

## ГЕНЕЗИС И ГЕОГРАФИЯ ПОЧВ

УДК 631.4

### ФАКТОРЫ СРЕДЫ В ПРЕДСКАЗАТЕЛЬНОМ КАРТОГРАФИРОВАНИИ ПОЧВ (ОБЗОР)

© 2023 г. П. А. Шарый\*

*Институт физико-химических и биологических проблем почвоведения РАН,  
ул. Институтская, 2, Пущино, Московская область, 142292 Россия*

\*e-mail: p\_shary@mail.ru

Поступила в редакцию 15.03.2022 г.

После доработки 08.11.2022 г.

Принята к публикации 09.11.2022 г.

В известной концептуальной модели SCORPAN анализируемое свойство почв рассматривается как зависящее от следующих факторов среды: других свойств почв, климата, организмов, рельефа, геологии, времени и пространства. Предсказательное картографирование почв в цифровой почвенной картографии основано на близких идеях, но факторы среды могут включать в себя не только факторы почвообразования, но и сведения, полученные при дистанционном зондировании. Его методы получили значительное распространение не только в почвоведении, но и в экологии, сельском хозяйстве и геоморфологии. В настоящей статье дается обзор факторов среды, используемых в предсказательном картографировании с особым вниманием к ситуациям, когда может применяться широкий набор факторов среды и/или когда часть их является неколичественными, как в случае типов растительности. Наиболее разработаны системы количественных показателей для описания рельефа и климата, поэтому им уделено особое внимание. Описание рельефа осуществляется как локальными, так и нелокальными, требующими интегрирования, величинами. При описании климата существенны величины, оценивающие сухость/влажность местности, такие как коэффициент увлажнения и дефицит воды. Они требуют оценки испаряемости, которая не измеряется метеостанциями, но может быть получена путем расчета. Описаны возможности учета этих и других факторов среды, включая неколичественные, в количественных статистических моделях предсказательного картографирования почв, принципы построения моделей, их проверки, сравнения, выбора подходящих моделей. Приведены примеры применения цифровой почвенной картографии к участкам различного масштаба, отмечена специфика моделей для разных масштабов. Обсуждаются некоторые аспекты использования данных дистанционного зондирования.

*Ключевые слова:* модель SCORPAN, цифровая почвенная картография, геоморфометрия, климат, пространственные связи

DOI: 10.31857/S0032180X22100380, EDN: HСYРW1

#### ВВЕДЕНИЕ

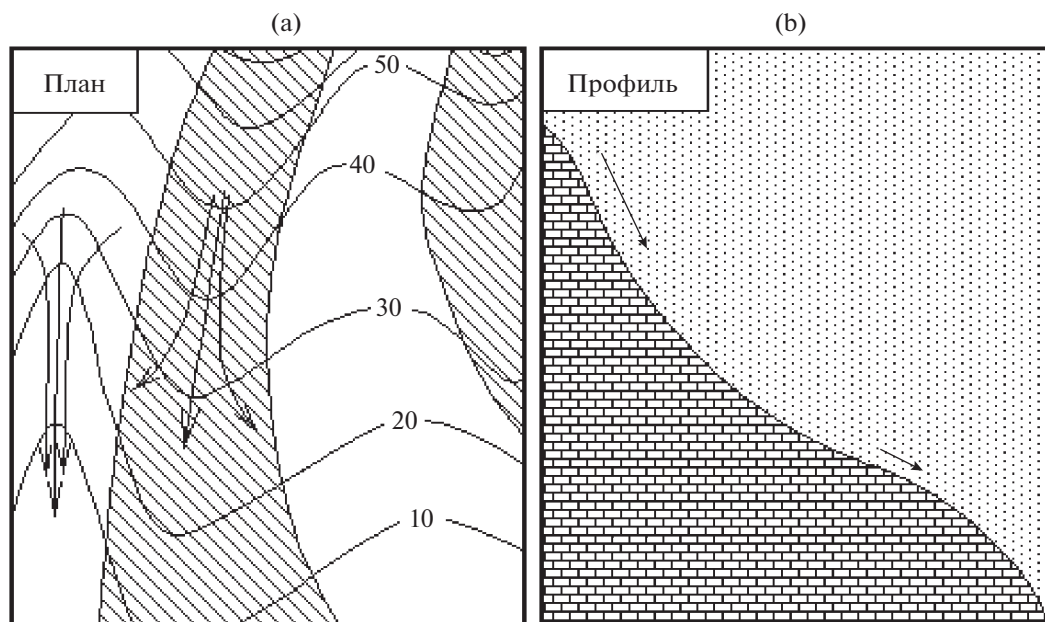
В общем концептуальном плане почву определяют пять факторов почвообразования Докучаева [5]: климат, рельеф, организмы, геология и время. Почва здесь понимается как вся совокупность ее свойств, а если рассматривать одно (изучаемое) свойство почв, то оно зависит также и от других свойств почв (например, содержание органического углерода может зависеть от гранулометрического состава почв), поэтому концептуальная модель SCORPAN [49] зависимости данного свойства почв  $S$  от факторов среды имеет вид

$$S = f(s, c, o, r, p, a, n), \quad (1)$$

где  $S$  – анализируемое свойство почв (или типы почв),  $s$  представляет другие свойства почв,  $c$  – климат,  $o$  – организмы (включая человека),  $r$  –

рельеф,  $p$  – почвообразующие породы,  $a$  – возраст почв, а  $n$  отвечает влиянию на  $S$  пространственных связей между значениями  $S$ . Например, в пространственно близких точках значения  $S$  могут быть ближе друг к другу, а в отдаленных – менее близки (иногда – наоборот), что и отражает показатель  $n$ .

Модель SCORPAN (уравнение (1)) является лишь концептуальной, поскольку функция  $f(\dots)$  в ней неизвестна. К этому подходу близок другой – предсказательное картографирование (или моделирование) – развивавшийся в почвоведении (обзор [56]), экологии (обзоры [31, 35]), геоморфологии (обзор [28]) и сельском хозяйстве [15]. Суть этого подхода в том, что анализируемое свойство почв в точках наблюдения статистически сравнивается с факторами среды, такими как



**Рис. 1.** Механизмы аккумуляции: а – первый, б – второй. Для а – кривые с цифрами – горизонтали (изолинии высоты), кривые со стрелками – линии тока. Области конвергенции (сближения потоков) показаны белым, области дивергенции заштрихованы. Для б – показан разрез профиля склона. Грунт заштрихован. В верхней части вогнутого профиля частицы движутся быстрее из-за большей крутизны, показанной длиной стрелки, в нижней – медленнее, “наползая” на лежащие ниже по склону частицы и формируя этим второй механизм.

климат, рельеф и данные дистанционного зондирования, что дает уравнение (например, множественной регрессии) или модель, которые затем можно применить для интерполяции и экстраполяции данного свойства почв на изучаемую местность с целью получения карты этого свойства. Кроме того, этот подход позволяет осуществить проверку модели, а также оценить тесноту связи и значимость (относительную роль или влияние) каждого фактора среды для  $S$ .

Настоящий обзор посвящен в первую очередь оценке доступных факторов среды для почв, но в качестве примеров рассматриваются и результаты в смежных дисциплинах: экологии, сельском хозяйстве и геоморфологии.

### ФАКТОРЫ СРЕДЫ

Концептуальная модель SCORPAN основана на использовании, в сущности, только факторов почвообразования и оставляет неизвестной функцию  $f(s, c, o, r, p, a, n)$ . В то же время изучаемое свойство почв  $S$  может быть тесно связано с более легкодоступными измерению показателями, такими как данные дистанционного зондирования [65], расстояния от рек и других объектов [28] или электропроводность почв [48]. Поэтому в предсказательном картографировании, вообще говоря, используются не обязательно факторы почвообразования, но в общем случае и косвенно

связанные с ними величины; те и другие называют факторами среды (environmental variables) или предикторами [48, 49]. Для разных факторов почвообразования наборы таких предикторов введены с разной степенью подробности. Отмечено, что доступность полезных предикторов важнее выбора той или иной статистической модели [50]. Наиболее представительные наборы факторов среды введены для описания рельефа и климата, поэтому рассмотрим характеристики последних подробнее.

**Описание рельефа.** Цифровая модель рельефа (ЦМР) обычно используется как базовая для получения предсказательной карты изучаемого свойства почв [56] или растительности [31, 35]. Поясним основные черты информации, получаемой из ЦМР.

Потоки веществ, движущихся под действием силы тяжести по земной поверхности, могут аккумулироваться в определенных местах потому что предпосылки для этого им создает рельеф. В геоморфометрии, науке о количественном анализе рельефа [30, 54, 58], интерес состоит, прежде всего, в количественном описании рельефа. Например, линии тока, по которым обычно течет вода, могут сближаться в одних участках местности и расходиться в других (рис. 1а).

