

DOI: <https://doi.org/10.37162/2618-9631-2022-3-144-160>

УДК 556.5.06+004.032.26

Моделирование запасов продуктивной влаги в верхних слоях почвы методом нейронных дифференциальных уравнений

В.А. Голов¹, Ф.Л. Быков²

¹*МИРЭА – Российский технологический университет, г. Москва, Россия*

²*Гидрометеорологический научно-исследовательский центр*

Российской Федерации, г. Москва, Россия

golov.v.a@yandex.ru, bphilipp@inbox.ru

Метод нейронных обыкновенных дифференциальных уравнений (ОДУ) основан на решении ОДУ с нейронной сетью в правой части и предназначен для параметризации ОДУ. Исследуется приложение метода нейронных ОДУ к задаче моделирования запаса продуктивной влаги в верхних 0–10 см и 0–20 см слоях почвы. Предлагается модификация метода нейронных ОДУ для задачи с дополнительными внешними метеорологическими (осадки, скорость ветра, температура воздуха и точка росы) и категориальными (выращиваемая культура, типы почвы и преобладающий тип подстилающей поверхности в окрестности) параметрами. Полученные параметризации демонстрируют среднюю абсолютную погрешность прогноза запасов продуктивной влаги на 10 дней, равную 3,20 и 5,53 мм для 0–10 см и 0–20 см слоёв почвы соответственно. Предложенный подход является перспективным для моделирования процессов в почве и принятия управленческих решений в агропромышленном комплексе.

Ключевые слова: запас продуктивной влаги, нейронные дифференциальные уравнения, машинное обучение, нейронные сети

The available water content in the upper layers of soil modelling by neural ordinary differential equations

V.A. Golov¹, Ph.L. Bykov²

¹*MIREA – Russian Technological University, Moscow, Russia*

²*Hydrometeorological Research Center of Russian Federation, Moscow, Russia*

golov.v.a@yandex.ru, bphilipp@inbox.ru

The neural ordinary differential equations (ODE) based on the solving ODE with temporal derivative calculated by neural network and intended for parameterization of the ODE. The available water content in the 0-10 cm and 0-20 cm layers of a soil model using the neural ODE. The modification for neural ODE takes into account additional meteorological (precipitation, wind speed, air temperature and dew point) and categorical (agricultural type, soil type and land cover type) variables problem was suggested. Obtained parameterizations, demonstrate the mean absolute error of 10 days forecast of available water content 3.20 and 5.53 mm for 0-10 cm and 0-20 cm soil layers, respectively. The proposed approach is promising for modeling processes in the soil and making management decisions in the agro-industrial complex.

Keywords: available water content, neural differential equations, machine learning, neural networks

Введение

В почве помимо гравитации действуют силы поверхностного натяжения, сорбционные и осмотические силы, а на процессы в верхних слоях почвы влияет растительность и взаимодействие с атмосферой. Процессы впитывания влаги в почву имеют масштаб капилляров (доли миллиметра), а для описания процессов влагообмена необходимо знать состав почвы, растительности, количество и размер капилляров в почве и другие физические параметры. Для физически корректного описания влагообмена растительного покрова с атмосферой необходимо знать площадь листьев и распределение метеорологических параметров в слое от поверхности Земли до высоты растительного покрова. Помимо количества жидкой воды в почве необходимо знать температуру почвы, количество водяного пара, содержащегося в почвенных капиллярах, и количество льда в почве. На практике для моделирования почвы и растительного покрова используют параметризацию [1, 3, 5, 7, 13]. Параметризация процессов, зависящих от множества внешних факторов, является сложной математической задачей. Недавно появившийся метод нейронных обыкновенных дифференциальных уравнений (ОДУ) [16] позволяет находить параметризацию в виде нейронных сетей.

Запасом продуктивной влаги (ЗПВ) называют ту часть влаги в почве, которую растения потребляют в процессе своей жизнедеятельности. ЗПВ измеряют в миллиметрах и может быть вычислен по формуле

$$W_{\text{пв}} = 10 \frac{\rho}{\rho_w} h \frac{W-K}{100\%}, \quad (1)$$

где 10 – коэффициент для перевода запасов влаги в миллиметры; ρ – плотность почвы, г/см³; $\rho_w = 1$ г/см³ – плотность воды; h – слой почвы, см; W – влажность почвы, % массы абсолютно сухой почвы; K – влажность устойчивого завядания, % массы абсолютно сухой почвы. Отметим, что результаты измерения ЗПВ на метеостанциях неотрицательны, хотя по формуле (1) можно получить и отрицательные значения ЗПВ.

Так как от количества потребляемой влаги зависит урожайность, а от динамики потребления влаги отчасти зависят траты на воду в агропромышленности, было бы полезно уметь прогнозировать данный параметр. От уровня увлажненности зависит эффективность таких сельскохозяйственных мероприятий, как внесение удобрений и обработка посевов от вредителей.

Зная влажность устойчивого завядания и объемную массу почвы, можно по ЗПВ спрогнозировать абсолютную влажность почвы и ее динамику во времени. Абсолютная влажность почвы сильно влияет на испарение влаги с поверхности Земли, а значит, и на выпадение осадков и конвективные явления в атмосфере.

Для моделирования запаса продуктивной влаги будем использовать метод нейронных ОДУ.

