

## ИСПОЛЬЗОВАНИЕ ДАННЫХ ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ В ИМИТАЦИОННЫХ МОДЕЛЯХ АГРОЭКОСИСТЕМ – ПРОБЛЕМЫ И ПЕРСПЕКТИВЫ

Д.Ю. Ерёмченко, А.Г. Топаж, С.А. Медведев (Санкт-Петербург)

### Введение

Одной из насущных проблем, решаемых в рамках направления цифровизации и автоматизации растениеводства, является апостериорная оценка и упреждающий прогноз урожайности сельскохозяйственной культуры в конкретный сезон и в конкретном месторасположении. Своевременное получение подобной объективной и достоверной информации имеет принципиальное значение в задачах экономического планирования, оперативного мониторинга и практике агрострахования. Одним из инструментов получения необходимых данных выступают математические модели продукционного процесса сельскохозяйственных растений. При этом в ходе развития математического моделирования в агроэкологии выделилось два качественно различающихся подхода к созданию подобных моделей: эмпирический и теоретический (Полуэктов и др., 1998).

*Эмпирические (статистические) модели или модели, основанные на данных,* характеризуются тем, что в них широко используется эвристическое описание определяющих процессов, применяются различные регрессионные соотношения. Подобное формальное описание на логическом уровне может хорошо отражать свойства реальной системы в терминах «вход-выход», но никак не привязываться к сути физических, химических и биологических процессов протекающих в реальных системах. При эмпирическом подходе моделируемый объект рассматривается как своеобразный серый ящик. Большинство определяющих соотношений в этом ящике являются эвристическими функциями, немногочисленные параметры которых идентифицируются по фактическим измерениям и наблюдениям.

*Теоретические (механистические) модели или модели, основанные на знаниях,* – это модели, предполагающие рассмотрение сути протекающих в реальных системах процессов и описание их динамики на языке физически интерпретируемых уравнений (обычно дифференциальных уравнений). Отражение в модели комплекса физических и биологических явлений, имеющих место в среде обитания и в самом рассматриваемом растении, может служить определённой гарантией ее работоспособности в широком классе внешних воздействий – более широком, чем тот, на котором она параметризировалась и верифицировалась в процессе создания. Иначе говоря, в то время как статистические модели позволяют только интерполировать результаты, теоретические модели позволяют их экстраполировать. Другим большим преимуществом имитационных моделей, основанных на знаниях, является возможность извлечения из них различных побочных данных и результатов на всех этапах моделирования, а не только конечных целевых показателей. Реализацией механистической методологии моделирования в агроэкологии выступают динамические имитационные модели агроэкосистем или так называемые модели типа «Погода-Урожай».

История имитационного моделирования продукционного процесса растений имеет богатую историю, как в России, так и за рубежом. К настоящему времени разработано большое число получивших широкую известность семейств моделей, таких как DSSAT (США), WOFOST (Нидерланды), MONICA (Германия), APSIM (Австралия), STICS (Франция), AGROTOOL (Россия). Но если изучить вопрос о том, в каких сферах находят основное применение данные модели, то окажется, что это в

основном теоретические научные исследования, характеризующиеся крупным временным или пространственным масштабом рассмотрения – оценка влияния на агроэкосистемы изменений климата, формирование требований к новым гибридам и сортам для их интродукции в новые зоны возделывания (так называемые идеотипы) и т.п. В то же время известно критически мало примеров, когда динамическая модель продуктивности с успехом применялась в той исходной задаче, для которой собственно и была разработана – предсказать или оценить урожай в конкретном сезоне и на конкретном поле. Чаще всего, точность соответствующих локальных оценок, то есть соответствие модельных урожаев фактическим значениям оказывается абсолютно неудовлетворительной. Типичный пример подобного расхождения приведён на рис. 1.

