

УДК 519.85, 519.65, 63

Технология сбалансированной идентификации для выбора математической модели транспирации сосны

Соколов А.В.^{*1}, Болондинский В.К.^{2}, Волошинов В.В.^{***1}**

¹*Институт проблем передачи информации, Москва, Россия*

²*Институт Леса КарНЦ РАН, Петрозаводск, Россия*

Аннотация. Демонстрируется применение численной технологии оценки соответствия математической модели экспериментальным данным методом сбалансированной (оптимизационной) идентификации для сравнения различных моделей транспирации сосны. Количественной мерой оценки является погрешность перекрестной проверки. Возможности программной реализации технологии позволяют исследователю сформулировать задачу в одном текстовом файле, содержащем: уравнения математической модели (включая интегро-дифференциальные); список параметров и/или функций, подлежащих идентификации; ссылку на источник данных и дополнительные настройки метода. В итоге программный комплекс определит неизвестные параметры, функции и погрешности моделирования. Данная технология успешно применяется для моделирования явлений различной природы в биологии, медицине, физике и т.п.

Ключевые слова: математическое моделирование, сбалансированная идентификация, регуляризация, распределенные вычисления, платформа Everest, транспирация сосны.

ВВЕДЕНИЕ

Выбор модели исследуемого объекта или явления по имеющимся измерениям и формулировка требований к эксперименту, проводимому для проверки модели – две актуальные задачи научного познания мира.

Один и тот же объект (процесс, явление) может быть описан различными математическими моделями. При этом, сложность, размерность и подробность математического описания определяются как наличием качественных знаний о функционировании объекта (которые могут быть формализованы в виде математических утверждений), так и доступностью количественных данных, их объемом, подробностью, надежностью и точностью. Чем сложнее исследуемое явление, чем сложнее предложенная для его описания математическая модель, тем тщательнее должен быть поставлен соответствующий “подтверждающий” эксперимент, тем более подробными и точными должны быть измерения.

Обычно для поиска баланса между сложностью модели и доступными измерениями используется перебор: испытываются различные модели и наборы данных до тех пор,

*alexander.v.sokolov@gmail.com

**bolond@yandex.ru

***vv_voloshinov@iitp.ru

пока не будет достигнуто удовлетворительное соответствие, позволяющее построить модель, отвечающую целям исследования.

Единого подхода для решения таких задач не существует – в каждом отдельном случае используются различные параметризации моделей, различные процедуры идентификации и верификации, используется разнообразное программное обеспечение. Кроме того, часто критерий выбора не очевиден и решение об использовании той или иной модели принимается субъективно.

Нужна технология, максимально упрощающая принятие решения о выборе той или иной модели с минимально необходимым участием исследователя. Это позволит для каждой модели получать по единой методике количественную оценку соответствия математического описания и имеющихся измерений. Такой количественной оценкой может служить погрешность описания измерений, полученная методом сбалансированной идентификации [1, 2], основанным на идеях регуляризации и оценки погрешности методом перекрестной проверки. Программную реализацию этого метода будем далее называть SvF-технологией (от английских слов *Simplicity vs Fitting*).

Для применения этой технологии достаточно задать математическую модель (в виде системы уравнений, включая интегро-дифференциальные) и подготовить таблицу измерений. Далее формализуется критерий выбора неизвестных функций, включающий регуляризирующие слагаемые с коэффициентами, которые определяют соотношения между сложностью модели и ее близостью к измерениям. Возникающие при этом задачи оптимизации (возможно, вариационные задачи в функциональном пространстве) автоматически заменяются на их конечномерные аналоги в виде задач математического программирования. Для решения этих задач можно применять сервисы оптимизации [3] на основе распределенной платформы Everest [4]. В ходе итерационного вычислительного процесса определяются коэффициенты регуляризации [5], минимизирующие погрешность перекрестной проверки. В итоге вычисляются искомые параметры, функции и оценка погрешности моделирования, которая и используется как численный критерий соответствия модели и измерений.

В статье также демонстрируются возможности описать все задание по оценке конкретной модели в одном текстовом файле, что упрощает практическое применение SvF-технологии широким кругом специалистов в различных научно-технических областях. Этот файл-задание содержит: уравнения математической модели (включая интегро-дифференциальные); список неизвестных параметров и/или функций, подлежащих идентификации; ссылку на файл данных; критерий оптимизации и дополнительные настройки метода. Разработанная для этого специальная нотация является достаточно наглядной, близкой к математической формулировке задачи, и может содержать помимо формул различные инструкции (в т.ч. на языке Python), связанные с численным решением (работа с таблицами, параметрами дискретизации и др.). Примеры файлов-заданий для рассматриваемых моделей приведены в Приложениях 1 и 2. Подробное описание этой нотации приводится в документе http://bit.ly/SvF_UserGuide27, который также содержит инструкцию по установке программного обеспечения, требуемого для применения SvF-технологии.

Таким образом, использование технологии не требует от исследователей навыков программирования, в отличие от других коммерческих (ACADO, Boeing SOCS, MATLAB, Tomlab PROPT) и свободно доступных [6–9] систем анализа динамических и оптимизационных моделей. Все они предполагают использование языков программирования (C++, Fortran, Python и др.). Перевод постановки задачи из математической записи на указанные языки требует значительных временных затрат и специальных знаний даже от квалифицированного программиста.

Технология успешно использовалась в различных предметных областях для построения моделей: динамики радона в атмосфере, цифровой модели местности,

радиоактивного загрязнения и эрозии почв, динамики глюкозы и инсулина, быстрого нелокального переноса тепла в установках магнитного удержания плазмы, осциллятора с вязким трением, потока CO_2 на сфагновом верховом болоте, атмосферного переноса загрязнений и др.

В данной статье технология используется для моделирования процесса транспирации (испарения воды) сосны. Построение такой модели поможет решению одной из актуальных проблем выявления механизмов устойчивости лесных экосистем к внешним воздействиям, включая оценку чувствительности к изменениям климата и антропогенной деятельности. Прогноз реакции растений на эти факторы имеет первостепенное значение для прогнозирования последствий воздействия глобальных климатических процессов [10, 11]. Надежными индикаторами состояния природных экосистем и растений, позволяющими оценить не только динамику их изменения и развития, но также и охарактеризовать их устойчивость, являются составляющие водного баланса растений, в частности, транспирации [12, 13]. Технология применяется к нескольким моделям различной сложности и позволяет сравнить их соответствие экспериментальным данным.

Прогноз отклика древесных растений на изменение условий внешней среды невозможен без применения математических моделей, позволяющих оценить будущие изменения фитоценозов при различных сценариях климатических изменений.

ОБЪЕКТ ИССЛЕДОВАНИЯ И ИСТОЧНИК ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

Исследования проводились на базе полевого стационара Института леса КНЦ РАН, расположенного в 50 км к северу от г. Петрозаводска ($62^\circ 13'$ с.ш. и $34^\circ 10'$ в.д.). Район исследований относится к средней подзоне европейской тайги и имеет следующие климатические характеристики [14]: среднегодовая температура воздуха $+2.4^\circ\text{C}$, 105 дней имеют температуру выше $+10^\circ\text{C}$. Годовое количество осадков 680 мм.