1. Нейронные ОДУ

Метод нейронных ОДУ [13] моделирует процесс в виде решения задачи Коши первого порядка с параметрически заданной правой частью:

$$\begin{cases} \frac{d\vec{z}(t)}{dt} = f(\vec{z}, t, \vec{w}), \\ \vec{z}(t_0) = \vec{z}_0, \end{cases} \quad (2)$$

где $\vec{z}(t)$ – вектор моделируемых параметров (в рассматриваемом случае – ЗПВ в 0–10 и 0–20 см слоях почвы); \vec{f} – векторная функция из параметрического семейства кусочно-гладких векторных функций (например, нейронная сеть [4]), аппроксимирующая производную $\vec{z}(t)$ по времени; \vec{w} – параметры параметрического семейства функций (нейронной сети). Далее для сравнения, помимо случая, когда f – нейронная сеть, рассмотрим и случай, когда f принадлежит семейству линейных функций.

При обучении (оптимизации параметров) нейронной сети f будем решать задачу Коши (2) явным методом Рунге-Кутты четвертого порядка точности $rk4$:

$$z_1 = z(t_0) + \int_{t_0}^{t_1} f(z, t, w) dt \approx \check{z}_1 = rk4(z_0, f(z, t, w), t_0, t_1).$$

Использовался метод четвертого порядка $rk4$, поскольку при его применении используются только значения функции f в моменты времени t , $t + \frac{\Delta t}{2}$, $t + \Delta t$, то есть для его применения на данных регулярных измерений внешних параметров (параметров атмосферы) необходима лишь одна интерполяция параметров атмосферы в точку $t + \frac{\Delta t}{2}$ (использовалась линейная интерполяция).

При обучении нейронной сети на конечной обучающей выборке Z объемом N минимизируется эмпирический риск $Q(w, Z)$, который вычисляется как среднее значение функции потерь \mathcal{L} (например, среднеквадратичной ошибки MSE) для прогноза ЗПВ от начального срока t_0 на срок t_1 в силу модели (2):

$$Q(w, Z) = \frac{1}{N} \sum_{i=1}^N \mathcal{L}(rk4(z_0^i, f(\vec{w}), t_0, t_1), z_1^i) \rightarrow \min_{\vec{w}}, \quad (3)$$

где z_0^i , z_1^i – значения ЗПВ, измеренные в моменты времени t_0 и t_1 соответственно.

Часто ОДУ параметризуют, решая уравнение регрессии для левой и правой частей уравнения. В этом случае ошибки аппроксимации могут накапливаться и нельзя гарантировать, что найденные параметры \vec{w} позволят достичь наилучшей точности среди параметрического семейства

моделей. Процесс минимизации функционала (3) гораздо сложнее, однако он минимизирует именно качество прогнозов, полученных с использованием выбранного метода интегрирования.

ЗПВ измеряются на метеостанциях раз в декаду, а метеорологические параметры – раз в $\Delta t = 3$ часа. Таким образом, возможна оценка прогноза на $t_1 - t_0 = 10$ суток, для вычисления которого будем делать $m = (t_1 - t_0)/\Delta t = 80$ шагов метода *rk4*.

Важной проблемой при минимизации функционала $Q(w, Z)$ при использовании, например, линейной функции f , является возможность возникновения экспоненциально растущих решений. Тогда использование функции потерь MSE может привести к слишком медленной сходимости градиентного спуска при минимизации функционала (3) в пространстве параметров w функции f . Поэтому далее отдельное внимание уделим выбору функции потерь \mathcal{L} . Видимо поэтому большинство работ, посвященных нейронным ОДУ, рассматривают либо замкнутые системы с диссипацией энергии, либо гамильтоновы системы. Рассматриваемая система влагообмена в почве незамкнута: существует приток извне в виде осадков.

1.1. Модификация метода

Изложенный выше метод имеет ряд положительных свойств, таких как возможность балансирования скорости и точности вычисления путем выбора метода решения задачи Коши, способность к аппроксимации сложных закономерностей с использованием обучаемых алгоритмов нейронных сетей.

Однако в данном виде метод способен прогнозировать явления, зависящие только от времени и начального значения ЗПВ. В задаче прогноза ЗПВ на целевой параметр влияют также и метеорологические параметры, такие как температура, осадки и скорость ветра, и агрометеорологические параметры, такие как выращиваемые культуры, типы почвы и преобладающие в окрестности типы подстилающей поверхности (пашня, пастбища, кустарник, лес, водные объекты и т. д.).

Для решения этой проблемы модель (1) была модифицирована следующим образом:

$$\begin{cases} \frac{d\vec{z}(t)}{dt} = f(\vec{z}(t), \vec{v}(t), \vec{w}), \\ z(t_0) = z_0, \end{cases}$$

где $\vec{v}(t)$ – вектор всех параметров, влияющих на целевую переменную z . То есть функция f зависит от времени не явно, а через переменные величины $\vec{v}(t)$, описывающие внешние факторы.

При такой постановке конечная обучающая выборка принимает вид

$$Z = (V_i, z_0^i, z_1^i)_{i=1}^N,$$

где V_i – набор векторов $\vec{v}_j(t)$ для каждого из $j = 1, \dots, m$ шагов интегрирования метода Рунге–Кутты *rk4*, то есть $V_i = (\vec{v}_j)_{j=0}^m$.

1.2. Оценка качества аппроксимации

В рамках исследования необходимо оценить качество модели. Обычно для оценки качества в задачах регрессии, которой является задача (3), в качестве оценки используется среднее значение выбранной функция потерь, что, однако, не всегда является достаточным.

В рамках исследования была выбрана следующая метрика относительной среднеквадратичной погрешности

$$RelMSE(\tilde{z}_1, z_1, z_0) \stackrel{def}{=} \frac{\|z_1 - \tilde{z}_1\|_2^2}{\|z_1 - z_0\|_2^2}, \quad (4)$$

где \tilde{z}_1 – прогнозируемое моделью значение целевой переменной; z_1 – истинное значение целевой переменной в момент времени t_1 ; z_0 – истинное значение целевой переменной в момент времени t_0 . Относительная среднеквадратичная погрешность *RelMSE* показывает, какова доля среднеквадратичной ошибки прогноза модели к среднеквадратичной ошибке прогноза начальным значением. При $RelMSE > 1$ модель бесполезна: использование модели в среднем хуже, чем использование последнего измеренного значения.