Эти два механизма аккумуляции, очерченные качественно на рис. 1, можно описать количественно, как показано в работах [57, 58], знаками

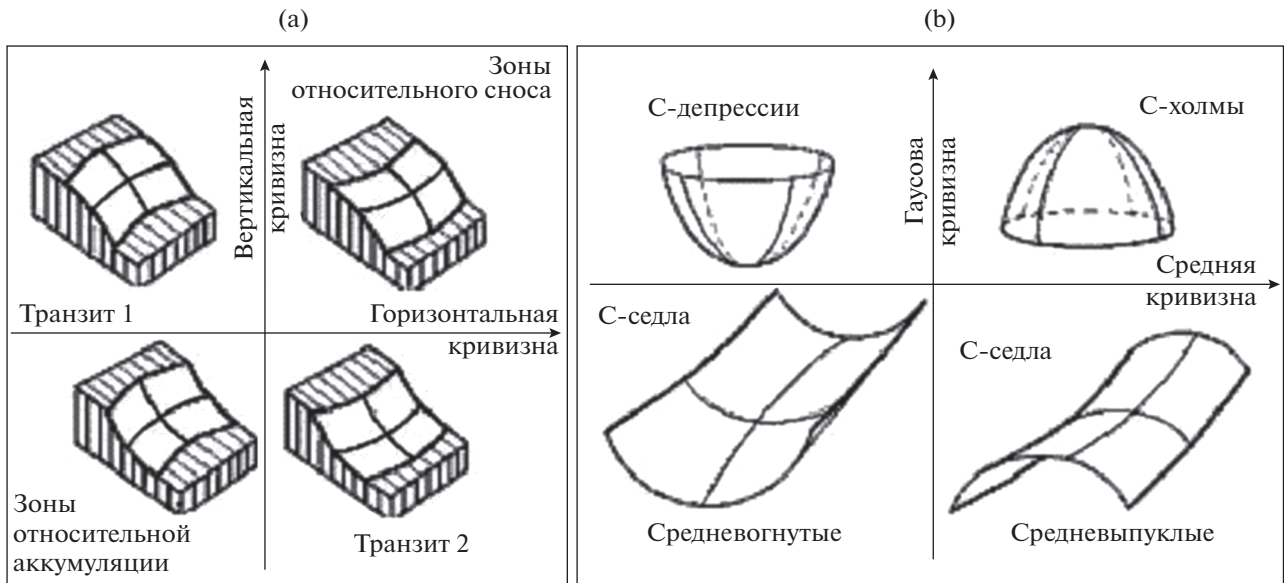


Рис. 2. Классификации форм рельефа Трёха [64] (а) и Гаусса [4] (б).

соответствующих количественных характеристик рельефа, называемых морфометрическими величинами (МВ). Именно областям сближения потоков (белым на рис. 1а), движущихся по земной поверхности под действием силы тяжести, отвечают значения горизонтальной (тангенциальной) кривизны  $kh < 0$ , а областям их расхождения —  $kh > 0$ . Областям относительного замедления потоков (рис. 1б) отвечает отрицательная вертикальная (профильная) кривизна  $k_v$ , а областям их относительного ускорения —  $k_v > 0$ . Основываясь на соображениях этого рода, Трёх [64] в почвоведении предложил классификацию форм рельефа, основанную на знаках  $kh$  и  $k_v$  (рис. 2а).

Классификация Трёха использует свойства потоков, то есть зависит от гравитации. Не зависящую от внешних полей классификацию форм рельефа ранее предложил Гаусс [4] (рис. 2б). Были предложены и другие классификации (обзор [46]), но здесь они не рассматриваются.

Скорость движения веществ, перемещаемых по земной поверхности под действием силы тяжести, обычно пропорциональна крутизне склонов  $G$ . Однако задержка движения веществ растительностью и труднопроходимыми участками, такими как торфяники, означает, что МВ действуют лишь как предпосылки этого движения, зависящего не только от рельефа [13]. Таким образом, значения МВ не описывают это движение полностью.

Помимо  $G$ ,  $kh$  и  $k_v$ , был предложен еще ряд МВ, описывающих свойства земной поверхности в гравитационном поле и в поле солнечного излучения, а также игнорирующих все векторные поля [58].

При учете солнечного излучения особенно важна освещенность склонов  $F(a, b)$ , описывающая перпендикулярность падения солнечных лучей на земную поверхность [58]. Ее выражают в процентах от максимальной интенсивности в ясный день (100%), достигаемой при перпендикулярном падении солнечных лучей на земную поверхность. На тневых склонах  $F = 0$ . Здесь  $a$  и  $b$  есть описывающие положение Солнца углы:  $a$  — склонения Солнца над горизонтом,  $b$  — азимута Солнца, отсчитываемого от севера по часовой стрелке (рис. 3).

Освещенность склонов зависит как от их крутизны, так и от экспозиции склонов. Практическое значение имеет эффективное положение Солнца, которое можно найти следующим образом. В средних широтах коэффициент корреляции такой характеристики ландшафта, как обилие трав, с  $F(a, b)$  очень слабо зависит от угла склонения Солнца  $a$  в практически важном диапазоне  $25^\circ - 45^\circ$  [15]. Поэтому здесь этот угол можно брать фиксированным,  $a_0 = 35^\circ$ . Зависимость  $F(a_0, b)$  от азимута  $b$  значительно сильнее. Значение *эффективного азимута*  $b_0$  можно найти путем построения кривой зависимости от азимута  $b$  коэффициента корреляции  $r$  между изучаемым свойством почв и  $F(a_0, b)$ . Точка максимума  $r$  на этой кривой дает эффективный азимут  $b_0$ , при котором связь данного свойства почв  $S$  и освещенности склонов наиболее тесная. Обычно эффективный азимут  $b_0$  отвечает в Северном полушарии юго-западным склонам ( $b_0 = 210^\circ - 250^\circ$ ), которые лучше прогреваются из-за отставания прогрева корнеобитаемого слоя почвы от мгновенного сол-

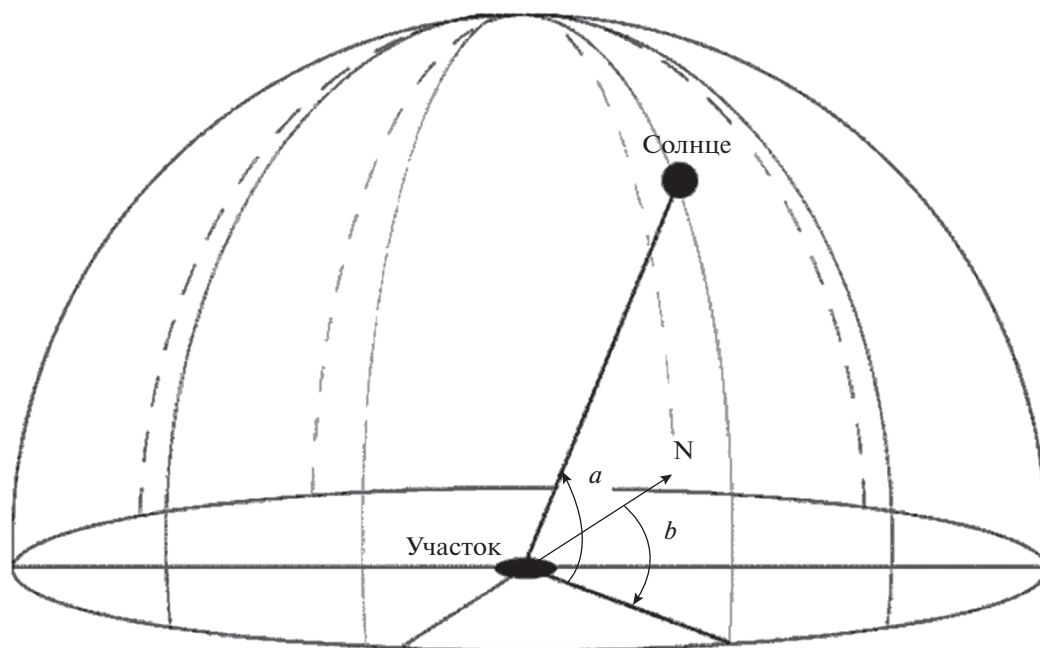


Рис. 3. Определяющие положение Солнца углы.

нечного излучения [24, 40]. Хотя углы в  $F(a_0, b_0)$  фиксированы, она по-прежнему зависит от крутизны и экспозиции склонов, меняясь от одной точки наблюдения к другой.