Изучение газообмена (CO_2 и H_2O) проводилось в сосняке черничном свежем [15], входящим в экологический ряд сосновых древостоев с различным уровнем залегания грунтовых вод [14]. Исследуемое 55-летнее дерево имело высоту 14 м, диаметр 16 см, хорошо развитую корневую систему с якорными корнями, достигающими водоносных слоев. Для непрерывной автоматической регистрации транспирации побегов сосны использовалась многоканальная автоматическая установка на базе дифференциального психрометра [16].

По каждому каналу снималось примерно 10 показаний в час. По ним рассчитывались средние получасовые значения транспирации. Относительная погрешность регистрации измеряемых показателей оценивается величиной около 15 %. Полиэтиленовые камеры полукрытого типа устанавливали с южной стороны кроны на побегах 4–5-ой мутовки, Микрометеорологические исследования в фитоценозе проводили по стандартным методикам.

Таким образом, результатом проведенного комплексного эксперимента являются ряды измерений с шагом 30 минут (k – номер группы измерений):

- E_k – транспирация ($\text{ммоль H}_2\text{O м}^{-2} \text{с}^{-1}$);
- Q_k – полная солнечная радиация ($\text{мкммоль фотонов м}^{-2} \text{с}^{-1}$);
- T_k – температура воздуха ($^\circ\text{C}$);
- H_k – относительная влажность воздуха (%);
- VPD_k – дефицит давления водяного пара воздуха (Па).

Для моделирования использовался фрагмент данных с 20 июня по 23 июля. Всего 1632 точки. Введем формальные обозначения для этого временного ряда

$$D: \{E_k, Q_k, T_k, VPD_k, H_k, t_k = 0.5k\}, k \in K, K = \{0, 1, \dots, 1631\}, \quad (1)$$

который будем называть множеством измерений. Данные по E неполны (1537 точек), из-за технических проблем в работе оборудования. Это обстоятельство следует учитывать ниже, где будут встречаться выражения с перечислением значений E_k .

Например, при вычислении среднеквадратичного отклонения $\frac{1}{|K|} \sum_{k \in K} (E_k - E(t_k))^2$, где

$E(t_k)$ есть результат моделирования транспирации как функции времени. Разработанный программный комплекс, который применяется в расчетах, предусматривает автоматическую обработку измерений с подобными “пропусками”. Если в файле-задании имеется опция “useNaN” (см. приложения), то для приведенного среднеквадратичного отклонения суммирование проводится только по тем индексам k , для которых известно значение E_k . Соответственно, будет правильно скорректирован знаменатель – число слагаемых в сумме квадратов.

Для верификации моделей использовались и результаты другого исследования, проведенного с тем же деревом в тот же период времени. Это измерения суточной динамики потока воды в стволе (т.н. ксилемный транспорт), полученные тепловым методом постоянного нагрева [17].

ПРИМЕНЕНИЕ ТЕХНОЛОГИИ СБАЛАНСИРОВАННОЙ ИДЕНТИФИКАЦИИ ДЛЯ ВЫБОРА МОДЕЛИ ТРАНСПИРАЦИИ

В основе рассматриваемой технологии лежит метод сбалансированной идентификации, содержанием которого является поиск компромисса между точностью воспроизведения набора измерений и некоторой мерой сложности, выраженной регуляризирующим слагаемым в критерии оптимизации. Цель – выбор математической модели, сложность которой соответствует количеству и качеству (погрешности) измерений. Прежде, чем переходить к последовательному анализу моделей-претендентов, рассмотрим два относительно простых примера, поясняющих применение технологии.

Пример 1. Аппроксимация суточного хода транспирации.

Выберем из полного множества измерений (1) подмножество, состоящее из данных по транспирации за одни сутки (всего 49 точек, от 0 до 24 часов с шагом 0.5 часа):

$$D': \{E_k, t_k = 0.5k\}, k \in K', K' = \{0, 1, \dots, 48\}$$

Здесь и далее для обозначения измерений используются названия переменных с нижним индексом. Те же символы без нижнего индекса используются для обозначения переменных математических моделей, соответствующих измеряемым показателям.

Будем искать функцию, аппроксимирующую измерения, т.е. выберем в качестве математической модели пространство функций, дважды дифференцируемых на отрезке $[0, 24 \cdot 3600]$, $E = E(t)$, $E(\cdot) \in C^2[0, 24 \cdot 3600]$ (здесь и далее все функции предполагаются дважды дифференцируемыми).

На рисунке 1 приведены результаты моделирования суточной динамики транспирации сосны. Из бесконечного множества возможных функций, аппроксимирующих данные, приведены три. В подписи к рисунку 1 используются обозначения: σ_{cv} – погрешность перекрестной проверки (корень среднеквадратичной ошибки перекрестной проверки); σ_{cv}^* – оптимальная погрешность перекрестной проверки (погрешность моделирования измерений); $rmse^1$ – среднеквадратичная ошибка. Здесь и далее все погрешности даны в процентах от величины стандартного отклонения измерений (E_k) (квадратный корень из дисперсии), что обеспечивает

¹ rmse - от английского термина root-mean-square-error

нормировку от 0 % (при совпадении модельных значений со всеми измерениями) до 100 % (для простейшей модели в форме постоянной функции, равной среднему значению измерений).

Общая вычислительная схема метода, включающего вычисление значений σ_{cv} , σ_{cv}^* и $rmse$, приводится в статьях [1, 2]. Кратко поясним эту схему, в основе которой – задача минимизации функционала, зависящего от коэффициента регуляризации α (возможно – не одного, а нескольких, как в Примере 2):

$$E(\cdot) \in C^2[0, 24 \cdot 3600],$$

$$F(E, K', \alpha) = \frac{1}{|K'|} \sum_{k \in K'} (E_k - E(t_k))^2 + \alpha \int_0^{24 \cdot 3600} \left(\frac{d^2 E}{dt^2} \right)^2 dt \rightarrow \min_{E(\cdot)} \quad (2)$$

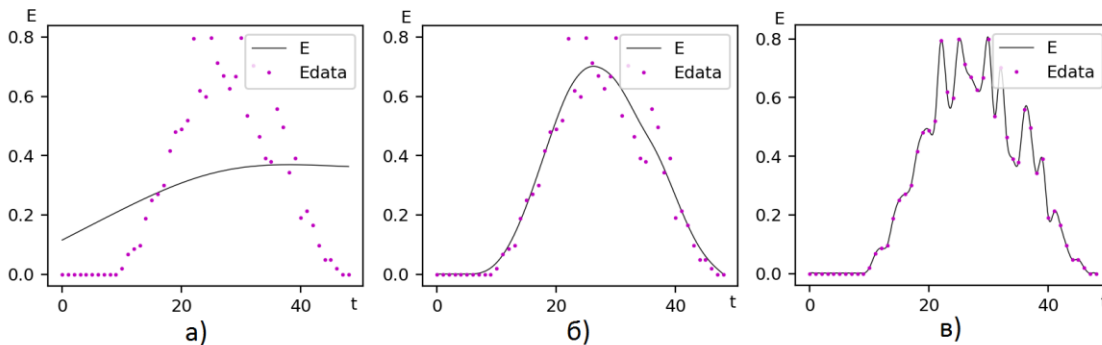


Рис. 1. Различные варианты аппроксимации функции транспирации $E(t)$: **а)** – слишком простая, “недообученная” модель ($\sigma_{cv} = 86.2\%$, $rmse = 85.3\%$); **б)** – сбалансированная модель ($\sigma_{cv}^* = 25.7\%$, $rmse = 21.8\%$); **в)** – слишком сложная, “переобученная” модель ($\sigma_{cv} = 35.0\%$, $rmse = 0\%$)

Первое слагаемое (2) – мера соответствия модели измерениям (среднеквадратичное отклонение), второе – регуляризирующая добавка, мера сложности модели [5], которой здесь является кривизна функции $E(t)$ (интеграл квадрата второй производной).