Обратное к значению метрики (4) число $1/RelMSE$ демонстрирует, во сколько раз модель ошибается меньше, чем если бы предсказывалось такое же значение, что и в начальный момент времени.

Для оценки качества прогноза низких ЗПВ (менее 5 мм в слое 0–10 см и менее 10 мм в слое 0–20 см), которые могут привести к засухе, критический индекс успеха *CSI* вычисляется по формуле:

$$CSI = \frac{TP}{TP + FN + FP},$$

где *TP* – количество угаданных случаев низких ЗПВ; *FN* – количество ложных тревог; *FP* – количество пропусков цели.

2. Численные эксперименты

2.1. Данные обучающей выборки

Используемая выборка включала данные 505 агрометеостанций России и Казахстана, ведущих агрометеорологические наблюдения, за период с 2007 по 2017 г., для которых доступны также данные наблюдений за приземными метеорологическими параметрами (рис. 1).

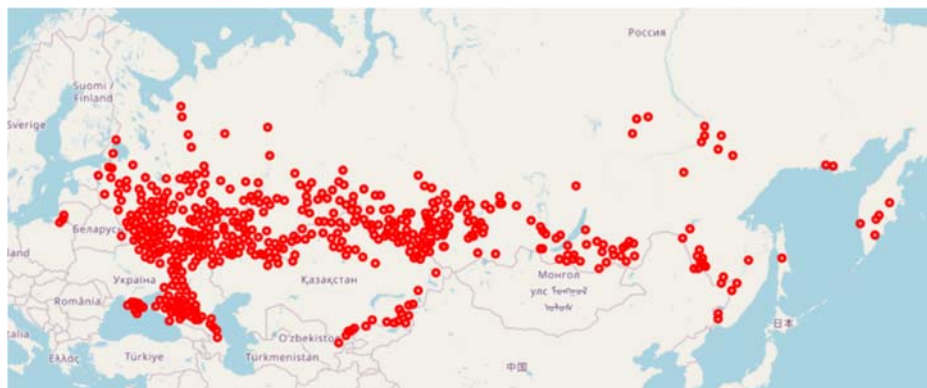


Рис 1. Карта расположения 505 используемых метеостанций, ведущих наблюдения за ЗПВ.

Fig 1. Map of the locations of 505 used weather stations, conducting observations of available water content.

В модели запасов продуктивной влаги в почве использовались следующие признаки \vec{v}

- Запас продуктивной влаги в слоях толщиной 10 и 20 см
- Температура воздуха на уровне 2 м
- Точка росы на уровне 2 м
- Скорость ветра на уровне 10 м
- Осадки, выпавшие за последние 12 часов
- Параметр сезонности (гармоническая функция от времени с периодом 1 год)
 - Среднемесячная температура поверхности
 - Среднемесячная влажность почвы
 - Среднемесячные осадки
 - Типы почвы
 - Типы подстилающих поверхностей
 - Типы выращиваемых культур

Сумма осадков за 12 часов разбивалась на четыре равные трехчасовые суммы. Запас продуктивной влаги измеряется раз в декаду (раз в 10 дней). На практике же (из-за месяцев длиной в 31 день) – раз в 10 или 11 дней. С целью распараллеливания вычислений желательно иметь равные временные отрезки, поэтому было принято, что ЗПВ на 10 день такой же, как и на 11 день с момента предыдущей записи. Время отбора проб почвы для замера ЗПВ неизвестно, поэтому будем считать, что все отборы проб происходят в 06 ч ВСВ. Это допущение не далеко от истины и с учетом периодичности измерений ЗПВ не должно оказывать заметного влияния.

Типы почв и типы подстилающей поверхности выбрали согласно картам Продовольственной и сельскохозяйственной организации ООН и Европейского Космического Агентства соответственно, доступным

в формате geotiff на сайтах [10] и [11] соответственно. Среднемесячные климатические характеристики получены из данных NOAA [15] в формате NetCDF, данные представлены в виде карт с шагом сетки $0,5^\circ$. При интерполяции карт и среднемесячных характеристик в точки станций использовался метод ближайшего соседа.

Типы почв, типы подстилающей поверхности и типы культур были преобразованы. Представленные в малом количестве типы были объединены в тип «Другие», а типы выращиваемых культур переразмечены в группы «Овощи», «Злаки озимые», «Злаки яровые», «Бобовые», «Травы», «Фрукты», «Другие».

Типы являются категориальными данными, поэтому для подачи на вход нейронной сети их необходимо преобразовать в вектора действительных чисел с помощью таблиц, причем числа в таблицах будем оптимизировать в процессе обучения модели. Такое преобразование называют вложением (Embedding) [8]. Размерности вложений p для каждой из рассматриваемых категориальных характеристик выбирались по правилу

$$p = \lceil k / 2 \rceil,$$

где k – количество различных категорий. Количество категорий k и используемые значения размерностей вложений p представлены в табл. 1.

Таблица 1. Размерности вложений для категориальных признаков
Table 1. Categorical features embeddings's dimentions

Признак	Количество типов	Размерность вложения
Тип почвы	7	3
Тип подстилающей поверхности	7	3
Тип выращиваемой культуры	8	4

Метеорологические характеристики записываются с периодичностью в $\Delta t = 3$ часа, поэтому на одну пару значений запасов продуктивной влаги (z_0, z_1) приходится 80 векторов значений метеорологических характеристик. В данной работе в качестве значений векторов \vec{v} использовались данные измерений на метеостанциях. В перспективе рассматриваемая модель может использовать данные кратко- и среднесрочного численного прогноза погоды. Ожидаемо (и это показано ниже), результат моделирования сильно зависит прежде всего от количества выпадающих осадков. Поэтому при использовании численного прогноза качество прогнозов будет напрямую зависеть от качества используемого прогноза осадков и стоит ожидать высокого качества только у краткосрочных прогнозов ЗПВ.