Освещенность склонов можно выразить не только в процентах, но и в энергетических единицах ( $\text{Вт}/\text{м}^2$ ), используя тот факт, что в ясный день мощность солнечной радиации на единицу площади перпендикулярной солнечным лучам площадки составляет в средних широтах  $\tau = 760 \text{ Вт}/\text{м}^2$  [10]. Тогда освещенность в энергетических единицах  $F_{\text{ЕН}}$  есть  $F_{\text{ЕН}}(a, b) = \tau F(a, b)/100 \text{ Вт}/\text{м}^2$ . Хотя в пасмурные дни преобладает рассеянное излучение, это лишь снижает контраст карты  $F$ , мало влияя на узор карты освещенности, но лишь несколько уменьшая тесноту связи между изучаемым свойством почв и  $F$  [55].

Интегрирование движения Солнца за определенное время (например, за вегетационный период) дает дозу солнечной радиации **PRR** (Potential Relative Radiation) в  $\text{Дж}/\text{м}^2$  [55], не зависящую от углов  $a$  и  $b$ . Однако для **PRR** нет возможности найти эффективные азимут и склонение (так как она не зависит от  $a$  и  $b$ ), вследствие чего **PRR** выделяет склоны, получающие наибольшее количество света (южные в Северном полушарии), а не наиболее теплые (юго-западные). Поэтому использование  $F(a_0, b_0)$  или  $F_{\text{ЕН}}(a_0, b_0)$  в предсказательном картографировании удобнее, чем **PRR**. Например, усредненная по годам урожайность озимой пшеницы в Окском бассейне была наибольшей для  $F(a_0 = 35^\circ, b_0 = 235^\circ)$ , а эта освещенность для

нее — наиболее значимым фактором среды [16, 59], возрастание которого на 5% приводило к увеличению урожайности в 1.5–2 раза.

Отметим, что в глобальном масштабе освещенность или доза солнечной радиации сильно влияют на распределение климата по Земле, поскольку падение солнечных лучей на земную поверхность становится все менее перпендикулярным по мере приближения к полюсам, определяя их более холодный климат; к этому добавляется более сильное поглощение солнечной радиации атмосферой при малых углах склонения  $a$  и другие эффекты [10].

Иногда для простоты вместо освещенности используют равную 1 для юго-западных склонов функцию экспозиции  $A_0$ ,  $-\sin(A_0 + 45^\circ)$  [24, 40], которая заранее предполагает юго-западные склоны (в Северном полушарии) самыми теплыми. Однако они не всегда наиболее теплые по многим причинам, например, из-за преобладания ветров определенного направления. Отметим, что саму экспозицию  $A_0$  некорректно использовать в статистических сравнениях из-за ее цикличности ( $0^\circ$  и  $360^\circ$  означают одно и то же — северные склоны), поэтому вместо  $A_0$  используют нециклические функции экспозиции, такие как  $\sin A_0$  (“восточность” склонов) и  $\cos A_0$  (“северность” склонов). В тайге близ р. Печоры отрицательная связь обилия таежной растительности с  $\sin A_0$  отвечала негативному влиянию преобладающих здесь холодных ветров с востока [18].

Для биотических свойств почв, таких как содержание органического углерода, в списке веду-

щих факторов среды нередко оказывается освещенность склонов — поскольку с ней связаны фотосинтез и тепловой режим почв, а для абиотических, таких как концентрации тяжелых металлов, важнее могут быть другие МВ, характеризующие механизмы аккумуляции, расчлененность рельефа или его формы [17].

Поскольку роль рельефа в предсказательном картографировании существенна, и он является одним из факторов почвообразования, кратко опишем смысл МВ. Как уже отмечалось, два механизма аккумуляции описывают горизонтальная кривизна  $kh$  и вертикальная кривизна  $kv$ . В англоязычной литературе их обычно называют тангенциальной кривизной и профильной кривизной соответственно [51]. Равная половине их суммы средняя кривизна  $H$  положительна в зонах относительной аккумуляции, отрицательна в зонах относительного сноса, но может иметь любой знак в зонах транзита (рис. 2а). Это уменьшает полезность  $H$  для различения форм рельефа Треха. Лапласиан высоты очень близок к  $-2H$  [58] и поэтому представляет небольшую самостоятельную ценность.

Расчлененность (изрезанность, сложность) рельефа описывается избыточной горизонтальной кривизной  $khe$  в плане и избыточной вертикальной кривизной  $kve$  в профиле, а также их произведением, полной кольцевой кривизной  $KR$  [58]. Введенный в [14] ротор  $rot$  обладает несколько необычными свойствами. Его модуль,  $|rot|$ , отвечает абсолютному значению кривизны линий тока, по которым течет вода, то есть описывает расчлененность в плане, но его знак отвечает направлению поворота (закручивания) этих линий:  $rot > 0$  при повороте этих линий по часовой стрелке и  $rot < 0$  в противном случае. Поэтому применение ротора пока ограничено, хотя ясно, что он может различать левые ( $rot < 0$ ) и правые ( $rot > 0$ ) борта долин.

Формы рельефа по Гауссу [4] описываются средней кривизной, максимальной кривизной и минимальной кривизной, а также несферичностью [58]. Несферичность обращается в ноль на участках поверхности сферы любого радиуса и возрастает на вытянутых формах рельефа, таких как долины и водоразделы.

Особое место занимают региональные (нелокальные) МВ, которые требуют для своего расчета интегрирования. К ним относится в первую очередь площадь сбора МСА [34, 49], которая показывает в каждом пикселе, с какой максимальной площади со всех вышележащих склонов в него может собираться вода, растворимые соли и другие жидкие и сыпучие вещества. При этом также рассчитываются глубины депрессий. Впитывание и испарение воды не учитываются, то есть МСА полностью определяется рельефом, не

являясь результатом гидрологического моделирования, где МСА также очень полезна.

Получили определенную популярность также составные МВ. Например, если вода в данный пиксель собирается с достаточно большой площади (велика МСА), а крутизна  $G$  в этом пикселе достаточно мала, то в таких участках (пикселях) вода может застаиваться, формируя зоны насыщения. Для описания последних используют составную МВ, возрастающую как с ростом площади сбора МСА, так и с уменьшением крутизны, так называемый топографический индекс  $TI = \ln(1 + MCA/G)$  [27], известный как индекс влажности почв.

На участках достаточно малой протяженности характеристики климата мало меняются, и наиболее тесная связь свойства почв обнаруживается с рельефом. Поэтому здесь на первый план выступают МВ. Использование их на этом масштабе в качестве факторов среды в предсказательном картографировании привело к построению ряда моделей, хорошо описывающих ряд свойств почв с помощью только одного фактора почвообразования — рельефа. Например, в работе [62] рельеф убедительно описывал пространственное распределение мощности горизонта А почв на участке запада штата Миннесота, США, с помощью МСА,  $TI$ ,  $kh$ ,  $kv$  и глубин депрессий (рис. 4).

В работе [17] на малом участке юга Московской области изучалась связь концентраций тяжелых металлов и органического углерода с рельефом (рис. 5).

Влияние климата обнаруживается как важное при изучении участков большего размера (~100 км и больше). Это не значит, что на малых участках климат не влияет на почву, но изменения характеристик климата на небольших участках малы по сравнению с ошибками их определения, измеренные значения характеристик климата в различных точках малого участка примерно совпадают, и поэтому их нельзя использовать как факторы среды, закономерно меняющиеся от точки к точке.

**Описание климата.** Значения осадков и температур еще не отвечают на вопрос о том, является климат сухим или влажным [6, 63]. Для этого нужно знать соотношение годовых осадков  $MAP$  и годовой испаряемости (потенциальной эвапотранспирации)  $PET$ , то есть коэффициент увлажнения  $MI = MAP/PET$  [6]. По определенным диапазонам значений безразмерного показателя  $MI$  приближенно выделяют известные природно-климатические зоны [7]: пустыню ( $MI < 0.12$ ), полупустыню (0.12–0.3), степь (0.3–0.6), лесостепь (0.6–1), тайгу и тундру ( $MI > 1$ ). Другой мерой оценки сухости/влажности климата является дефицит воды, определяемый как разность испаряемости и испарения:  $WD = PET - AET$  [60]. Пола-