Замечание. В формуле (2) коэффициент регуляризации α имеет размерность (сек³). В дальнейшем размерности коэффициентов регуляризации не указываются, так как они используются только в алгоритме идентификации и сами по себе (и их размерности) не представляют интереса (см., напр. [5]).

Минимизируя функционал (2) для разных значений α , получим семейство решений

$$E_K^\alpha = \text{Arg min}_{E(\cdot)} \{F(E, K', \alpha)\}, \quad (3)$$

с различным соотношением “отклонение от измерений”/“сложность модели” (в зависимости от значения α). Например, так получены все кривые на рисунке 1. При больших значениях α второе слагаемое функционала (2) подавляет первое, и соответствующее решение является слишком простым, проходит далеко от измерений, т.е. плохо описывает исследуемое явление (рис. 1,а). При малых α наблюдается обратная ситуация: решение проходит слишком близко к измерениям, повторяет содержащиеся в них ошибки, является слишком сложным (рис. 1,в).

Естественно, возникает задача выбора “наилучшей” кривой из параметрического семейства решений (3). Для этого необходим критерий оценки качества решения, например, среднеквадратичная ошибка перекрестной проверки (cross-validation) [1, 2]. Значения σ_{cv} для трех кривых рисунка 1 приведены в подписи к рисунку. Сбалансированное решение (рис. 1,б) соответствует такому значению α^* , при котором достигается минимум погрешности перекрестной проверки σ_{cv}^* .

Опишем на данном примере процедуру перекрестной проверки и поиска оптимального решения. Заметим, что вместо всего множества K' в (2) можно использовать некоторое его подмножество, например, исключая из K' некоторый элемент k . Введем обозначение для соответствующей функции-решения:

$$E_{K' \setminus k}^{\alpha} = \text{Arg min}_{E(\cdot)} \{F(E, K' \setminus k, \alpha)\}, \quad k \in K' \quad (4)$$

Эти решения используются для расчета отклонений модели от измерений (на исключенной из множества K' точке) $E_k - E_{K' \setminus k}^{\alpha}(t_k)$, $k \in K'$, и среднеквадратичной погрешности перекрестной проверки (при фиксированном α) по всем точкам:

$$\sigma_{cv}(\alpha) = \sqrt{\frac{1}{|K'|} \sum_{k \in K'} (E_k - E_{K' \setminus k}^{\alpha}(t_k))^2} \quad (5)$$

Решение задач (4) является нижним уровнем двухуровневой задачи минимизации, где на верхнем уровне требуется найти значение α^* коэффициента регуляризации (здесь – одного, в общем случае, как будет видно из дальнейшего, – нескольких), при котором погрешность (5) минимальна. Положим $\sigma_{cv}^* = \sigma_{cv}(\alpha^*)$:

$$\alpha^* = \text{Arg min}_{\alpha} \sqrt{\frac{1}{|K'|} \sum_{k \in K'} (E_k - E_{K' \setminus k}^{\alpha}(t_k))^2}; \quad \sigma_{cv}^* = \sqrt{\frac{1}{|K'|} \sum_{k \in K'} (E_k - E_{K' \setminus k}^{\alpha^*}(t_k))^2} \quad (6)$$

Значение величины σ_{cv}^* , которую будем называть *погрешностью моделирования данных*, можно трактовать как меру предсказательной точности модели. Наконец, для найденного α^* , определим оптимальное решение задачи (2) на полном наборе измерений K' и соответствующее среднеквадратичное отклонение от измерений:

$$E^* = E_{K'}^{\alpha^*} = \text{Arg min}_{E(\cdot)} \{F(E, K', \alpha^*)\}, \quad k \in K'; \quad \text{rmse}^* = \sqrt{\frac{1}{|K'|} \sum_{k \in K'} (E_k - E^*(t_k))^2} \quad (7)$$

Численный метод поиска оптимального значения α^* (решения задачи верхнего уровня) описан в статьях [1, 2]. Он основан на идеях суррогатной оптимизации.

Пример 2. Транспирация E как функция солнечной радиации Q и влажности воздуха H . Здесь и далее используется полный набор измерений (1), где для небольшого подмножества индексов отсутствуют значения E_k , что автоматически учитывается программой при вычислении выражений, включающих E_k . В качестве математической модели рассмотрим функцию двух переменных: $E = E(Q, H)$.

Критерий выбора аналогичен Примеру 1, но содержит теперь частные производные и зависит от двух коэффициентов α_1 и α_2 :

$$F(E, K, \alpha_1, \alpha_2) = \frac{1}{|K|} \sum_{k \in K} (E_k - E(Q_k, H_k))^2 + \int_{Q^{\min}}^{Q^{\max}} \int_{H^{\min}}^{H^{\max}} \left[\alpha_1^2 \left(\frac{\partial^2 E}{\partial Q^2} \right)^2 + 2\alpha_1 \alpha_2 \left(\frac{\partial^2 E}{\partial Q \partial H} \right)^2 + \alpha_2^2 \left(\frac{\partial^2 E}{\partial H^2} \right)^2 \right] dQ dH$$

Для оценки погрешности моделирования здесь и далее используется процедура перекрестного оценивания с разбиением множества измерений на 34 подмножества: каждое содержит измерения за одни сутки (до 48 точек). Полученная сбалансированная модель приведена на рисунке 2,б.

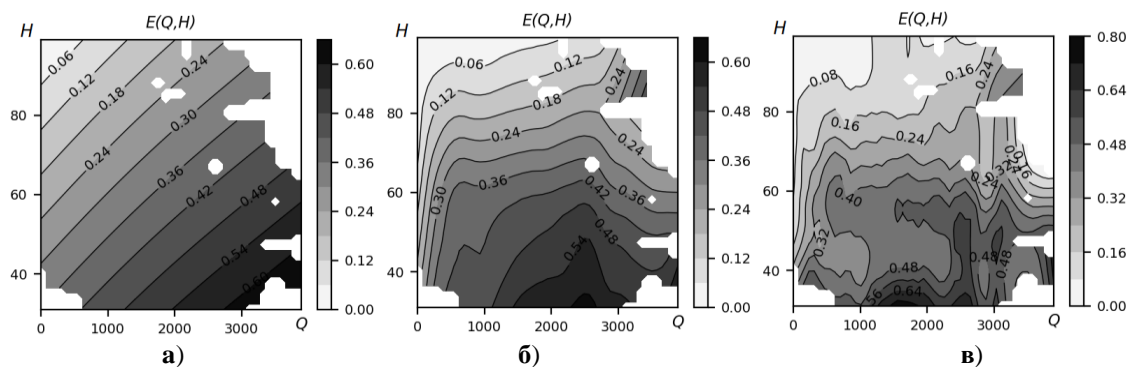


Рис. 2. Изолинии зависимости транспирации (E) от радиации (Q) и влажности воздуха (H). Различные варианты аппроксимации: **а)** – слишком простая, “недообученная” модель ($\sigma_{cv} = 57.8\%$, $rmse = 56.9\%$); **б)** – сбалансированная модель ($\sigma_{cv}^* = 51.5\%$, $rmse = 48.4\%$); **в)** – слишком сложная, “переобученная” модель ($\sigma_{cv} = 51.9\%$, $rmse = 43.8\%$). Здесь и далее на рисунках отображаются только те области (пространства аргументов), рядом с которыми имеются измерения. Белые области (и пятна) соответствуют зонам отсутствия данных.