2.2. Параметры обучения

В ходе экспериментов все параметры, кроме функции потерь, семейства алгоритмов и функции активации, были фиксированными. Это было необходимо, чтобы оценить вклад не зафиксированных параметров в качество модели.

Для обучения и тестирования модели выборка была разбита на обучающую (2007–2014 гг.) и тестовую (2015–2017 гг.). Разбиение осуществлялось по времени. Такое разбиение, в отличие разбиения по пространству, позволяет выявить перепопдгонку модели f . При разбиении по пространству в обе выборки могут попасть близкие станции со схожим процессом.

Объем N обучающей выборки был равен 39000 декад. На каждом шаге оптимизации параметров w из общей выборки случайным образом выбирается подвыборка (так называемый батч) объемом 500 декад. Для рассматриваемого батча с использованием алгоритма градиентного спуска AMSGrad [17] в пространстве параметров w с убывающим размером шага $\eta(i) < \eta_0 = 0.005$ оптимизируются параметры w нейронной сети f . Один проход по всему архиву называют эпохой обучения. Каждая эпоха состояла из 78 батчей, а количество эпох было равно 100.

Численные эксперименты проводились с использованием пакета PyTorch версии 1.9.1. На системе с GPU NVidia Tesla V100 (модули суперкомпьютера V6000 ГВЦ Росгидромета) обучение каждой из рассмотренных моделей заняло около 1 ч 15 мин. По нашим предварительным оценкам, поскольку при применении предлагаемой модели в оперативном режиме пропадет необходимость хранить в памяти результаты промежуточных вычислений, размер батча может быть увеличен как минимум в 300 раз, что позволит применять модель на родных сетках региональных численных моделей прогноза погоды.

Для уменьшения шага градиентного спуска $\eta(i)$ использовался планировщик, модифицирующий скорость обучения η в зависимости от номера эпохи обучения i по правилу

$$\eta(i) = 0.01^{i/150} \eta_0,$$

где i – номер эпохи обучения.

Во всех экспериментах все характеристики, подающиеся на вход нейронной сети (за исключением категориальных), нормируются. Нормировки вычисляются по обучающей выборке

$$\bar{v}_i = \frac{x - \mathbb{E}v_i}{\sigma(v_i)},$$

где $\mathbb{E}v_i$ – среднее значение признака v_i по архиву; $\sigma(v_i)$ – среднеквадратическое отклонение признака v_i .

Если в параметрическом семействе функций f содержатся неограниченные функции, то некоторые из решений могут экспоненциально расти.

Тогда при использовании классической среднеквадратичной функции потерь $\mathcal{L} = MSE$ градиенты ошибок могут оказаться большими и со сходимостью к минимуму функционала (3) могут возникнуть сложности. Рассмотрим следующие функции потерь с ограниченным градиентом:

$$RMSE = \|\tilde{z}_1 - z_1\|_2; MAE = \|\tilde{z}_1 - z_1\|_1;$$

$$SmoothMAE = \begin{cases} \frac{1}{4} \|\tilde{z}_1 - z_1\|_2^2, & \|\tilde{z}_1 - z_1\|_1 < 2, \\ \|\tilde{z}_1 - z_1\|_1 - 1, & \|\tilde{z}_1 - z_1\|_1 \geq 2. \end{cases}$$

Другим недостатком функции MSE является плохое качество прогнозов редких малых значений. Малые значения ЗПВ представляют особый интерес для прогнозирования, поскольку приводят к засухе, поэтому рассмотрим дополнительно среднюю относительную ошибку WAPE и симметричную относительную ошибку SMAPE:

$$WAPE = \frac{\|\tilde{z}_1 - z_1\|_1}{\|z_1\|_1}; SMAPE = \frac{2\|\tilde{z}_1 - z_1\|_1}{\|z_1\|_1 + \|\tilde{z}_1\|_1}.$$

Для нейронных сетей с функцией активации \tanh использовалось правило инициализации весов Xavier [12], то есть коэффициенты выбирались из равномерного распределения на отрезке

$$\left[-\sqrt{\frac{6}{n_{in} + n_{out}}}, \sqrt{\frac{6}{n_{in} + n_{out}}} \right],$$

где n_{in} – размерность вектора, подаваемого на вход линейному слою; n_{out} – размерность вектора на выходе линейного слоя.

Для сетей с функцией активации ReLU использовалось правило Kaiming He [14], то есть коэффициенты выбирались из нормального распределения с дисперсией $\frac{2}{n_{in}}$.

2.3. Линейная аппроксимация производной

В качестве семейства функций, аппроксимирующих производную, рассмотрим семейство линейных функций

$$f(\vec{z}, \vec{v}, \{w_z, w_v, b_1\}) = w_z \vec{z} + w_v \vec{v} + \vec{b}_1,$$

где w_z и w_v – матрицы размеров 2×2 и 2×18 соответственно; $\vec{b}_1 \in \mathbb{R}^2$ – вектор. Вектор оптимизируемых параметров w модели f состоит из 42 коэффициентов матриц w_z, w_v и вектора \vec{b}_1 .