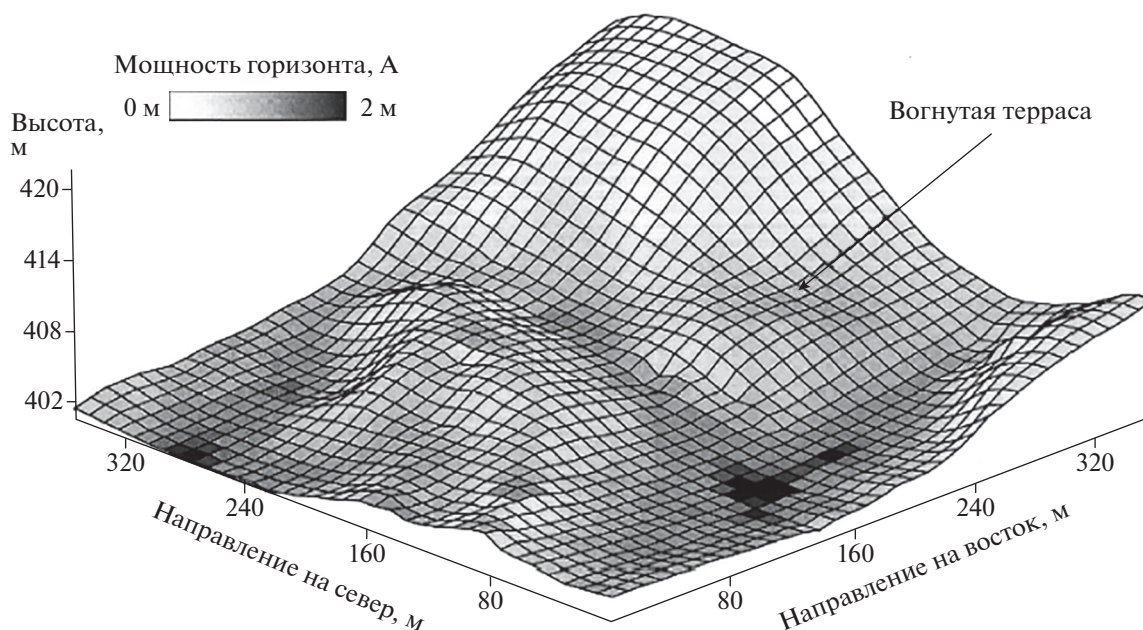


Рис. 4. Трехмерная предсказательная карта мощности горизонта А почвы [62], рассчитанная по рельефу с разрешением 10 м (размер квадратика на карте).

гают, что для каждого вида деревьев существует наибольшее значение среднегодового дефицита воды, при превышении которого этот вид не растет [45]. Оценки таких критических значений WD для основных типов леса в Волжском бассейне сделаны недавно в работе [19]. WD быстро увеличивается с ростом температуры даже при постоянных осадках; его расчет на 2050 г. для парка Yosemite в Калифорнии по данным прогноза изменения климата дал увеличение WD на 23%, что может привести здесь к исчезновению двух видов деревьев, находящихся на южной границе своей области распространения [45].

Испаряемость PET, то есть наибольшее возможное испарение, достигаемое при увлажнении поверхности, например, в болотах, не измеряется метеостанциями. Поэтому в работе [63] предложена эмпирическая методика расчета PET по данным метеостанций; относительная ошибка расчета PET оценена в этой работе как 2.6%. Предложены и более точные методики оценки PET [21], но они требуют не всегда доступных дополнительных параметров, таких как скорость ветра или запасы влаги в метровом слое почвы [45]. Методику оценки суммарного испарения АЕТ (actual evapotranspiration) предложил Будыко [3]. Описание обеих этих методик воспроизведено в работе [19]. При расчете годового АЕТ накопление и уход влаги из почвы за год приблизительно компенсируются [3], поэтому влияние запаса и ухода влаги из почвы на годовое испарение АЕТ невелико. Однако при расчете месячных значений испарения это влияние уже

не мало, и необходимы более точные методы, такие как метод Пенмана—Монтейта [21], учитывающий влияние солнечной радиации, растительности, ветра и др., то есть требующий трудоемкого и не всегда доступного измерения ряда дополнительных характеристик.

Помимо этих характеристик климата, важными для ряда регионов могут быть и другие, такие как континентальность климата IC, определяемая как амплитуда температуры года, то есть обычно как разность максимальной и минимальной среднемесячных температур за год,  $IC = T_{\max} - T_{\min}$  [42]. Отмечено, например, что в европейской России виды деревьев последовательно перестают расти в направлении примерно с запада на восток с ростом IC, сначала бук, затем граб, ясень, далее дуб черешчатый, который перестает расти при  $IC > 42^\circ\text{C}$  [11]. В то же время в направлении с севера на юг исчезновение видов деревьев, например, ели, ясеня и дуба, определяется в основном возрастом дефицита воды WD [11, 19]. С ростом WD в этом направлении также закономерно меняются и типы почв.

В различных целях используют также составные характеристики климата, такие как гидротермический коэффициент Селянинова, сумму активных температур и другие. Зимой снежный покров защищает растительность от вымерзания корней, но часто его толщина под лесом неизвестна; для его оценки используют индекс снежности, определенный как сумма осадков холодного периода [42]. Для востока России и Японии было пока-

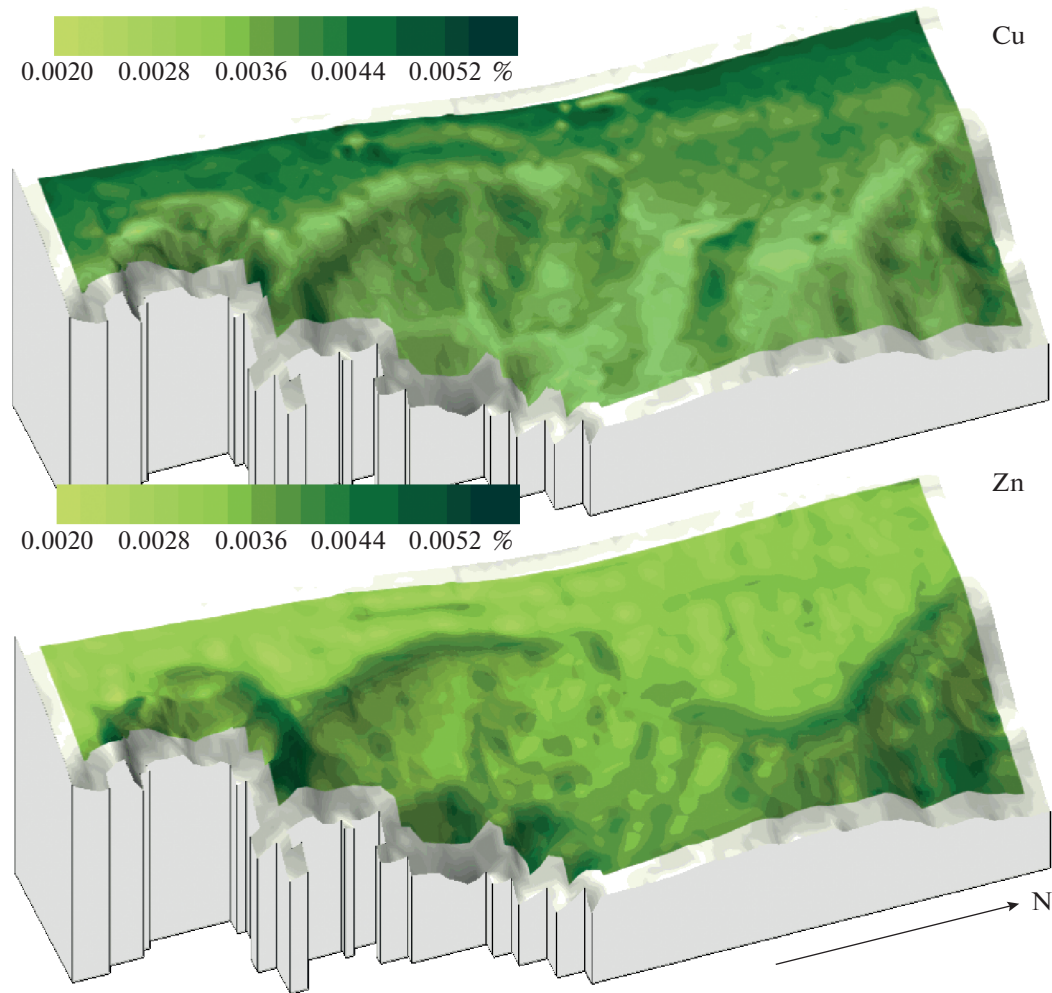


Рис. 5. Трехмерные предсказательные карты распределения Cu и Zn в слое 0–20 см почв участка на юге Московской области [17], разрешение 1 м.

зано, что определение границ групп леса здесь лучше проводить по индексам Кира [34]. Тепловой индекс Кира есть  $WK = \sum \max\{0, (T_i - 5)\}^\circ C$ , где  $T_i$  – средняя температура  $i$ -го месяца, а холодовой индекс Кира есть  $СК = -\sum \max\{0, (5 - T_i)\}^\circ C$  [19, 42]. Однако индексы Кира носят, по-видимому, частный (региональный) характер и их значимые применения вне указанного региона нам пока неизвестны.