Перейдем к более сложным моделям процесса транспирации сосны. Будем использовать следующие обозначения:

$VPD = VP_{100\%} - VP_{атм}$ – дефицит давления водяного пара воздуха (Па);

$VP_{100\%}$ – давление насыщенного водяного пара (Па);

$VP_{атм}$ – давление водяного пара в атмосфере (Па);

Ust – устьичная проводимость (моль·м⁻²·с⁻¹);

$A = 0.010017$ (ммольH₂O·моль⁻¹·Па⁻¹) – коэффициент согласования единиц измерения;

W – запас влаги в стволе (л);

WD – дефицит воды в стволе (сколько не хватает до максимума) (л);

F_{WF} – (ксилемный) поток воды в стволе (л·час⁻¹);

WF – коэффициент потока воды в стволе (час⁻¹);

Reg – безразмерный множитель (зависимость проводимости устьиц от WD),

$Sneedles$ – площадь хвои (м²) (оценивается величиной 32 м²).

Рассмотрим 11 моделей, описывающих процесс транспирации с различной степенью подробности. По единой методике для каждой модели найдем оптимально сбалансированное решение (определим коэффициенты регуляризации α , неизвестные функции и параметры моделей) и рассчитаем погрешность моделирования измерений. В качестве функционала, определяющего критерий выбора, используем сумму среднеквадратичного отклонения и меру сложности модели, состоящую из суммы взвешенных мер сложности неизвестных функций, входящих в модель. Мерами сложности функций выберем интегралы квадратов вторых производных.

ФУНКЦИОНАЛЬНЫЕ ЗАВИСИМОСТИ

Начнем с простых моделей – различных функциональных зависимостей транспирации от внешних параметров.

Модель $E(Q)$. Транспирация как функция солнечной радиации. Этой простой модели соответствует файл-задание (для расчета) из двух строк:

DataFile = test.txt

Var: E (Q)

Здесь первая строка определяет имя файла с измерениями E_k и Q_k , а вторая – объявляет неизвестную функцию. В файле не задан критерий идентификации, поэтому

используется критерий по умолчанию и при заданном коэффициенте регуляризации α функция $E(Q)$ определяется как решение следующей оптимизационной задачи:

$$E = E(Q) \in C^2[0, Q^{\max}],$$

$$F(E, K, \alpha) = \frac{1}{|K|} \sum_{k \in K} (E_k - E(Q_k))^2 + \alpha \int_0^{Q^{\max}} \left(\frac{d^2 E}{dQ^2} \right)^2 dQ \rightarrow \min_{E(\cdot)}$$

Полученная в результате функция изображена на рисунке 3, а оценка погрешности составляет: $\sigma_{cv}^* = 67.06\%$.

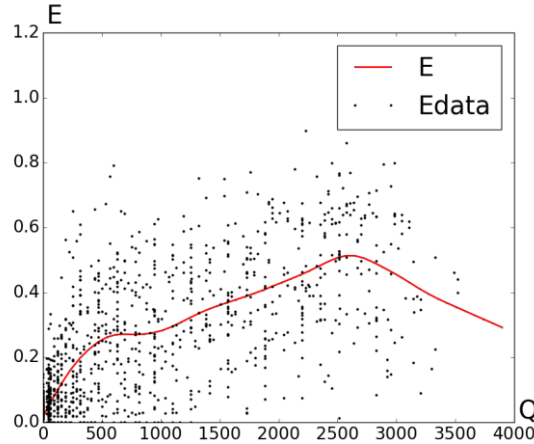


Рис. 3. Транспирация E как функция радиации (Q): функция (E) и измерения (E_{data}).

Модель $E(Q, T)$. Транспирация как функция солнечной радиации и температуры. Приведем соответствующую модель и критерий выбора:

$$E = E(Q, T) \in C^2([0, Q^{\max}] \times [T^{\min}, T^{\max}]),$$

$$F(E, K, \alpha_1, \alpha_2) = \frac{1}{|K|} \sum_{k \in K} (E_k - E(Q_k, T_k))^2 +$$

$$+ \int_0^{Q^{\max}} \int_{T^{\min}}^{T^{\max}} \left[\alpha_1^2 \left(\frac{\partial^2 E}{\partial Q^2} \right)^2 + 2\alpha_1 \alpha_2 \left(\frac{\partial^2 E}{\partial Q \partial T} \right)^2 + \alpha_2^2 \left(\frac{\partial^2 E}{\partial T^2} \right)^2 \right] dQ dT \rightarrow \min_{E(\cdot)}$$

Файл-задание для расчета по этой модели приведен в Приложении 1. Изолинии полученной функции изображены на рисунке 4,а; погрешность $\sigma_{cv}^* = 58.13\%$.

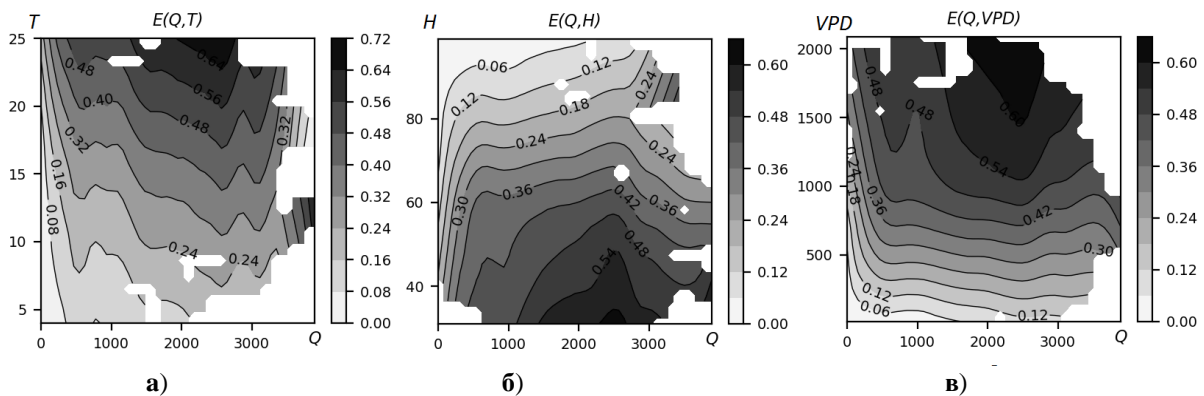


Рис. 4. Изолинии зависимости транспирации (E) от: а) – радиации (Q) и температуры (T); б) – радиации (Q) и влажности (H); в) – радиации (Q) и дефицита упругости водяного пара (VPD).

Заметим, что с добавлением температуры точность заметно возросла.

Модель $E(Q, H)$. Транспирация как функция радиации и влажности. Результатом применения технологии является функция, изображенная на рисунке 4,б, и погрешность $\sigma_{cv}^* = 51.51\%$. Можно сделать вывод, что радиация и влажность лучше описывают измерения (процесс транспирации), чем радиация и температура.

Модель $E(Q, VPD)$. Транспирация как функция радиации и дефицита упругости водяного пара. Результатом является функция на рисунке 4,в и погрешность $\sigma_{cv}^* = 49.13\%$. Это наименьшая погрешность среди моделей в виде функции двух переменных.