В соответствии с результатами численных экспериментов, модель f , обученная на функции потерь RMSE, демонстрирует наилучшее значение относительной среднеквадратичной погрешности RelMSE. Линейная

модель при заданных параметрах обучения менее чем за 20 эпох асимптотически сходится к определенному значению параметров w . Использование линейной модели f по сравнению с использованием последнего известного значения уменьшает MSE в

$$1.49 < \frac{1}{RelMSE} < 1.57$$

раз для слоя 0–10 см и в

$$1.38 < \frac{1}{RelMSE} < 1.57$$

раз для слоя 0–20 см соответственно.

Из табл. 2 и 3 видно, что среди линейных моделей по относительной среднеквадратичной погрешности RelMSE лучшей является модель, обученная на функции потерь RMSE, однако по среднему абсолютному отклонению MAE лучшей является модель, обученная на функции потерь SmoothMAE.

Таблица 2. Оценки для слоя 0-10 см на тестовой выборке 2015–2017 гг. различных моделей, обученных на разных функциях потерь
Table 2. Estimations for 0-10 cm layer on the test 2015-2017 years dataset for different models trained on various loss functions

Функция потерь	Оценка			
	RelMSE	MAE	WAPE	CSI
Линейная модель				
MAE	0.6561	3.4834	0.2362	0.5396
MSE	0.6384	3.4968	0.2376	0.5152
RMSE	0.6494	3.5411	0.2401	0.4982
SmoothMAE	0.6555	3.4831	0.2362	0.5391
WAPE	0.6562	3.4835	0.2362	0.5399
SMAPE	0.6687	3.5282	0.2392	0.5367
Нелинейная модель с $g = \tanh$				
MAE	0.5597	3.2291	0.2190	0.5779
MSE	0.5623	3.2722	0.2219	0.5474
RMSE	0.5605	3.2646	0.2214	0.5393
SmoothMAE	0.5586	3.2259	0.2187	0.5793
WAPE	0.5600	3.2298	0.2190	0.5785
SMAPE	0.5567	3.2429	0.2199	0.5625
Нелинейная модель с $g = ReLU$				
MAE	0.5554	3.2006	0.2170	0.5702
MSE	0.5550	3.2462	0.2201	0.5503
RMSE	0.5551	3.2507	0.2204	0.5463
SmoothMAE	0.5533	3.2007	0.2170	0.5725
WAPE	0.5565	3.2045	0.2173	0.5649
SMAPE	0.5575	3.2334	0.2192	0.5680

Таблица 3. Оценки для слоя 0-20 см на тестовой выборке 2015–2017 гг. различных моделей, обученных на разных функциях потерь

Table 3. Estimations for 0-20 cm layer on the test 2015-2017 years dataset for different models trained on various loss functions

Функция потерь	Оценка			
	RelMSE	MAE	WAPE	CSI
Линейная модель				
MAE	0,7203	6.3231	0.2127	0.5457
MSE	0.6384	5.9804	0.2018	0.5265
RMSE	0.6687	6.1730	0.2076	0.5227
SmoothMAE	0.7203	6.3247	0.2127	0.5457
WAPE	0.7206	6.3245	0.2127	0.5454
SMAPE	0.6658	5.9975	0.2017	0.5606
Нелинейная модель с $g = \tanh$				
MAE	0.5716	5.5598	0.1870	0.5838
MSE	0.5603	5.5637	0.1871	0.5647
RMSE	0.5612	5.5776	0.1876	0.5626
SmoothMAE	0.5716	5.5564	0.1869	0.5826
WAPE	0.5720	5.5627	0.1871	0.5844
SMAPE	0.5848	5.6390	0.1897	0.5799
Нелинейная модель с $g = ReLU$				
MAE	0.5701	5.5265	0.1859	0.5927
MSE	0.5562	5.5535	0.1868	0.5726
RMSE	0.5579	5.5579	0.1869	0.5701
SmoothMAE	0.5687	5.5323	0.1861	0.5867
WAPE	0.5700	5.5336	0.1861	0.5895
SMAPE	0.5748	5.5590	0.1870	0.5854

2.4. Нелинейная аппроксимация производной

Теперь в качестве семейства функций, аппроксимирующих производную, рассмотрим множество нейронных сетей типа двухслойный перцептрон [4] с $k = 64$ нейронами

$$f(\vec{z}, \vec{v}, \{w_z, w_v, \vec{b}_1, w_2, \vec{b}_2\}) = w_2 g(w_z \vec{z} + w_v \vec{v} + \vec{b}_1) + \vec{b}_2,$$

где $g(\cdot)$ – функция активации (которая применяется поэлементно); w_z, w_v и w_2 – действительные матрицы размеров $k \times 2$, $k \times 18$ и $2 \times k$ соответственно; $b_1 \in \mathbb{R}^k$, $b_2 \in \mathbb{R}^2$. Вектор оптимизируемых параметров w модели f состоит из $23k + 2 = 1474$ коэффициентов матриц w_z, w_v, w_2 и векторов b_1, b_2 . Эксперименты проводились с использованием функций активации $g(x) = ReLU(x) = \max(x, 0)$ и $g(x) = \tanh(x)$.

Согласно теореме Цыбенко [9], любая непрерывная функция на n -мерном кубе может быть приближена двухслойным перцептроном с функцией активации \tanh с любой наперед заданной точностью в метрике C . Кроме того, при использовании \tanh функция f ограничена сверху

значением $\|w_2\| + |\vec{b}_2|$, а значит, ограничены и прогнозы модели, и градиенты ошибок.

Функция активации ReLU рассмотрена, поскольку она, во-первых, быстрее вычисляется, а во-вторых, во многих практических задачах достигает более высоких оценок по сравнению с \tanh .