Можно выделить несколько причин, объясняющих подобные негативные результаты. Прежде всего, большая проблема связана с качеством используемых данных. Причём, здесь под данными сомнительной точности понимается как информация, используемая на входе модели (погода, почва, сортовые особенности культуры, технология, начальные условия и т.д.), так и сведения о фактической урожайности, с которыми сравнивается результат моделирования. Вторая существенная проблема связана с неточностью или неадекватностью самой модели, которая, в свою очередь, обуславливается двумя обстоятельствами – внутренними логическими и алгоритмическими ошибками (динамическая модель агроэкосистемы – это сложный программный продукт, написанный, как правило, непрофессионалами в области информационных технологий) и учётом в модели только самых основных факторов, лимитирующих продуктивность (водный стресс, уровень минерального питания, температура и т.д.). А дьявол, как известно, прячется в деталях, и урожайность на соседних полях может отличаться в разы вследствие множества различных обстоятельств, которые в прикладной модели все учесть невозможно (болезни, сорняки, вредители, стресс по микроэлементам и т.д.).

Рис. 1. Пример соответствия рассчитанных по модели WOFOST и фактических урожаев риса для случайной тестовой выборки фермерских хозяйств в Республике Индия

Однако в последние годы появился совершенно новый источник данных о текущем состоянии агроландшафтов. С внедрением и развитием совершенно новых, дистанционных и даже космических средств мониторинга оказалось, что многие

характеристики состояния системы «почва – растение – атмосфера» сегодня допускают прямое оперативное измерение с практически произвольно задаваемым уровнем временного и пространственного разрешения. Причем, для этого зачастую не требуется никаких специальных усилий или дорогостоящих измерений, достаточно иметь доступ к частично или полностью открытым, оперативно обновляющимся базам данных дистанционного спутникового зондирования. Появление больших объёмов информации о динамике оптических свойств посевов, получаемых по обработкам спутниковых снимков, дало новую жизнь направлению, связанному с построением статистических или эмпирических моделей. Только вместо простых регрессионных соотношений для оценки урожайности теперь используются модные современные подходы, связанные с методами обработки больших данных (Big Data), искусственного интеллекта и машинного обучения. При этом открытым остаётся вопрос о роли и месте «наукоёмких» механистических моделей в этом процессе. Иными словами, ставится задача интеграции моделей, основанных на знаниях с моделями, основанными на данных, и с новыми источниками данных как таковыми, то есть использования нового типа информации (дистанционного зондирования) в динамических моделях агроэкосистем.

### Материалы и методы

В последние годы вопрос использования данных прямых измерений или, говоря шире, данных из произвольных альтернативных источников в динамические модели производственного процесса привлекает большое внимание исследователей и практиков-модельеров. Как правило, идея состоит в том, чтобы интегрировать (связать) как в пространстве, так и во времени текущие переменные состояния посева с различной информацией, полученной с использованием методов и средств оперативного удаленного мониторинга. Для того чтобы продемонстрировать все возможности и аспекты этой операции, рассмотрим формальное представление динамической модели производственного процесса как многомерной дискретной системы автоматического управления. Внутренняя логика любой подобной динамической модели представляет собой рекуррентный алгоритм:

$$\mathbf{x}_{k+1} = f(\mathbf{x}_k, \mathbf{u}_k, \mathbf{w}_k, \mathbf{a}), k=0, 1 \dots T \quad (1)$$

где:  $\mathbf{x}$  – вектор динамических переменных состояния исследуемого объекта;

$\mathbf{a}$  – вектор постоянных параметров (например, сортовые физиологические параметры культуры);

$\mathbf{u}$  – вектор контролируемых внешних воздействий (применяемая агротехника);

$\mathbf{w}$  – вектор неконтролируемых внешних воздействий (погода);

$k$  – временной шаг модели (принимаемый в большинстве динамических моделей равным одним суткам),

$f$  – эволюционный оператор, составляющий логическую суть имитационного алгоритма (Полуэктов, 1991).