Данные о среднемноголетних осадках и температурах каждого месяца (средние за 50 лет, 1950–2000 гг.) открыто доступны с высоким разрешением 1 км в глобальной базе данных WorldClim [37], сайт [www.worldclim.org](http://www.worldclim.org). Они получены интерполяцией данных от метеостанций (по осадкам – всего 44 554 метеостанции).

Помимо климата и рельефа, в предсказательном картографировании могут использоваться и другие факторы среды, такие как расстояния от

различных объектов (рек, геологических разломов, дорог, вулканов, гор и др.), данные спутниковых снимков и другие. Однако расстояния часто очень косвенно отражают факторы почвообразования, и они обычно более эффективны для частных приложений. Например, расстояние от разломов существенно для оценки вероятности оползней [28].

#### МЕТОДЫ ПРЕДСКАЗАТЕЛЬНОГО КАРТОГРАФИРОВАНИЯ

Концептуальная модель SCORPAN (уравнение (1)), отражающая связь изучаемого свойства почв  $S$  с другими величинами, является скорее “памяткой” о том, какие факторы среды могут быть теоретически важны для оценки связи  $S$  с факторами среды. Непосредственно в этой модели функция  $f(s, c, o, r, p, a, n)$  остается неизвестной. Время  $a$ , хотя оно и входит в SCORPAN,

практически не используется в моделях предсказательного картографирования [49]. Как показано выше, наиболее хорошо описаны характеристики рельефа и климата, факторов почвообразования. Труднее описать такие факторы, как геология и организмы, включая антропогенный фактор [49], поскольку они очень многоплановы и их влияние на почвы качественно разнообразно, порой быстро меняется (антропогенное воздействие) и труднее поддается формализации, то есть количественному описанию. Тем не менее, они тоже являются факторами почвообразования, которые надлежит учитывать в моделях предсказательного и цифрового картографирования почв, даже если их описание непрерывной числовой переменной, как освещенности, нередко затруднительно (например, типы растительности). Поэтому методы предсказательного картографирования могут значительно выигрывать от включения в них возможности учета неколичественных характеристик окружающей среды (см. ниже).

**Общий принцип предсказательного картографирования** можно пояснить следующим образом. Пусть данное свойство почв  $S$  зависит от факторов среды,  $S = F(c, r, o, \dots)$ , где функция  $F$  нам пока неизвестна. Разлагая эту функцию в степенной ряд и ограничиваясь квадратичными членами, получаем

$$S = a_0 + a_1c + a_2c^2 + a_3r + a_4r^2 + a_5cr + a_6o + a_7o^2 + \dots + a_m \text{Factor}_m, \quad (2)$$

где  $a_i$  есть неизвестные, подлежащие определению коэффициенты. Уравнение (2) называется уравнением множественной регрессии [52].

Будем считать, что значения  $S$ , а также факторов среды  $c, r, o, \dots$  измерены и известны нам в каждой из  $n$  точек наблюдения. Тогда у нас есть  $n$  уравнений (2) с  $m + 1$  неизвестными  $a_0, \dots, a_m$ , причем примем, что  $n > m + 1$ , то есть точек наблюдения достаточно много (обычно  $n \geq 25$ , но может достигать и сотен). Уравнение (2) решают методом наименьших квадратов (ordinary least squares, **OLS**), то есть используя критерий минимальности суммы квадратов отклонений левых частей (2) от правых  $\sum_n (S_k - \text{ПЧ}_k)^2 = \min$ , где  $\text{ПЧ}_k$  есть правая часть (2) в  $k$ -ой точке наблюдения. Условия минимальности, то есть  $\partial[\sum_n (S_k - \text{ПЧ}_k)^2] / \partial a_i = 0$ , дают  $m + 1$  линейных по  $a_i$  уравнений с  $m + 1$  неизвестными  $a_0, \dots, a_m$ , которые известным способом находят из этой системы уравнений [52].

Теперь, зная коэффициенты  $a_0, \dots, a_m$  и факторы среды в точках карты, мы знаем, таким образом, правую часть (2) и поэтому, хотя значения  $S$  измерены только в  $n$  точках, мы можем построить карту изучаемого свойства почв  $S$  по уравнению (2) во всех точках карты (их обычно значительно

больше, чем  $n$ ), поскольку факторы среды в них известны. Задача нахождения функции  $F(c, r, o, \dots)$  в указанном приближении, таким образом, решена, и формально можно построить предсказательную карту, хотя мы еще не рассмотрели вопросы о тесноте связи между  $S$  и  $F(c, r, o, \dots)$ , о статистической значимости отдельных факторов среды и о проверке полученной модели (см. ниже).

Для иллюстрации принципа предсказательного или цифрового картографирования почв использовали уравнение множественной регрессии (2), решаемое методом OLS. Это не значит, что данный метод является единственным или наилучшим, хотя он наиболее прост и популярен [49]. Например, при изучении большой местности, сильно различающейся в различных своих частях, например, условиями климата, может оказаться, что неявно принятое в OLS предположение о постоянстве регрессионных коэффициентов  $a_0, \dots, a_m$  для всей местности не является адекватным задаче. В таких случаях используют другие статистические подходы, такие как модель географически взвешенной регрессии, в которой принимается, что  $a_0, \dots, a_m$  медленно меняются на местности. Например, при разработке карты содержания полного фосфора в слое 0–30 см почв Австралии [65] и карт почвенного органического углерода в Австралии [66], использовалась модель Cubist этого рода. Дополнительно к значениям  $S$ , эта модель давала карты ошибок в оценке  $S$ , позволяющие выделять области с малыми и большими неопределенностями в оценке  $S$ . В работе [66] в качестве предикторов использовали данные дистанционного зондирования почв спектроскопическими сенсорами и сенсорами гамма-излучения, а также данные карт растительности. Рассмотрение различных статистических моделей дано, например, в работах [28, 35, 48, 49, 56]. Таких моделей для разных ситуаций немало и здесь мы не будем останавливаться на них, концентрируясь главным образом на описании факторов среды и принципов анализа. Для определенности будем использовать с этой целью OLS, но основные принципы остаются справедливыми и для других статистических подходов.

**Значимость факторов среды и их отбор.** Изучаемое свойство почв  $S$  может не полностью определяться учтенными факторами среды, доступный набор которых всегда ограничен. Поэтому *все* факторы среды в уравнении  $S = F(\text{factor}_1, \dots, \text{factor}_m, \dots)$  на практике недоступны. В силу этого уравнение типа (2) описывает лишь некоторую *долю* изменчивости  $S$ , объясненную вошедшими в (2) факторами среды. Если бы изучаемое свойство почв  $S$  точно описывалась учтенными в (2) факторами среды, то эта доля равнялась 1.

В любых моделях предсказательного картографирования описание  $S$  является по своей природе



статистическим, и та статистика, которая описывает эту долю, называется коэффициентом детерминации  $R^2$ . При решении уравнения (2) находят  $R^2$ , причем  $R^2$  можно найти для любой статистической модели [35]. Полезно также знать, какие из факторов среды значимы для  $S$ , а какие – незначимы. В методе OLS для этого используются так называемые  $t$ -статистики Стьюдента [52], рассчитываемые для каждого фактора среды. С их помощью можно удалить из уравнения (2) все незначимые факторы среды, а остающиеся расположить по убыванию значимости (первый в уравнении – самый значимый). Процедура выбора наиболее значимых факторов среды в уравнении (2) называется отбором предикторов [52]. Она важна прежде всего потому, что точность модели уменьшается с ростом числа предикторов в ней [52]. При отборе предикторов выбираются те факторы среды (предикторы), которые дают наибольший  $R^2$ ,  $t$ -статистики или их аналоги позволяют оценить также значимость (относительную роль) каждого из вошедших в финальную модель факторов среды.

**Независимость факторов среды.** Если повторить один и тот же фактор среды дважды в уравнении (2), то регрессионные коэффициенты не будут найдены – система уравнений не имеет решения [52]. Факторы среды должны быть *линейно независимыми*, чтобы уравнение (2) имело решение. Более того, если два фактора среды линейно независимы, но близки к линейно зависимым (как это бывает, например, с температурами соседних месяцев, которые часто близки к пропорциональным друг другу), то статистики могут начать противоречить друг другу. Например, чтобы связь левой и правой частей уравнения (2) была значимой, нужно, чтобы связь  $S$  хотя бы с одним из факторов среды была значимой. В случае близких к линейно зависимым факторам среды этого может не быть. Это известно в предсказательном картографировании как проблема *мультиколлинеарности* [35, 52].