Учет устьичной проводимости

Усложним модель, используя качественную информацию о моделируемом объекте: знание об устьичной природе регуляции потока водяного пара [18, 19]. В простейшем случае, когда температура хвои близка к температуре воздуха, поток водяного пара (E) из растения в атмосферу определяется по формуле [20]:

$$E = Ust \cdot (VP_{100\%} - VP_{атм}) \cdot A = Ust \cdot VPD \cdot A,$$

где Ust – проводимость устьицы, VPD – дефицит давления водяного пара, $A = 0.010017$ (ммоль H_2O ·моль $^{-1}$ ·Па $^{-1}$) – коэффициент согласования единиц измерения.

Расчет A проведен в метрической системе СГС. При этом использовались следующие соотношения: ммоль H_2O ·м $^{-2}$ ·с $^{-1}$ = $10^{-3} \cdot 18 \text{ г} / (10^4 \text{ см}^2 \cdot \text{с}) = 18 \cdot 10^{-7} \text{ г} \cdot \text{см}^{-2} \cdot \text{с}^{-1}$, моль м $^{-2}$ ·с $^{-1}$ = 2.404 см·с $^{-1}$, 1 Па приблизительно соответствует $7.5 \cdot 10^{-9} \text{ г} \cdot \text{см}^{-3}$ (с точностью до двух значащих цифр из психрометрических таблиц [21]). В результате получим $18 \cdot 10^{-7} E = 2.404 \cdot Ust \cdot 7.5 \cdot 10^{-9} \cdot VPD$. Отсюда $A = 2.404 \cdot 7.5 \cdot 10^{-9} / (18 \cdot 10^{-7}) = 0.010017$.

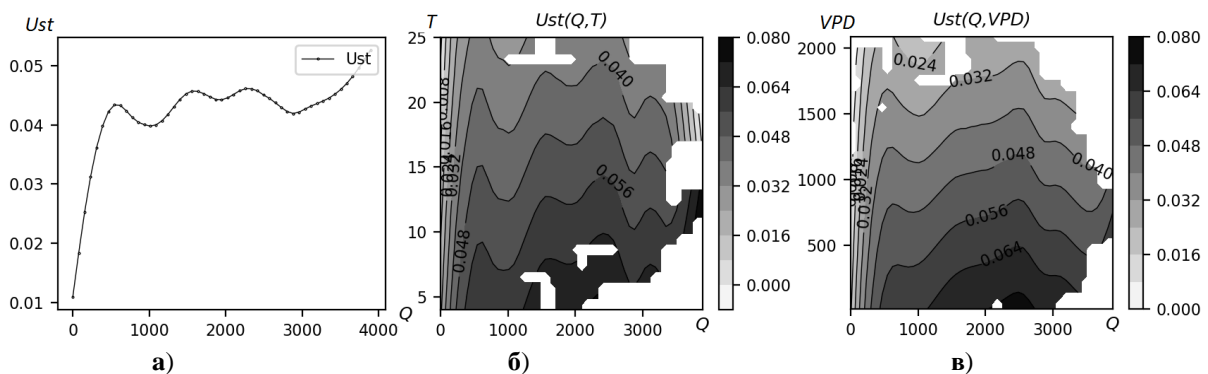


Рис. 5. Устьичная проводимость (изолинии) (Ust) как функция: **а** – радиации (Q); **б** – функция радиации (Q) и температуры (T); **в** – радиации (Q) и дефицита упругости водяного пара (VPD).

Модель $E = Ust(Q) \cdot VPD \cdot A$. Устьичная проводимость как функция радиации. Результат: функция, изображенная на рисунке 5,а и погрешность: $\sigma_{cv}^* = 55.35\%$. Мы получили заметное уменьшение погрешности относительно модели $E(Q)$, т.е. учет устьичной проводимости – правильное решение.

Модель $E = Ust(Q, T) \cdot VPD \cdot A$. Устьичная проводимость как функция радиации и температуры. Результат: функция на рисунке 5,б и погрешность: $\sigma_{cv}^* = 53.74\%$. Таким образом, учет устьичной проводимости заметно уменьшает погрешность по сравнению с моделью $E(Q, T)$, где погрешность была равной 58.13 %.

Модель $E = Ust(Q, VPD) \cdot VPD \cdot A$. Устьичная проводимость как функция радиации и дефицита упругости водяного пара. Результат: функция, изображенная на рисунке 5, в и погрешность $\sigma_{cv}^* = 48.55 \%$. Таким образом, учет устьичной проводимости уменьшает погрешность моделирования.

Динамические модели

Перейдем к динамическим моделям, описывающим динамику воды в стволе, через процессы транспирация и поглощения воды из почвы. Как известно, [22], на величины транспирации влияют не только внешние параметры, но и внутреннее состояние растения. В частности, имеется обратная связь уменьшающая транспирацию (путем частичного закрытия устьиц) при недостатке влаги в растении [20].

Динамическая модель $Ust(Q, VPD) \cdot Reg(WD) \cdot F_{WF}(WD)$. Функции регулирования транспирации и потока воды из почвы зависят от дефицита воды в растении.

$$\left\{ \begin{array}{l} \frac{dW}{dt} = F_{WF}(WD) - E(t) \cdot Sneedles \cdot 6.48 \cdot 10^{-2} \\ E = Ust(Q, VPD) \cdot Reg(WD) \cdot VPD \cdot 0.010017 \\ W \leq W^{\max} = 75 \\ W \geq W^{\min} = 68 \\ WD = W^{\max} - W \\ WF(0) = 0; \quad WF \geq 0 \\ Reg(0) = 1; \quad Reg \leq 1; \quad Reg \geq 0 \\ Sneedles = 32 \end{array} \right.$$

Для описания динамики запасов воды в растении используем дифференциальное уравнение (первое равенство), где: первое слагаемое описывает поступление воды из почвы (предполагается, что влажность почвы меняется несущественно), второе – расход воды в процессе транспирации (множитель $6.48 \cdot 10^{-2} = 0.001 \cdot 18 / 1000 \cdot 3600$ соответствует переходу от единиц измерения транспирации (ммоль $H_2O \cdot m^{-2} \cdot c^{-1}$) в эксперименте к единицам, используемым в модели (кг $\cdot m^{-2} \cdot час^{-1}$)). Кроме того, добавим мультипликативный множитель, регулирующий интенсивность транспирации в зависимости от дефицита воды (второе уравнение).

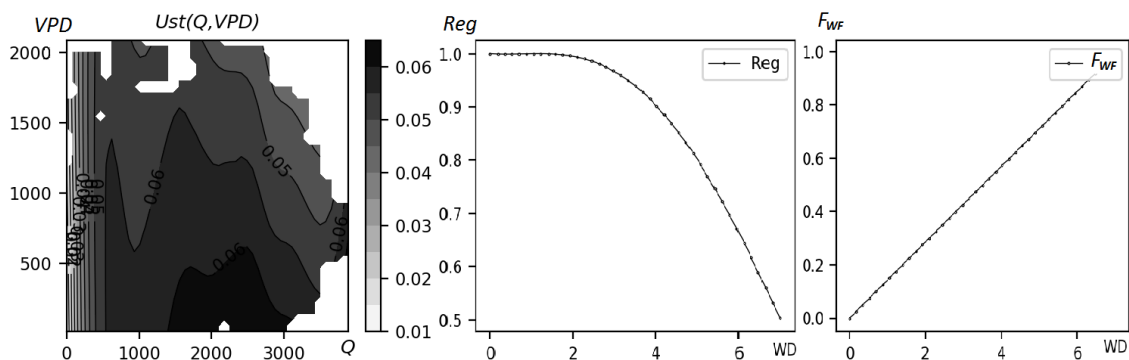


Рис. 6. Функции, определяющие динамику воды: устьичная проводимость (Ust) как функция радиации (Q) и дефицита упругости водяного пара (VPD), функция регулирования устьичной проводимости стволе (Reg) в зависимости от дефицита воды в стволе (WD), функция скорости поступления воды из почвы (F_{WF}) в зависимости от дефицита воды в стволе (WD).