В ходе решения задачи ассимиляции необходимо чётко различать друг от друга данные наблюдений или измерений  $\mathbf{z}$  (поступающие из внешних по отношению к модели информационных ресурсов), характеристики состояния модели (её внутренние динамические переменные  $\mathbf{x}$ ), модельные параметры  $\mathbf{a}$ , а также выходные характеристики модели (в качестве основной подобной характеристики в большинстве задач выступает урожай). Основные принципы и подходы к ассимиляции внешних данных в модели продуктивности описаны в различных обзорных статьях (Delécolle et al., 1992; Jin et al., 2018) и могут быть разбиты на четыре большие группы: метод рекалибровки (*Recalibration*), метод прямого присвоения (*Forcing*), метод адаптивного обновления (*Updating*) и метод апостериорной коррекции (*Corrective approach*).

Метод прямого присвоения (*Forcing*) заключается в том, что на каждом временном шаге модели часть вектора состояния  $\mathbf{x}_k$  искусственно заменяется на соответствующие значения, напрямую вычисленные из данных дистанционного зондирования  $\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k(\mathbf{z}_k)$ . Например, модельная величина листового индекса LAI заменяется на ту, которая может быть оценена из текущего вегетационного индекса NDVI. Этот подход вызывает много справедливой критики. Во-первых, при этом происходит своеобразная подмена причин и следствий, вызванная неполной наблюдаемостью исходной модели. Действительно, обратная зависимость переменных состояния от измеренного оптического индекса может быть неоднозначной или слабо выраженной в том или ином интервале значений (тот же NDVI может зависеть одновременно от нескольких характеристик текущего состояния, а, кроме того, практически не меняется при достижении LAI значений, больших 2). Во-вторых, частота поступления данных ДЗЗ, как правило, гораздо меньше, чем шаг модели, то есть в промежутках между датами измерений величины используемых оптических индексов приходится интерполировать. И, наконец, при этом нарушается внутренняя логика модели, так как некоторые, рассчитанные на каждом шаге переменные состояния директивно заменяются на следующем шаге на новые значения, никак не коррелирующие с другими переменными состояния.

Адаптивное обновление (*Updating*) заключается в последовательном переприсвоении части вектора состояния модели не на каждом шаге её расчёта, а только в те даты, когда доступными делаются новые наблюдения. В основе этого подхода лежит предположение о том, что скорректированные подобным образом результаты моделирования в текущий день улучшат точность моделирования в последующие дни. Часто алгоритмы подобного рода называются методами усвоения или ассимиляции данных фактических измерений. Этот подход более гибкий и вычислительно эффективный, нежели метод прямого присвоения, кроме того он является менее разрушительным с точки зрения поддержания логической связности модели. Но к нему можно предъявить те же претензии, связанные с необходимостью обратного пересчёта переменных состояния из косвенных и опосредованных данных измерений.

Метод перекалибровки (*Recalibration*) принципиально отличается от двух описанных выше подходов тем, что мы не используем обратный пересчёт характеристик состояния из косвенных измерений и не нарушаем логическую стройность используемой модели.

Сущность перекалибровки состоит в том, чтобы, используя алгоритмы алгебраической оптимизации, подобрать те или иные параметры модели так, чтобы обеспечить наилучшую согласованность между оптическими индексами, рассчитанными из текущих переменных состояния и полученных обработкой снимков ДЗЗ. То есть подгон модели под измерения осуществляется путём идентификации её параметров в ходе процедуры направленного поиска решения оптимизационной задачи

$$\hat{\mathbf{a}} = \arg \min_{\mathbf{a}} \sum_k \|\mathbf{z}_k - \hat{\mathbf{z}}(\mathbf{x}_k)\| \quad (2)$$

При этом в качестве варьируемых параметров модели могут выступать не только статические параметры  $\mathbf{a}$ , но и не полностью известные нам характеристики применяемой технологии  $\mathbf{u}$ , или параметры начального состояния  $\mathbf{x}_0$ . Так, одной из модификаций описанного подхода может служить метод усвоения данных методом минимизации корректирующих возмущений (Топаж, Митрофанов, 2017). Безусловно, данный метод обладает самой большой вычислительной сложностью, однако при этом нам не требуется переписывать логику модели, то есть мы можем использовать любую модель внешней разработки как «чёрный ящик».