Поэтому факторы среды должны быть близки к линейно независимым. Ясно, что мало смысла использовать температуры соседних месяцев, если они пропорциональны друг другу. Тогда один из этих факторов среды просто исключают из рассмотрения. Для оценки линейной независимости предикторов в OLS используют так называемые факторы вспухания дисперсии (variance inflation factors) **VIF**, наибольший из которых не должен превосходить 5.15 [15, 17]. В итоге, для отбора факторов среды используют только те комбинации предикторов, для которых  $\max\{VIF_i\} \leq 5.15$  (все комбинации предикторов, для которых  $\max\{VIF_i\} > 5.15$ , исключают из рассмотрения), и доля объясненной факторами среды дисперсии  $S$ , то есть  $R^2$ , является наибольшей.

В связи с этим в предсказательном картографировании принято называть  $S$  зависимой переменной, а факторы среды – независимыми переменными или предикторами. Если, например,  $R^2 = 0.68$ , то говорят, что 68% дисперсии зависимой переменной объясняется учтенными в модели предикторами [52].

Следует отметить, что некоторые “самообучающиеся” непараметрические статистические методы предсказательного картографирования не требуют независимости предикторов. Таковы, например, методы искусственной нейронной сети, применяемые иногда в цифровой почвенной картографии [26]. Однако, хотя они могут давать предсказательные карты  $S$ , они не дают никакого уравнения и не выявляют ведущих предикторов [35].

**Неколичественные переменные.** Некоторые факторы среды, такие как типы растительности или почв, трудно выразить одной или немногими непрерывными количественными переменными. Между тем они могут существенно влиять на свойство почв  $S$ , и потому их учет в моделях предсказательного картографирования весьма желателен. Для этого используют так называемые индикаторы (dummy variables), принимающие обычно значения 0 или 1 (в общем случае – два разных числа). Например, если  $Ind_1 = 1$  отвечает черноземам, а  $Ind_2 = 1$  – серым лесным почвам, то ясно, что в данной точке не может быть  $Ind_1 = Ind_2 = 1$ , так как почва в ней не является одновременно и черноземом, и серой лесной. Значения же  $Ind_1 = Ind_2 = 0$  возможны и отвечают тому, что почва не является ни черноземом, ни серой лесной (например, дерново-подзолистая). Поэтому для описания  $N$  типов почв нужен  $N-1$  индикатор [15, 52].

Таким образом, в предсказательном картографировании могут учитываться как количественные, так и неколичественные предикторы (такие как типы растительности, типы антропогенного воздействия или наличие/отсутствие оглеения). Это расширяет возможности подходов. Также вместо количественного предиктора  $X$  можно рассматривать нелинейную функцию  $f(X)$ , которая, например, более обоснована физически и увеличивает  $R^2$ .

Если в качестве  $S$  рассматривается наличие/отсутствие данного свойства почв, то может изучаться вероятность  $p$  наличия этого свойства. В таких ситуациях используют логит-регрессию (logistic regression), то есть решают уравнение вида  $\ln[p/(1-p)] = F(s, c, o, \dots)$ , где функция  $\ln[p/(1-p)] = \text{logit}(p)$  называется логит. В таких случаях решение находят не методом наименьших квадратов, а методом максимального правдоподобия [39].

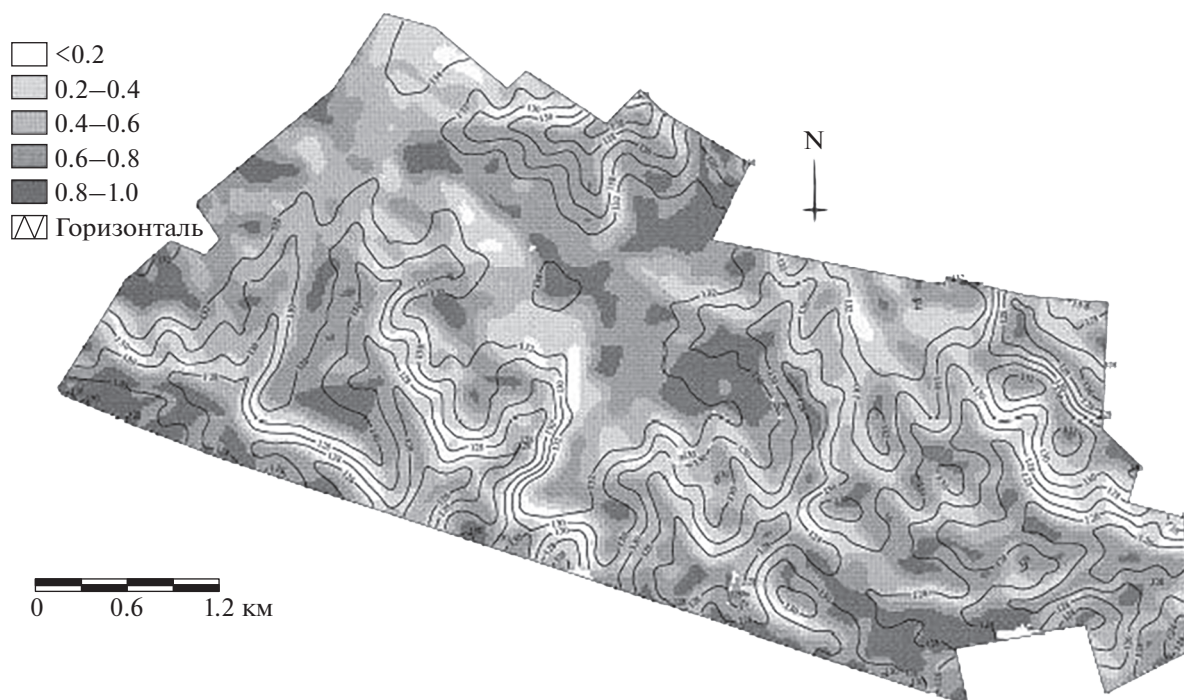


Рис. 6. Предсказательная карта вероятности обнаружения иловато-суглинистого горизонта почв [41], рассчитанная по модели (3).

Например, в работе [41] изучалось наличие/отсутствие некарбонатного иловато-суглинистого горизонта почв на небольшом (размером  $<10$  км) участке близ Парижа. Из-за малости участка характеристики климата не использовались. Для этого пологого участка (максимальный уклон  $<1.1^\circ$ ) роль горизонтальной и вертикальной кривизны была незначимой, а основную роль играли крутизна в градусах  $GA$  и экспозиция склонов  $A_0$ . Используя измерения в 246 точках наблюдения, авторы получили уравнение логит-регрессии в виде

$$\ln \left[ \frac{p}{1-p} \right] = 1.69 - 1.62 \sin GA + 0.60 \sin A_0 + 1.01 \cos A_0, \quad (3)$$

где  $p$  есть вероятность наличия иловато-суглинистого горизонта. Предсказательная карта показана на рис. 6.

Сравнение этой карты с наблюдениями на местности показало [41], что ее можно было бы улучшить с помощью учета сухих долин, то есть площади сбора МСА, расчет которой в то время еще не был доступен авторам [41].

**Пространственные связи.** В модели SCORPAN  $S = f(s, c, o, r, p, a, n)$  фактор  $n$  отвечает пространству. Под этим понимается то, что в расположенных поблизости точках наблюдения значения свойства почв  $S$  нередко (но не всегда) ближе, чем в более отдаленных друг от друга точках. Это яв-

ление отвечает автокорреляции  $S$  (то есть корреляции между значениями  $S$  в различно удаленных точках) и изучалось еще в работе Козловского [8] по элементарному почвенному ареалу, который впервые использовал автокорреляционную функцию (спектр) для оценки размеров почвенного индивидуума. Эту или подобные функции стали в дальнейшем использовать многие авторы, в том числе с использованием моделей геостатистики [9, 48], где предполагается наличие значимой автокорреляции. Используют также различные варианты геостатистики (регрессионный кригинг и др.), которые наряду с автокорреляцией учитывают и связь почв с факторами среды [48].

В авторегрессионных моделях учитывается только автокорреляция, то есть вообще не учитываются факторы среды [25]. Однако эти модели не нашли заметного использования в предсказательном картографировании. Различные статистические модели с учетом как автокорреляции, так и факторов среды описаны в обзорах [23, 29].

Общим свойством автокорреляции является то, что она обычно быстро ослабевает с ростом расстояния между точками наблюдения. Поэтому при расстоянии между ними 250 м в лесах Австрийских Альп уже не было значимой автокорреляции [44]. В моделях OLS автокорреляция в предикторах не имеет значения, важна лишь автокорреляция в остатках [52], то есть в разности между левой и правой частями (2). По этим при-

чинам фактор  $n$  в модели SCORPAN учитывают не всегда, что может, однако, приводить к объяснению меньшей доли свойства почв  $S$  факторами среды (то есть к меньшим  $R^2$ ), чем при учете пространственного фактора  $n$ .