Рассматриваемая модель содержит пять неизвестных функций: $E(t)$, $W(t)$, $Ust(Q,VPD)$, $Reg(WD)$ и $F_{WF}(WD)$. Результаты расчетов представлены на рисунке 6, погрешность $\sigma_{cv}^* = 47.72\%$. Погрешность заметно снизилась.

Динамическая модель $Ust(Q,VPD) \cdot Reg(WD) \& WF \cdot WD$. Из рисунка 6 видно, что полученная зависимость $F_{WF}(WD)$ потока воды в растение из почвы от водного дефицита близка к линейной. Заменяем ее на $F_{WF}(WD) = WF \cdot WD$, где коэффициент потока воды в стволе WF имеет размерность (час^{-1}) и тем самым упростим модель. Ниже приводится только измененное первое уравнение модели

$$\begin{cases} \frac{dW}{dt} = WF \cdot WD - E \cdot Sneedles \cdot 6.48 \cdot 10^{-2} \\ \dots \end{cases}$$

В результате функции $Ust(Q,VPD)$ и $Reg(WD)$ почти не изменились, и было определено значение параметра $WF = 0.142167$, которое в дальнейших моделях почти не изменялось. Погрешность $\sigma_{cv}^* = 47.71\%$ даже немного уменьшилась, что подтверждает корректность сделанного упрощения.

Динамическая модель $Ust(Q) \cdot Hmult(VPD) \cdot Reg(WD) \& WF \cdot WD$. Анализ графика зависимости $Ust(Q,VPD)$ позволяет сделать предположение о мультипликативной зависимости устьичной регуляции от переменных Q и VPD . Проверим эту гипотезу. Рассмотрим модель (приводятся только два первых уравнения):

$$\begin{cases} \frac{dW}{dt} = WF \cdot WD - E \cdot Sneedles \cdot 6.48 \cdot 10^{-2} \\ E = Ust(Q) \cdot Hmult(VPD) \cdot VPD \cdot Reg(WD) \cdot 0.010017 \\ \dots \end{cases}$$

Результат приведен на рисунке 7, $\sigma_{cv}^* = 47.45\%$. Погрешность уменьшилась. Примем гипотезу о мультипликативной зависимости.

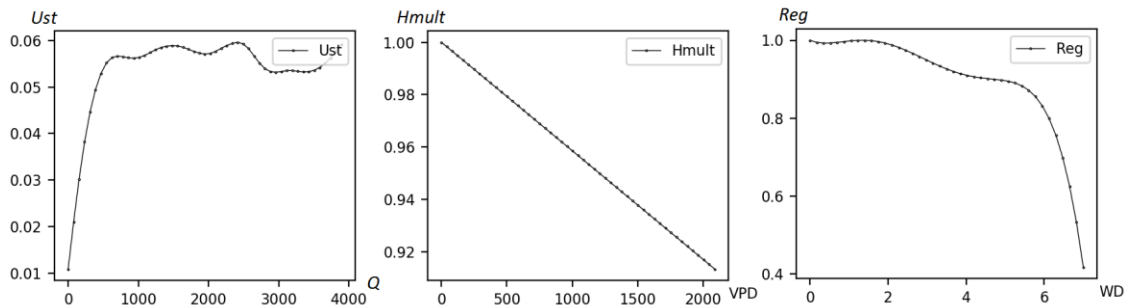


Рис. 7. Функции, определяющие динамику воды: устьичная проводимость (Ust) как функция радиации (Q), мультипликативный множитель ($Hmult$) зависимости от дефицита упругости водяного пара (VPD), функция регулирования устьичной проводимости (Reg) в зависимости от дефицита воды в стволе (WD).

Динамическая модель $Ust(Q) \cdot Reg(WD) \& WF \cdot WD$. Изменения мультипликативного множителя $Hmult$ невелики, положим его равным единице и рассмотрим модель:

$$\left\{ \begin{array}{l} \frac{dW}{dt} = WF \cdot WD - E \cdot Sneedles \cdot 6.48 \cdot 10^{-2} \\ E = Ust(Q) \cdot Reg(WD) \cdot VPD \cdot 0.010017 \\ W \leq W^{\min} = 75 \\ W \geq W^{\max} = 68 \\ WD = W^{\max} - W \\ Reg(0) = 1; Reg \leq 1; Reg \geq 0 \\ Sneedles = 32 \end{array} \right.$$

Модель содержит пять неизвестных функций: $E(t)$, $W(t)$, $WD(t)$, $Ust(Q)$, $Reg(WD)$ и неизвестный числовой параметр WF . В процедуре перекрестного оценивания для определения неизвестных модели и двух параметров α используется критерий выбора:

$$F(E, W, Ust, Reg, WF, K, \alpha_1, \alpha_2) = \frac{1}{|K|} \sum_{k \in K} (E_k - E(t_k))^2 + \\ + \alpha_1 \int_0^{Q^{\max}} \left(\frac{d^2 Ust}{dQ^2} \right)^2 dQ + \alpha_2 \int_{WD^{\min}}^{WD^{\max}} \left(\frac{d^2 Reg}{dWD^2} \right)^2 dWD$$

Текстовое описание задания расчета для этой динамической модели приведено в Приложении 2. Графики функций аналогичны тем, что представлены на рисунке 7, но с меньшей оценкой погрешности: $\sigma_{cv}^* = 47.33\%$.

Таким образом, для описания процесса транспирации целесообразно выбрать модель $Ust(Q) \cdot Reg(WD) \cdot WF \cdot WD$ – она лучше других соответствует имеющимся измерениям, о чем свидетельствует наименьшая погрешность моделирования измерений.

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Созданная программная реализация технологии сбалансированной идентификации применялась для сравнения набора моделей, описывающих природное явление транспирации сосны. Расчеты проводились на основе реальных экспериментальных данных полевого стационара Института леса КНЦ РАН. В качестве численного критерия оценки использовалась погрешность моделирования данных.

В начале выбор моделей делался от простого к сложному: функция одного переменного, различные функции двух переменных, модель устьичной проводимости и, наконец, динамическая модель воды в стволе. Направление усовершенствования (усложнения) модели выбиралось на основе изменения погрешности моделирования измерений – уменьшение погрешности рассматривалось как подтверждение правильности выбора (см. табл. 1). Наконец, когда для описания измерений было выбрано самое широкое семейство моделей – динамическая модель воды в стволе, начался обратный процесс – этап упрощения модели, состоящий в редукции некоторых функций, определяющих динамику воды. Адекватность редукции проверялась по изменению погрешности моделирования измерений.

Наконец, сравним полученные модельные кривые с результатами другого независимого исследования, проведенного с тем же деревом в тот же период времени. На рисунке 8 приводятся суточные графики потока воды в стволе (т.н. ксилемный транспорт), полученные экспериментально [17] и рассчитанные по модели $Ust(Q) \cdot Reg(WD) \cdot WF \cdot WD$. Нетрудно увидеть качественное соответствие графиков.