Определение и перспективы использования методов апостериорной коррекции приводятся в разделе «Заключение» настоящей статьи.

Описываемое в докладе исследование выполнялось в рамках работ по договору с компанией ООО «КонтролТуту.Ру» по теме «Модификация сервисов вычислительного контура для модельной оценки и прогнозирования урожайности сельскохозяйственных культур на базе массовых расчётов прикладных имитационных моделей агроэкосистем и их развёртывание в составе масштабируемой вычислительной архитектуры информационной системы CLASS». Эта система разрабатывается компанией-заказчиком для внедрения на государственном уровне для целей централизации и автоматизации системы агрострахования в Республике Индия. В рамках этого договора были выполнены тестовые расчёты для адаптации методов адаптивной идентификации к массовым расчётам продуктивности по нескольким динамическим моделям. В качестве расчётных имитационных моделей были выбраны семейства моделей ORYZA, WOFOST, MONICA и AGROTOOL, а моделируемых культур – пшеница и рис в зимнем сезоне Раби 2021 года. В качестве наблюдаемой характеристики, по которой рассчитывалось расхождение между модельным и реальными данными использовался вегетационный индекс NDVI. Это обуславливается как наличием достаточно информативных временных рядов этого показателя, которые можно легко получить обработкой спутниковых снимков, так и соображениями о том, что этот индекс наиболее явно отражает плотность растительного покрова в рассматриваемой области, а следовательно, можно ожидать, что его динамика в течение периода вегетации может быть наиболее тесно связано с получающимся в результате урожаем.

Для получения характеристик спектра отражения модельного посева по текущей величине модельного LAI нами использовалась широко известная и распространённая модель оптических свойств посева PROSAIL. Дополнительным соображением в пользу применения данной модели может служить то, что её исходные коды на языке Python находятся в открытом доступе и могут быть без ограничений использованы во внешних приложениях. В основе модели PROSAIL лежат подходы, изложенные в работах (Jacquemoud et al., 2009; Ding et al., 2017; Berger et al., 2018). Принципиальная схема проведённых расчётов для интеграции данных дистанционного зондирования в имитационные модели продуктивности методом перекалибровки приведена на рис. 2.

Рис. 2. Алгоритм усвоения данных ДЗЗ в расчёты имитационных моделей агроэкосистемы методом перекалибровки.

В качестве первого скалярного варьируемого параметра модели выступала дата сева. Поскольку в предоставленных исходных данных соответствующие сведения отсутствовали, то подбором оптимального значения этой технологической переменной достигалось наилучшее соответствие фазы пика временного хода NDVI на оси физического времени. Вторым варьируемым параметром модели был выбран вектор фиктивных или виртуальных азотных подкормок (даты и нормы внесения азотсодержащих удобрений), задаваемых как входное технологическое воздействие в исследуемых динамических моделях продукционного процесса. Подбором расписания азотных подкормок обеспечивалась «тонкая подстройка» формы временного хода модельного NDVI в каждой конкретной дате измерений.

### Результаты

Выборочные результаты использования описанной методики для улучшения модельной оценки урожайности сельскохозяйственных культур представлены на рис. 3. Всего расчёты проводились для 20 тестовых участков, занятых посевом риса, в штатах Орисса и Карнатака и 20 тестовых участков, занятых посевами пшеницы, в штатах Харьяна и Мадхья Прадеш. Анализ полученных результатов позволяет заключить, что применение алгоритма перекалибровки модели для интеграции данных спутниковых измерений позволяет существенно увеличить согласованность модельного и фактического хода показателя NDVI в течение периода вегетации (в среднем наблюдается падение среднеквадратичного отклонения модельных от измеренных величин в два с половиной раза). При этом в большинстве случаев наблюдается и улучшение модельного прогноза урожайности, хотя этот эффект оказывается не столь значительным, как этого можно было бы ожидать.