**Роль масштаба.** В работе [61] при изучении ряда свойств почв во Франции найдено, что 70% дисперсии этих свойств объяснялось рельефом при разрешении 50 м, а при разрешении 120 м – лишь 55%. Отчасти это связано с зависимостью характеристик рельефа от масштаба [58]. Отметим, что при изучении протяженных местностей увеличивается роль климата (который незначимо меняется на участках малой протяженности), так что свойства почв или растительности на участках субконтинентальной протяженности могут оказываться более тесно связанными с климатом. Например, более 90% дисперсии разнообразия сосудистых растений объясняется климатом в глобальном масштабе при разрешении 100 км [36], но лишь 50% – в региональном масштабе при разрешении 5.6 км [53]. Однако некоторые характеристики почв, такие как педогенный прирост магнитной восприимчивости в верхнем слое развитых на лёссах степных почв, могут быть тесно связаны с климатом, например, с коэффициентом увлажнения ( $R^2 = 0.85$ ), и в региональном масштабе [1].

**Проверка моделей.** Для проверки предсказательной модели используют часть (~1/3) точек наблюдения, которые не употреблялись при формировании модели. Измеренные в них значения изучаемого свойства почв  $S$  сравнивают с предсказанными в этих точках моделью значениями. Наиболее популярной для этой цели величиной является среднеквадратичное отклонение RMSE [48], которое определяется формулой

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n [S(x_i) - S^*(x_i)]^2}, \quad (4)$$

где  $S(x_i)$  – измеренное значение  $S$ , а  $S^*(x_i)$  – предсказанное моделью. Используют также [48] среднюю ошибку

$$\text{ME} = \frac{1}{n} \sum_{i=1}^n [S(x_i) - S^*(x_i)], \quad (5)$$

которая позволяет оценить преобладающий знак ошибки. Помимо того, что эти показатели дают оценку неточности (неопределенности) предсказания, они позволяют сравнивать различные предсказательные модели, если они были применены к одному и тому же участку [48].

Для некоторых участков исследования число точек наблюдения может быть слишком мало ( $n \sim 30-50$ ) для разделения его на две части, для калибровки модели и для ее проверки. В таких случаях для проверки используют те же точки

наблюдения, последовательно удаляя их по одной и получая “предсказательный” коэффициент детерминации  $R_{pr}^2$  [52], который обычно меньше, чем обычный  $R^2$ . Тогда величину  $\text{Degr} = 100 \times [(R^2/R_{pr}^2) - 1]$  можно использовать как показатель деградации модели в новых (где не было измерений) точках наблюдения того же участка исследования. Эмпирически показано [15], что при  $\text{Degr} < 50\%$  модель OLS успешно предсказывает изучаемое свойство почв.

Если данный фактор среды имеет определенный диапазон на данном участке исследования, и пытаемся применить предсказательную модель в области значений вне этого диапазона, то модель может дать значительные ошибки [52]. С этим связано то, что разработанная для данного участка модель может плохо работать на другом участке, где, например, другие климатические условия [35].

Модели искусственной нейронной сети, а также генерализованные аддитивные модели, обладают нежелательным свойством преувеличивать роль нелинейностей в связи  $S$  с факторами среды, в то время как OLS и их обобщение, генерализованные линейные модели, такого свойства не имеют [38], как и модель “случайного леса” (Random Forest) [33].

Детальные данные о рельефе разрешения 1–5 м, полученные путем лазерного сканирования LiDAR с воздушных летательных аппаратов, становятся все более доступными [12], в том числе под лесом [43]. С использованием таких данных строились предсказательные карты высокого разрешения (~1 м) ряда свойств почв, например, pH в лесу [22]. Как уже отмечалось, данные о среднемесячных осадках и температурах доступны сегодня с разрешением 1 км [37].

Если набор факторов среды, относящихся к описанию климата и рельефа, относительно хорошо развит (как показано выше), то набор факторов среды для описания организмов и почвообразующих пород развит пока еще недостаточно. Поиски в этом направлении продолжаются. Например, на малых участках для описания почвообразующих пород используют наземные растровые измерения электропроводности [48]. Однако этот метод трудно применить к большим участкам, хотя некоторые дистанционные данные (гамма-спектроскопия и др.) дают определенную информацию о почвообразующих породах, которая уже оказалась полезна при составлении цифровых матриц содержания полного фосфора и органического углерода в почвах Австралии [65, 66]. Сегодня доступны и полученные обработкой спутниковых данных ежегодно (с 2000 г.) обновляемые карты характеристик состава и обилия лесов России разрешения 250 м, причем осуществляет-

ся работа по увеличению разрешения таких карт до 110 м и выше [2]. Это делает все более реалистичным построение достаточно подробных предсказательных карт свойств почв по России. Например, с использованием дистанционных и наземных данных сформирована матрица (карта) запасов органического углерода в верхнем метре почв России разрешения 1 км [20].

Поскольку данные дистанционного зондирования (ДДЗ) покрывают большие территории и ежегодно обновляются [2], их использование в почвоведении становится все более частым. Однако, за небольшими исключениями, такими как гамма-спектроскопия, они дают мало информации о почвах и почвообразующих породах, больше отражая свойства растительности, которая тоже является одним из факторов почвообразования. Поэтому в последние годы ДДЗ все чаще используются для предсказательного картографирования почв. В ДДЗ информация о свойствах почв и растительности часто смешана, и эти сигналы трудно разделить. В некоторой мере для оценки свойств почв могли бы быть полезны данные гиперспектральных снимков, содержащие информацию от сотен каналов (длин волн) и позволяющие оценивать ряд физико-химических свойств почв, но пока прогресс в этом отношении ограничен большим объемом данных и трудностями идентификации конкретных свойств почв. Поэтому расширенные наборы характеристик климата и рельефа остаются сегодня главными из доступных факторов среды, которым и посвящен настоящий обзор.

## ЗАКЛЮЧЕНИЕ

Рельеф и климат характеризуются сегодня наиболее богатой системой количественных показателей, вводимых на единой основе. Все морфометрические величины рельефа рассчитываются с помощью единой цифровой модели рельефа. При этом различаются локальные и региональные, требующие интегрирования, МВ. Одни МВ описывают свойства потоков веществ, перемещающихся под действием силы тяжести по земной поверхности, другие – формы рельефа, третьи, такие как освещенность склонов, связаны с описанием радиационного режима почв, теплового и светового. Для некоторых задач используют также составные МВ. Все МВ являются количественными характеристиками рельефа, и их учет важен в предсказательном картографировании. Учет рельефа особенно важен для участков небольшой протяженности, когда закономерное изменение характеристик климата трудно измерить из-за его малости.

Это не значит, что рельеф важен только на крупных масштабах. Например, закономерное уменьшение освещенности от экватора к полю-

сам приводит к холодным условиям в высоких широтах.

Важно отметить, что климат, помимо обычных характеристик типа осадков и температуры, характеризуется балансовыми показателями, такими как коэффициент увлажнения и дефицит воды. В настоящее время их использование в предсказательном картографировании явно недостаточно, в то время как ясно, что не только природно-климатические зоны сменяются при их изменении, но меняются также свойства почв, включая смену типов почв с изменением климатических показателей. В ряде ситуаций можно наблюдать критические значения климатических показателей, при превышении которых почвы теряют или приобретают определенные свойства. Это может быть связано как с заменой лесов степями, полупустынями, пустынями при большом дефиците воды, так и с явлениями, вызванными избытком влаги в водном балансе, например, заболачиванием или оглеением.

Развитие систем количественных показателей для других факторов почвообразования, организмы (включая человека), геология и время (возраст почв), более затруднительно, но, тем не менее, ведется. Однако время пока практически не используется в моделях предсказательного картографирования.

Концепция SCORPAN позволяет использовать предсказательное картографирование для данного свойства почв, когда предикторы включают в себя не только факторы почвообразования, но также другие, отличные от изучаемого, свойства почв и пространственные связи при значимой автокорреляции.

Все еще важны трудности с расчетом, включая недостаток общедоступных программ, а значит и с использованием ряда важных характеристик рельефа (площадь сбора, освещенность и др.) и климата (дефицит воды и др.). Поэтому представленное здесь описание их особенностей, как надеется автор, будет полезно для развития соответствующих баз данных и программ.

## КОНФЛИКТ ИНТЕРЕСОВ

Автор заявляет об отсутствии конфликта интересов.