Из табл. 2 и 3 можно заключить, что нелинейная модель f демонстрирует лучшее качество по сравнению с линейной моделью. Согласно численным экспериментам, модели с одним скрытым слоем еще имеют потенциал к обучению, однако на тестовой выборке оценки сходятся. Дополнительное обучение может перерасти в перепогонку модели. Использование нелинейных моделей f по сравнению с использованием последнего известного значения уменьшает MSE

$$1.78 < \frac{1}{RelMSE} < 1.81$$

раз для слоя 0–10 см и

$$1.71 < \frac{1}{RelMSE} < 1.80.$$

раз для слоя 0–20 см.

Более точные оценки приведены в табл. 2 и 3. Из таблиц видно, что хотя модель, обученная на функции потерь MSE, демонстрирует наилучший результат по относительной среднеквадратичной погрешности RelMSE, модель, обученная на функции SmoothMAE, является лучшей по среднему абсолютному отклонению MAE.

3. Анализ и сравнение результатов

Сравним значения оценок для линейной и нелинейной аппроксимаций производной. Для этого построим график, на котором точками отметим позиции линейных и нелинейных моделей. По осям отложены оценки RelMSE и CSI для прогноза низких ЗПВ в слоях 0–10 см и 0–20 см.

На рис. 2 видно, что в зависимости от архитектуры и используемой при обучении функции потерь модели разбиваются на группы с близкими значениями полученных на тестовой выборке оценок. По метрике RelMSE наилучшие оценки демонстрирует модель с функцией активации ReLU, а наихудшие – линейная модель.

Заметим, что выбор функции активации практически не влияет на значение метрики RelMSE, а для слоя 0–10 см на это не сильно влияет и выбор функции потерь. Выбранная функция потерь сильнее сказывается на значениях метрики RelMSE в слое 0–20 см. Модели, демонстрирующие наименьшее значение метрики RelMSE, соответствуют функциям потерь MSE и RMSE. Модели, демонстрирующие лучшие значения индекса CSI, соответствуют остальным рассмотренным функциям потерь. Последняя закономерность о значениях индекса CSI верна и для слоя 0–10 см.

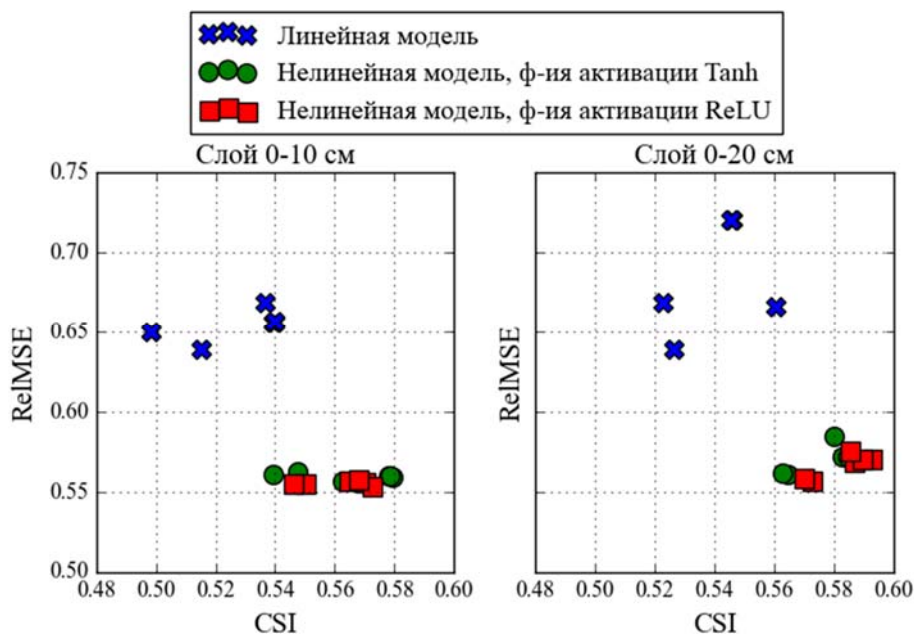


Рис 2. Значения оценок CSI и ReIMSE для нелинейных и линейных моделей на тестовой выборке 2015–2017 гг.: для слоя 0-10 см (а); для слоя 0-20 см (б).
Fig 2. Estimations CSI and ReIMSE for linear and non-linear models on the test 2015-2017 years dataset: for layer 0-10 cm (a); for layer 0-20 cm (б).

На рис. 3 показана месячная динамика ЗПВ согласно линейной и нелинейной модели с функцией активации ReLU. Обе модели обучены на функции потерь RMSE. Видно, что нелинейная модель точнее воспроизводит динамику ЗПВ, причем для нелинейной модели рост ЗПВ может заметно отличаться при том же количестве выпавших осадков. Отметим, что качество прогнозов падает с ростом заблаговременности, что говорит о необходимости использования системы усвоения данных о ЗПВ в почве [2].

Выводы

По результатам экспериментов можно сделать вывод, что модели, основанные на методе нейронных дифференциальных уравнений, являются перспективными для моделирования таких сложных незамкнутых систем, как влагообмен в почве. Хотя рассматривались нейронные ОДУ общего вида, то есть никакая эмпирическая информация о физическом процессе не использовалась, наилучшей оказалась параметризация, в которой рост ЗПВ в почве происходит только в периоды выпадения осадков (рис. 3).

