не только вопрос повышения эффективности. Можно показать, что параллельный, но синхронизированный расчет нескольких экземпляров стандартной одномерной модели продукционного процесса посева может быть использован для эмуляции простейших процессов перетока вещества и энергии в пространственно-неоднородных экосистемах. В результате, вовлечение средств и методов параллельных и распределенных вычислений в систему поливариантного анализа может послужить первым шагом к созданию на ее базе имитационно-моделирующего комплекса для описания агроландшафта. Последнее особенно актуально в свете задач и вызовов точного земледелия, позиционируемого в качестве мейнстрима прикладных исследований в современной сельскохозяйственной науке [3].

Литература

1. Acock, B., Ya. A. Pachepsky, E.V. Mironenko, F. D. Whisler, and V. R. Reddy. 1999. GUICS: A Generic User Interface for On-Farm Crop Simulations. *Agronomy J.*, 91:657-665.
2. Richardson, C. W., Wright, D. A.: WGEN: A Model for Generating Daily Weather Variables. In: U. S. Department of Agriculture, Agricultural Research Service, ARS-8, 83 p. (1984)
3. Полуэктов Р.А., Топаж А.Г., Кобылянский С.Г. Система поливариантных расчетов динамической модели продукционного процесса в задачах точного земледелия // Доклады РАСХН, 2005, №6, стр. 58-62.