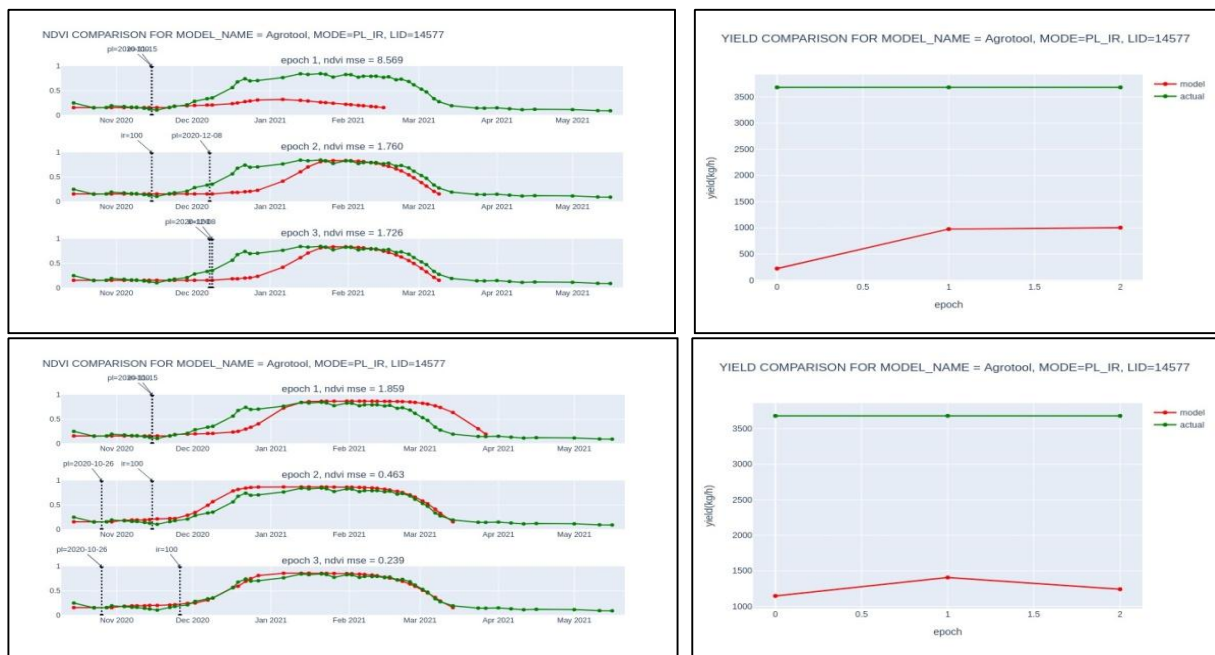


Рис. 3. Временная динамика индекса NDVI (левые графики) и урожайности (правые графики) для различных этапов процедуры интеграции данных фактических измерений (расчёт по базовому варианту® подбор даты сева® подбор расписания азотных подкормок). Посевы пшеницы, тестовый участок в штате Мадхья Прадеш. Зелёные кривые – фактические измерения, красные – данные модельных расчётов. Набор физиологических параметров для интенсивного (верхние графики) и позднеспелого тропического (нижние графики) идеотипов.

В целом, можно заключить, что, несмотря на недостаточную информативность спутниковых индексов, описанный алгоритм усвоения данных дистанционного

зондирования в некоторых случаях позволяет улучшить прогноз урожая в конкретной точке, полученный расчётами по динамической модели продукционного процесса сельскохозяйственного посева.