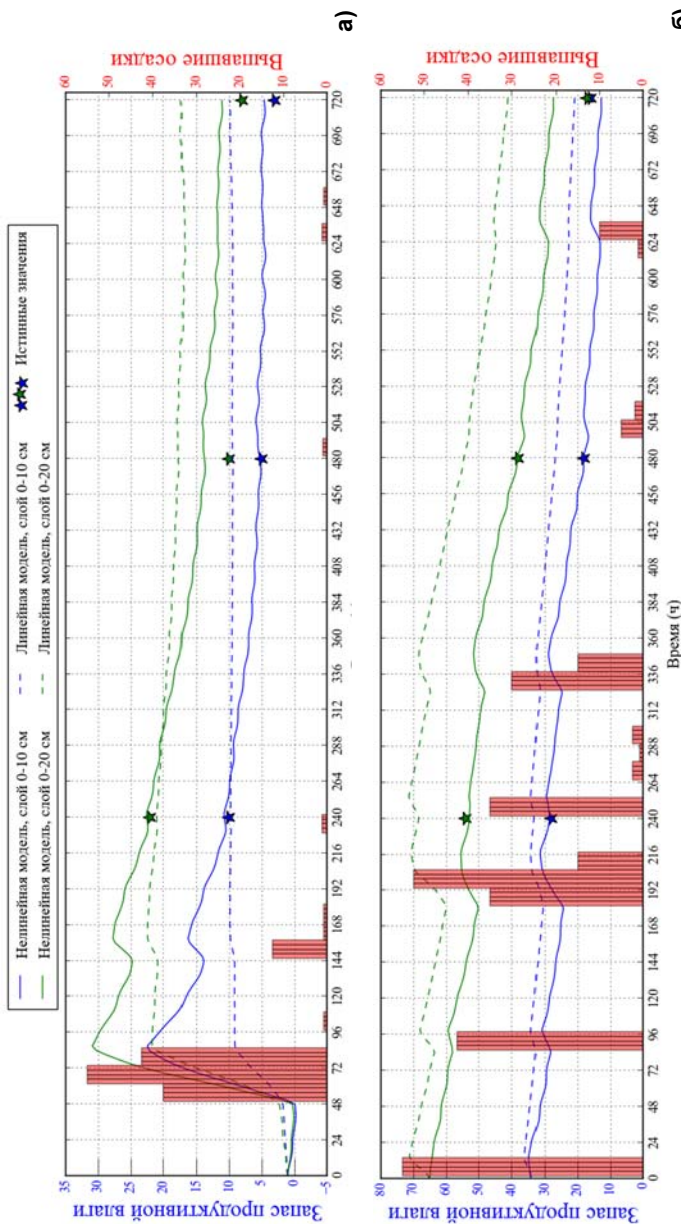


Рис. 3. Примеры динамики ЗГВ (линии, левая шкала, мм) следующие на 30 суток согласно линейной модели (пунктир) и не-линейной модели с функцией активации ReLU, обученных на функции потерь RMSE: для станции 27393 Нолинск (Кировская область) от начального срока 7 июля 2016 г. (а); для станции 27877 Городище (Пензенская область) от начального срока 27 июня 2017 г. (б). Показано количество выпавших за 12 ч осадков (правая шкала, мм).
Fig 3. Samples of available water content (lines, left y-axis, mm) 30 days modelling with linear model (dashed) and non-linear model with ReLU activation function learned on RMSE loss function: a) for station 27393 Nolinisk (Kirov region) started on 7 July 2016 (a); for station 27877 Gorodische (Penza region) started on 27 June 2017 (b). The 12 hours precipitation's sums are shown (right scale, mm).

Рассмотренные нелинейные модели демонстрируют результаты существенно лучшие, чем линейные. Полученные нелинейные параметризации демонстрируют (табл. 2 и 3) среднюю абсолютную погрешность MAE прогноза запасов продуктивной влаги на 10 дней, равную 3.20 и 5.53 мм в 0–10 см и 0–20 см слоях почвы соответственно. Согласно Руководящему документу [4], погрешность измерений ЗПВ составляет 10 %, но не более 5 мм. Полученные оценки средней относительной погрешности WAPE (табл. 2 и 3) дают оценку ошибки около 22.0 % и 18.6 % для слоев 0–10 см и 0–20 см соответственно.

Продемонстрировано (рис. 2), что наибольшие значения критического индекса успешности (CSI = 0.58–0.59) прогноза низких ЗПВ, приводящих к засухе; наименьшая среднеквадратичная ошибка RMSE прогноза ЗПВ достигаются на различных параметризациях модели.

Перед практическим применением предложенных математических моделей для прогнозирования процессов в почве необходимо провести дополнительные исследования. Прежде всего необходимо протестировать качество применения предложенных моделей на прогностических данных об атмосфере, а не на данных измерений. В случае удачного тестирования на прогностических данных предложенные модели могут быть использованы для принятия более обоснованных решений в агропромышленном комплексе.

Для использования данной модели совместно с атмосферными и гидрологическими моделями необходимо выполнить дополнительный комплекс работ, включающий в себя: 1) моделирование ЗПВ при увеличении числа рассматриваемых слоев почвы; 2) учет дополнительных переменных (величина испарения, речной сток).

Список литературы

1. Будаговский А.И. Впитывание воды в почву. 1955. 136 с.
2. Быков Ф.Л., Василенко Е.В., Гордин В.А., Тарасова Л.Л. Статистическая структура поля влажности верхнего слоя почвы по данным наземных и спутниковых наблюдений // Метеорология и гидрология. 2017. № 6. С. 68-84.
3. Роде А. А. Избранные труды. Т.3. Основы учения о почвенной влаге // Под ред. Г.В. Добровольского. М.: Почвенный ин-т им. В. В. Докучаева Россельхозакадемии, 2008. 664 с.
4. РД 52.33.217–99. Наставления гидрометеорологическим станциям и постам. Вып. 11. Агрометеорологические наблюдения на станциях и постах. Часть 1 Основные агрометеорологические наблюдения. Кн. 1. СПб.: Гидрометеоиздат, 2000. 348 с.
5. Сиротенко О.Д. Математическое моделирование воднотеплового режима и продуктивности агроэкосистем. Л.: Гидрометеоиздат, 1981. 167 с.
6. Хайкин С. Нейронные сети. Полный курс: 2-е изд. М.: Изд. дом «Вильямс», 2006. 1104 с.
7. Baver L.D. Soil Physics. New York: John Wiley Sons Inc., 1948. 398 p.
8. Guo C., Berkahn F. Entity Embeddings of Categorical Variables. 2016, arXiv. <https://arxiv.org/abs/1604.06737>.
9. Cybenko G. Approximation by superpositions of a sigmoidal function // Mathematics of Control, Signals and Systems. 1989. Vol. 2, no. 4. P. 303-314.