Таблица 1. Точность различных моделей: σ_{cv}^* – погрешность моделирования измерений (в % дисперсии); *rmse* – среднеквадратичная ошибка (в % дисперсии)

| Модель | σ_{cv}^* , % | rmse, % |
|--|---------------------|---------|
| $E(Q)$ | 67.06 | 64.03 |
| $E(Q,T)$ | 58.13 | 54.79 |
| $E(Q,H)$ | 51.51 | 48.40 |
| $E(Q,VPD)$ | 49.13 | 44.86 |
| $Ust(Q) \cdot VPD \cdot A$ | 55.35 | 51.37 |
| $Ust(Q,T) \cdot VPD \cdot A$ | 53.74 | 49.48 |
| $Ust(Q,VPD) \cdot VPD \cdot A$ | 48.55 | 45.22 |
| $Ust(Q,VPD) \cdot Reg(WD) \& F_{wf}(WD)$ | 47.72 | 44.90 |
| $Ust(Q,VPD) \cdot Reg(WD) \& WF \cdot WD$ | 47.71 | 44.91 |
| $Ust(Q) \cdot Hmult(VPD) \cdot Reg(WD) \& WF \cdot WD$ | 47.45 | 44.41 |
| $Ust(Q) \cdot Reg(WD) \& WF \cdot WD$ | 47.33 | 44.42 |

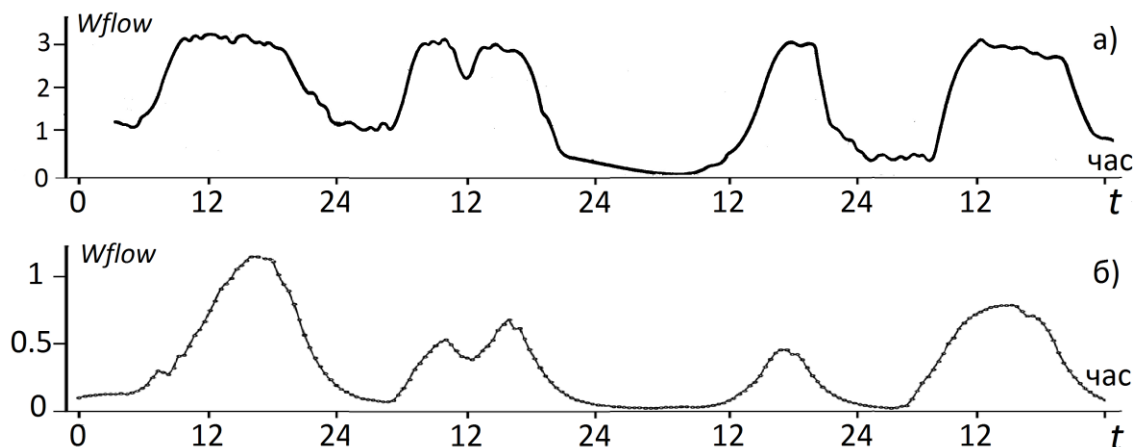


Рис. 8. Поток воды в растение (фрагмент за четверо суток): **а)** Поток воды (*Wflow*) в стволе (т.н. “ксилемный транспорт”) в условных единицах. Независимые (неиспользованные при построении модели) экспериментальные данные [17]; **б)** Поступление воды из почвы ($Wflow = WF \cdot WD$, определяется дефицитом влаги в растении *WD* умноженным на константу *WF*) л·час⁻¹. Расчеты по модели $Ust(Q) \cdot Reg(WD) \& WF \cdot WD$.

Построенная в итоге модель соответствует современным представлениям о регуляции влаги в растении [10–20, 22], она показала хорошее качественное совпадение с независимыми экспериментальными данными.

ЗАКЛЮЧЕНИЕ

Проведенное исследование по моделированию процесса транспирации, кроме решения актуальной задачи исследования лесных экосистем, позволило усовершенствовать программную реализацию SvF-технологии и подтвердило ее эффективность – без нее эта работа потребовала бы гораздо больше времени и усилий.

Предлагаемая технология сбалансированной идентификации является достаточно универсальной и востребованной в различных прикладных исследованиях. Однако, ее практическое применение требует многократного решения нелинейных задач математического программирования и значительных вычислительных ресурсов. Отдельной проблемой также является автоматизация подготовки исходных данных

задач для отправки их решателям и обработка результатов оптимизации. К настоящему времени указанные проблемы “в первом приближении” решены:

- разработан несложный синтаксис (нотация), формализующий описание исследуемой математической модели;
- разработана Python-программа, которая с помощью пакета Ruomo преобразует указанную нотацию в описание задач математического программирования и готовит NL-файлы для отправки AMPL-совместимым решателям, а также извлекает результаты решения из SOL-файлов, полученных от решателей;
- проблема привлечения вычислительных ресурсов решается за счет отправки указанных NL-файлов в распределенную среду сервисов оптимизации на платформе Everest, причем независимые задачи могут решаться параллельно пулом подключенных с Everest ресурсов (персональные компьютеры, отдельные серверы, кластеры, и т.п.).

Работа выполнена сотрудниками ИППИ РАН при финансовой поддержке РФФИ проект 16-11-10352 (апробация нотации записи интегро-дифференциальных уравнений, в разд. «Применение технологии сбалансированной идентификации для выбора модели транспирации» подраздел «Динамические модели»); РФФИ проект 18-07-01175а (разд. «Применение технологии сбалансированной идентификации для выбора модели транспирации» подразделы «Пример 1», «Пример 2», «Функциональные зависимости» и «Учет устьичной проводимости»); РФФИ проект 17-04-01087а (разд. «Объект исследования и источник экспериментальных данных», предложения по составу и структуре моделей транспирации).

ПРИЛОЖЕНИЕ 1.

ПРИМЕР ФАЙЛА-ЗАДАНИЯ РАСЧЕТА ПО SvF-ТЕХНОЛОГИИ ДЛЯ МОДЕЛИ $E(Q, T)$

```
CVNumOfIter 21                # Number of bi-level optimization iteration

useNaN                        # use NaN values in experimental data

Select ROWNUM AS t, I AS Q, Ta AS T, WPD AS VPD, Enan AS E, Pnan AS P, Dat, NN, PRel, ERel \
  from ../../DATA/Phot-5short.xlsx

MakeSets_byParam Dat 11      # CV procedure parameters

Grid:  gQ = [ , , -50, iq, Q ]
       gT = [ , , -50, it, T ]
Grid:  QT = DataSet (gQ,-2.6, gT,-2.6)

Var:  E ( QT );

Obj:  E.Complexity ( Penal[0],Penal[1] ) + E.MSDnan()

Draw

EOF
```

ПРИЛОЖЕНИЕ 2.