### **Заключение**

Описанные выше приёмы и алгоритмы были направлены на то, чтобы улучшить прогноз урожайности, даваемый стандартной динамической механистической моделью, используя дополнительные внешние данные фактических измерений текущего состояния изучаемого объекта (сельскохозяйственного посева). Но при этом сама механистическая модель может рассматриваться как источник данных, которые можно использовать как обучающую выборку в различных эмпирических моделях статистического характера. Действительно, каждый прогон динамической модели (например, в ходе процедуры перекалибровки) порождает новую логическую запись (кортеж), содержащих величины модельных величин измеряемых характеристик (например, NDVI) в даты фактических измерений и соответствующих им величин модельных урожаев. Набор данных формальных зависимостей можно использовать для вычисления дополнительной поправки к величине получаемой модельной оценки урожайности, вычисляемой по невязкам фактических и модельных величин индексов NDVI в даты фактических измерений (некий статический аналог фильтра Калмана). Схожие идеи составляют суть метода апостериорной коррекции, упомянутого выше.

Развивая концепцию «модель как источник данных», можно предложить различные перспективные идеи по возможному объединению двух парадигм – статистического и механистического моделирования в едином подходе, позволяющем взять наиболее ценное из каждого компонента.

Например, можно предложить обучать формальную нейронную сеть предсказывать по фактическим данным не сам урожай, а отклонение фактического урожая от его модельной оценки, даваемой механистической моделью. При этом, в идеале, ответственностью механистической модели становится учёт наиболее значимых и явных факторов, а статистическая модель берёт на себя учёт трудноформализуемых и тонких эффектов.

Другой перспективной идеей выглядит «подмешивание» в обучающую выборку алгоритмов искусственного интеллекта, состоящую изначально только из фактических измерений, некоторой доли кортежей, полученных в модельных расчётах. Наличие в общей выборке подобных «физически значимых» элементов может сгладить негативный эффект переобучения за счёт ошибок измерения, содержащихся в исходных данных. Наконец, значения некоторых внутренних переменных модели в промежуточных точках расчёта можно использовать в качестве дополнительных предикатов в моделях машинного обучения, так как они имеют большую связанность с предсказываемой величиной урожайности, чем различные косвенные показатели, доступные прямому инструментальному измерению.

Так или иначе, представляется, что лишь совместное использование моделей, основанных на знаниях и моделей, основанных на данных, может привести к успеху в такой сложной предметной области, как информатизация земледелия.

### **Литература**

1. **Р.А. Полуэктов, А.Г. Топаж, В. Миршель**, Сравнение эмпирического и теоретического подходов в математическом моделировании агросистем на примере описания процесса фотосинтеза. Математическое моделирование, 1998, том 10, номер 7. С.25-36.
2. **Полуэктов Р.А.** Динамические модели агроэкосистемы. Л.: Гидрометеиздат. 1991. – 312 с.

3. **Delécolle R., Maas S., Guérif M., Baret F.**, 1992. Remote sensing and crop production models: present trends. *ISPRS J. Photogramm.* 47, 145-161.
4. **Jin X., Kumar L., Li Z., Feng H., Xu X., Yang G., Wang J.**, 2018. A review of data assimilation of remote sensing and crop models. *Eur. J. Agron.* 92, 141-152.
5. **Топаж А.Г., Митрофанов Е.П.** Ассимиляция данных в имитационном моделировании экологических процессов методом минимизации корректирующих возмущений // Вестник СПбГУ. Серия «Прикладная математика. Информатика. Процессы управления», 2017. Т. 13. Вып. 3. С.326-338.
6. **Ding Y., Zhanga H., Zhaob K., Zheng X.** Investigating the accuracy of vegetation index-based models for estimating the fractional vegetation cover and the effects of varying soil backgrounds using in situ measurements and the PROSAIL model // *International Journal Of Remote Sensing*, 2017, vol. 38, №. 14, 4206-4223.
7. **Jacquemoud S., Verhoef W., Baret F., Vacour C., Zarco-Tejada P.J., Asner G.P., François C., Ustin S.L.** // PROSPECT + SAIL models: a review of use for vegetation characterization. *Remote Sensing of Environment*, 2009, 113, S56-S66.
8. **Berger K., Atzberger C., Danner M., D'Urso G., Mauser W., Vuolo F., et al.** Evaluation of the PROSAIL model capabilities for future hyperspectral model environments: A review study // *Remote Sensing*, 2018, 10(1), 85.