10. European Space Agency. Global Land Cover Map, 2010. http://due.esrin.esa.int/page_globcover.php.
11. FAO-UNESCO. Soil Map of the World, digitized by ESRI. Soil climate map. USDA-NRCS, Soil Science Division, World Soil Resources, Washington D.C., 2005, https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/use/?cid=nrcs142p2_054013.
12. Xavier G., Bengio Y. Understanding the difficulty of training deep feedforward neural networks // Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, PMLR. 2010. Vol. 9. P. 249-256.
13. Hillel D. Environmental Soil Physics: Fundamentals, Applications and Environmental Considerations. London: Academic Press, Oval Road, UK, 1998. 775 p.
14. He K., Zhang X., Ren S., Sun J. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. 2015. <https://arxiv.org/abs/1502.01852>.
15. National Oceanic and Atmospheric Administration (NOAA). CPC Soil Moisture data provided by the NOAA/OAR/ESRL PSL. <https://psl.noaa.gov/data/gridded/data.cpcsoil.html>.
16. Ricky T., Chen Q., Rubanova Yu., Bettencourt J., Duvenaud D. Neural Ordinary Differential Equations. 2018. <https://arxiv.org/abs/1806.07366>.
17. Reddi S.J., Kale S., Kumar S. On the Convergence of Adam and Beyond. 2019, <https://arxiv.org/abs/1904.09237>.

References

1. Budagovsky A.I. Vpityvanie vody v pochvu [Water absorption into the soil], 1955, 136 p. [in Russ.].
2. Bykov Ph.L., Vasilenko E.V. Gordin V.A. Tarasova L.L. The Statistical Structure of the Field of Surface Soil Layer Moisture from Ground-based and Satellite Observations. *Russ. Meteorol. Hydrol.*, 2017, vol. 42, no. 6, pp. 403-414. DOI: 10.3103/S1068373917060061.
3. Rode A. A. Izbrannyye trudy. T.3. Osnovy ucheniya o pochvennoy vlage [Selected Research Works, Vol. 3: Basic Theory on Soil Moisture]. Dokuchaev Soil Science Inst., Moscow, 2008, 664 p. [in Russ.].
4. RD 52.33.217–99. Nastavleniya gidrometeorologicheskim stancyam i postam. Vyp. 11. Agrometeorologicheskie nablyudeniya na stancyah i postah. Chast' 1 Osnovnyye agrometeorologicheskie nablyudeniya [Guidance document RD 52.33.217–99 Instructions to hydrometeorological stations and posts. Issue 11 Agrometeorological observations at stations and posts. Part 1 Basic agrometeorological observations], 2000, Saint Petersburg, Gidrometeoizdat publ., 348 p. [in Russ.].
5. Sirotenko O. D. Matematicheskoe modelirovanie vodnoteplovogo rezhima i produktivnosti agroekosistem [Mathematical modeling of the water-thermal regime and productivity of agroecosystems] Leningrad, Gidrometeoizdat publ., 1981, 167 p. [in Russ.].
6. Haykin S. Neural Networks: A Comprehensive Foundation, 2nd Edition. 1999, McMaster University, Ontario Canada, 823 p.
7. Baver L.D. Soil Physics. New York: John Wiley Sons Inc., 1948, 398 p.
8. Guo C., Berkhahn F. Entity Embeddings of Categorical Variables. 2016. arXiv, available at: <https://arxiv.org/abs/1604.06737>.
9. Cybenko G. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*. 1989, vol. 2, no. 4, pp. 303-314.
10. European Space Agency. Global Land Cover Map, 2010, available at: http://due.esrin.esa.int/page_globcover.php.
11. FAO-UNESCO. Soil Map of the World, digitized by ESRI. Soil climate map. USDA-NRCS, Soil Science Division, World Soil Resources, Washington D.C., 2005, available at: https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/use/?cid=nrcs142p2_054013.
12. Xavier G., Bengio Y. Understanding the difficulty of training deep feedforward neural networks. Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, PMLR, 2010, vol. 9, pp. 249-256.
13. Hillel D. Environmental Soil Physics: Fundamentals, Applications and Environmental Considerations. London: Academic Press, Oval Road, UK, 1998, 775 p.

14. He K., Zhang X., Ren S., Sun J. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. 2015. Available at: <https://arxiv.org/abs/1502.01852>.

15. National Oceanic and Atmospheric Administration (NOAA). CPC Soil Moisture data provided by the NOAA/OAR/ESRL PSL. Available at: <https://psl.noaa.gov/data/gridded/data.cpesoil.html>.

16. Ricky T., Chen Q., Rubanova Yu., Bettencourt J., Duvenaud D. Neural Ordinary Differential Equations, 2018. Available at: <https://arxiv.org/abs/1806.07366>.

17. Reddi S.J., Kale S., Kumar S. On the Convergence of Adam and Beyond. 2019, available at: <https://arxiv.org/abs/1904.09237>.

*Поступила 31.05.2022; одобрена после рецензирования 07.09.2022;
принята в печать 23.10.2022.*

*Submitted 31.05.2022; approved after reviewing 07.09.2022;
accepted for publication 23.10.2022.*