ПРИМЕР ФАЙЛА-ЗАДАНИЯ РАСЧЕТА ПО SvF-ТЕХНОЛОГИИ ДЛЯ ДИНАМИЧЕСКОЙ МОДЕЛИ $Ust(Q) \cdot Reg(WD) \& WF \cdot WD$

```

RunSolver ServerParallel      # To solve independent opt. problems in parallel
minH2O = 68.0                 # kg
maxH2O = 75.0                 # kg
delH2O = maxH2O-minH2O       # 7 kg
Sneedles = 32                 # m2 tree needles area
transMult = Sneedles * 6.48e-2 # kg/h 0.001*18*3600/1000
t_max = 1631                  # 34 days
bNN = 2257                    # 20.06
eNN = bNN + t_max             # 23.07
useNaN                         # use NaN values

```

```

Select ROWNUM AS t, I AS Q, Ta AS T, WPD AS VPD, Enan AS E, Dat, NN \
from ../DATA/Phot-4.xlsx \
where NN >= bNN and NN <= eNN

```

```

MakeSets_byParam Dat          # CV procedure parameters

```

```

GRID: t = [ 0, t_max, 1 ]     # grid for Time (first=0)
      gWD = [ 0, 7, -40 ]     # grid for WD

```

```

Var: E ( t )                  # Traspiration mmolH2O/m2/c
      H2O( t )                 # Water kg
      WD ( t ); >=0; <=delH2O # Water deficit kg

```

```

Param: VPD( t )
        Q( t )

```

```

Var: Ust ( Q ); Q ∈ [0, , -50]
      Reg ( gWD ); PolyPow = 5
      WF

```

```

EQ: E(t) = Ust(Q(t)) · Reg(WD(t)) · VPD(t) · 0.010017
     d/dt(H2O(t)) = - E(t) · transMult + WF · WD(t)
     H2O(0) = maxH2O-0.5 # initial condition
     WD(t) = maxH2O-H2O(t)
     Reg(0.0) = 1
     Reg(gWD) <= 1

```

```

Obj: Ust.Complexity (Penal[0]) + Reg.Complexity (Penal[1]) + E.MSDnan()

```

```

EOF

```


СПИСОК ЛИТЕРАТУРЫ

1. Соколов А.В., Волошинов В.В. Выбор математической модели: баланс между сложностью и близостью к измерениям. *International Journal of Open Information Technologies*. 2018. Т. 6. № 9. С. 33–41.
URL: <http://injoit.org/index.php/j1/article/view/612> (дата обращения: 25.12.2019).
2. Sokolov A., Voloshinov V. Balanced Identification as an Intersection of Optimization and Distributed Computing. *arXiv:1907.13444 [math.OC]*. 2019.
3. Smirnov S., Voloshinov V., Sukhosroslov O. Distributed Optimization on the Base of AMPL Modeling Language and Everest Platform. In: *Procedia Computer Science*. 2016. V. 101. P. 313–322.
4. Sukhoroslov O., Volkov S., Afanasiev A. A Web-Based Platform for Publication and Distributed Execution of Computing Applications. In: *Parallel and Distributed Computing: Proc. 14th International Symposium on IEEE*. Cambridge, USA; 2015. P. 175–184.
5. Тихонов А.Н. О математических методах автоматизации обработки наблюдений. В: *Проблемы вычислительной математики*. М.: Изд-во МГУ, 1980. С. 3–17.
6. Nicholson B., Sirola J.D., Watson J.-P., Zavala V.M., Biegler L.T. pyomo.dae: a modeling and automatic discretization framework for optimization with differential and algebraic equations. *Mathematical Programming Computation*. 2018. V. 10. No. 2. P. 187–223.
7. Schoukens J., Ljung L. Nonlinear System Identification: A User-Oriented Roadmap. *arXiv: 1902.00683 [cs.SY]*. 2019.
8. Clewley R. Hybrid models and biological model reduction with PyDSTool. *PLoS computational biology*. 2012. V. 8. No. 8. Article No. e1002628. doi: [10.1371/journal.pcbi.1002628](https://doi.org/10.1371/journal.pcbi.1002628).
9. Guzmán J.L., Rrivera D.E., Dormido S., Berenguel M. ITSIE: An interactive software tool for system identification. *Adv. Eng. Softw.* 2012. V. 45. No. 1. P. 115–123. doi: [10.1016/j.advengsoft.2011.09.013](https://doi.org/10.1016/j.advengsoft.2011.09.013).
10. Niinemets Ü., Anten N.P.R. Packing the photosynthetic machinery: from leaf to canopy. In: *Photosynthesis in silico Advances in Photosynthesis and Respiration*. Eds. A. Laisk, L. Nedbal, Govindjee. Netherlands: Springer, 2009. P. 363–399.
11. Olchev A.V., Deshcherevskaya O.A., Kurbatova Yu.A., Molchanov A.G., Novenko E.Yu., Pridacha V.B., Sazonova T.A. CO₂ and H₂O exchange in the forest ecosystems of southern taiga under climate changes. *Doklady Biological Sciences*. 2013. V. 450. P. 173–176.
12. Wieser G., Leo M., Oberhuber W. Transpiration and canopy conductance in an inner alpine Scots pine (*Pinus sylvestris* L.) forest. *Flora*. 2014. V. 209. No. 9. P. 491–498. doi: [10.1016/j.flora.2014.06.012](https://doi.org/10.1016/j.flora.2014.06.012).
13. Wieser G., Gruber A., Oberhuber W. Growing season water balance of an inner alpine Scots pine (*Pinus sylvestris* L.) forest. *IForest*, 2018. V. 11. P. 469–475. doi: [10.3832/ifor2626-011](https://doi.org/10.3832/ifor2626-011).
14. Казимиров Н.И., Волков А.Д., Зябченко С.С., Иванчиков А.А., Морозова Р.М. *Обмен веществ и энергии в сосновых лесах Европейского Севера*. Ленинград: Наука, 1977. 304 с.
15. Кайбияйнен Л.К. Эколого-физиологические исследования сосны и сосновых древостоев. *Тр. КарНЦ РАН*. 2003. № 5. С. 65–73.
16. Веселков Б.М. Автоматическая регистрация транспирации у древесных растений в естественных условиях с помощью дифференциального психрометра. В: *Биофизические методы исследований в экофизиологии древесных растений*. Наука, Ленингр. отд-ние, 1979. С. 50–67.

17. Веселков Б.М., Тихов П.В. Связь транспорта воды по ксилеме с интенсивностью транспирации у сосны обыкновенной. *Физиология растений*. 1984. Т. 31. С. 1099–1107.
18. Urban J., Ingwers M.W., McGuire M.A., Teskey R.O. Increase in leaf temperature opens stomata and decouples net photosynthesis from stomatal conductance in *Pinus taeda* and *Populus deltoides* x *nigra*. *Journal of Experimental Botany*. 2017. V. 68. No. 7. P. 1757–1767. doi: [10.1093/jxb/erx052](https://doi.org/10.1093/jxb/erx052).
19. Rodríguez-Gamir J., Xue J., Clearwater M.J., Meason D.F., Clinton P.W., Domec J.-C. Aquaporin regulation in roots controls plant hydraulic conductance, stomatal conductance, and leaf water potential in *Pinus radiata* under water stress. *Plant Cell Environ.* 2019. V. 42. P. 717–729.
20. Барри Д.А., Даунтон У.Д.С. Зависимость фотосинтеза от факторов окружающей среды. В: *Фотосинтез*. Том 2. Под ред. Говинджи. М.: Мир, 1987. С. 273–364.
21. *Психрометрические таблицы*: справочник. Ленинград: Гидрометеиздат, 1972. 235 с.
22. Korpilahti E. Photosynthetic production of Scots pine in the natural environment. *Acta Forestalia Fennica*. 1988. No. 202. Article id 7649. doi: [10.14214/aff.7649](https://doi.org/10.14214/aff.7649).

Рукопись поступила в редакцию 21.08.2019, переработанный вариант поступил 23.12.2019.
Дата опубликования 26.12.2019.