## СПИСОК ЛИТЕРАТУРЫ

1. *Алексеев А.О., Митенко Г.В., Шарый П.А.* Количественные оценки палеоэкологических изменений в позднем голоцене на юге Восточно-Европейской равнины на основе магнитных свойств почв // Почвоведение. 2020. № 12. С. 1425–1435. <https://doi.org/10.31857/S0032180X20120023>
2. *Барталев С.А., Егоров В.А., Жарко В.О., Лупян Е.А., Плотников Д.Е., Хвостиков С.А., Шабанов Н.В.*

- Спутниковое картографирование растительного покрова России. М.: ИКИ РАН, 2016. 208 с.
3. Будыко М.И. Тепловой баланс земной поверхности. Л.: Гидрометеиздат, 1956. 256 с.
  4. Гаусс К.Ф. Общие исследования о кривых поверхностях // Об основаниях геометрии. М.: Гос. изд-во технико-теоретической литературы, 1956. С. 123–161.
  5. Докучаев В.В. Русский чернозем // Избр. соч. М.: Сельхозгиз, 1948. Т. 1. 480 с.
  6. Иванов Н.Н. Ландшафтно-климатические области земного шара // Записки географического общества. М.–Л., 1948. Т. 1. 223 с.
  7. Исаченко А.Г. Ландшафтоведение и физико-географическое районирование. М.: Высшая школа, 1991. 366 с.
  8. Козловский Ф.И. Почвенный индивидуум и методы его определения // Закономерности пространственного варьирования свойств почв и информационно-статистические методы их изучения. М.: Наука, 1970. С. 42–59.
  9. Кузякова И.Ф., Романенков В.А., Кузяков Я.В. Метод геостатистики в почвенно-агрохимических исследованиях // Почвоведение. 2001. № 9. С. 1132–1139.
  10. Матвеев Л.Т. Курс общей метеорологии. Физика атмосферы. Л.: Гидрометеиздат, 1984. 752 с.
  11. Погребняк П.С. Общее лесоводство. М.: Колос, 1968. 440 с.
  12. Рьельский И.А. Лазерное сканирование и цифровая аэрофотосъемка: новый уровень детальности // Geomatics. 2015. № 4. С. 53–56.
  13. Степанов И.Н., Флоринский И.В., Шарый П.А. О концептуальной схеме исследований ландшафта // Геометрия структур земной поверхности. Пушкино, 1991. С. 9–15.
  14. Шарый П.А. Топографический метод вторых производных // Геометрия структур земной поверхности. Пушкино, 1991. С. 28–58.
  15. Шарый П.А., Рухович О.В., Шарая Л.С. Методология анализа пространственной изменчивости характеристик урожайности пшеницы в зависимости от условий агроландшафта // Агрохимия. 2011. № 2. С. 57–81.
  16. Шарый П.А., Рухович О.В., Шарая Л.С. Предсказательное моделирование характеристик урожая озимой пшеницы // Цифровая почвенная картография: теоретические и экспериментальные исследования. М.: Почв. ин-т им. В.В. Докучаева, 2012. С. 310–326.
  17. Шарый П.А., Пинский Д.Л. Статистическая оценка связи пространственной изменчивости содержания органического углерода в серой лесной почве с плотностью, концентрацией металлов и рельефом // Почвоведение. 2013. № 11. С. 1344–1356. <https://doi.org/10.7868/S0032180X13090104>
  18. Шарый П.А., Смирнов Н.С. Механизмы влияния солнечной радиации и анизотропии местности на растительность темнохвойных лесов Печоро-Ильичьего заповедника // Экология. 2013. № 1. С. 11–19. <https://doi.org/10.7868/S0367059713010113>
  19. Шарый П.А., Шарая Л.С., Сидякина Л.В. Связь NDVI лесов и характеристик климата Волжского бассейна // Современные проблемы дистанционного зондирования Земли из космоса. 2020. Т. 17. № 4. С. 154–163. <https://doi.org/10.21046/2070-7401-2020-17-4-154-163>
  20. Щепашенко Д.Г., Мухортова Л.В., Швиденко А.З., Ведрова Э.Ф. Запасы органического углерода в почвах России // Почвоведение. 2013. № 2. С. 123–132. <https://doi.org/10.7868/S0032180X13020123>
  21. Allen R.G., Pereira L.S., Raes D., Smith M. Crop evapotranspiration (guides for computing crop water requirements). FAO irrigation and drainage paper No. 56. 1998. 300 p.
  22. Baltensweiler A., Walthert L., Ginzler C., Sutter F., Purves R.S., Hanewinkel M. Terrestrial laser scanning improves digital elevation models and topsoil pH modelling in regions with complex topography and dense vegetation // Environmental Modelling and Software. 2017. V. 95. P. 13–21.
  23. Beale C.M., Lennon J.J., Yearsley J.M., Brewer M.J., Elston D.A. Regression analysis of spatial data // Ecology Letters. 2010. V. 13. № 2. P. 246–264.
  24. Beers T.W., Dress P.E., Wensel L.C. Aspect transformation in site productivity research // J. Forestry. 1966. V. 64. P. 691–692.
  25. Besag J. Spatial interaction and the statistical analysis of lattice systems // J. Royal Statistical Society. 1974. Series B. V. 36. № 2. P. 192–236.
  26. Behrens T., Förster H., Scholten T., Steinrücken U., Spies E.-D., Goldschmitt M. Digital soil mapping using artificial neural networks // J. Plant Nutrition Soil Sci. 2005. V. 168. № 1. P. 21–33.
  27. Beven K.J., Kirkby M.J. A physically based, variable contributing area model of basin hydrology // Hydrological Sci. Bull. 1979. V. 24. № 1. P. 43–69.
  28. Bishop M.P., James L.A., Shroder J.F. Jr., Walsh S.J. Geospatial technologies and digital geomorphological mapping: Concepts, issues and research // Geomorphology. 2012. V. 137. P. 5–26.
  29. Dormann C.F., McPherson J.M., Araújo M.B., Bivand R., Bolliger J., Carl G., Davies R.G., Hirzel A., Jetz W., Kissling W.D., Kühn I., Ohlemüller R., Peres-Neto P.R., Reineking B., Schröder B., Schurr F.M., Wilson R. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review // Ecography. 2007. V. 30. № 5. P. 609–628.
  30. Evans I.S. General geomorphometry, derivatives of altitude, and descriptive statistics // Spatial Analysis in Geomorphology. London: Methuen & Co., Ltd., 1972. Ch. 2. P. 17–90.
  31. Franklin J. Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients // Progress in Physical Geography. 1995. V. 19. № 4. P. 474–499.
  32. Freeman T.G. Calculating catchment area with divergent flow based on a regular grid // Computers and Geosciences. 1991. V. 17. № 3. P. 413–422.
  33. Grimm R., Behrens T., Märker M., Elsenbeer H. Soil organic carbon concentrations and stocks on Barro Colorado Island – Digital soil mapping using Random Forests analysis // Geoderma. 2008. V. 146. P. 102–113.

34. *Grishin S. Yu.* The boreal forests of north-eastern Eurasia // *Vegetatio*. 1995. V. 121. P. 11–21.
35. *Guisan A., Zimmermann N.E.* Predictive habitat distribution models in ecology // *Ecological Modelling*. 2000. V. 135. № 2–3. P. 147–186.
36. *Hawkins B.A., Rodríguez M.Á., Weller S.G.* Global angiosperm family richness revisited: linking ecology and evolution to climate // *J. Biogeography*. 2011. V. 38. P. 1253–1266.
37. *Hijmans R.J., Cameron S.E., Parra J.L., Jones P.J., Jarvis A.* Very high resolution interpolated climate surfaces for global land areas // *Int. J. Climatology*. 2005. V. 25. № 15. P. 1965–1978.
38. *Hjort J., Marmion M.* Effects of sample size on the accuracy of geomorphological models // *Geomorphology*. 2008. V. 102. P. 341–350.
39. *Hosmer D.W., Lemeshow S.* *Applied Logistic Regression*. N.Y.: John Wiley & Sons, Inc., 2000. 375 p.
40. *Hwang T., Song C., Vose J.M., Band L.E.* Topography-mediated controls on local vegetation phenology estimated from MODIS vegetation index // *Landscape Ecology*. 2011. V. 26. P. 541–556.
41. *King D., Bourennane H., Isampert M., Macaire J.J.* Relationship of the presence of a non-calcareous clay-loam horizon to DEM attributes in a gently sloping area // *Geoderma*. 1999. V. 89. № 1–2. P. 95–111.
42. *Krestov P.V., Omelko A.M., Nakamura Y.* Phytogeography of higher units of forests and krummholz in North Asia and formation of vegetation complex in the Holocene // *Phytocoenologia*. 2010. V. 40. P. 41–56.
43. *Lim K., Treitz P., Wulder M., St-Onge B., Flood M.* LiDAR remote sensing of forest structure // *Progress in Physical Geography*. 2003. V. 27. P. 88–106.
44. *Lischke H., Guisan A., Fischlin A., Bugmann H.* Vegetation responses to climate change in the Alps - Modeling studies // *A View from the Alps: Regional Perspectives on Climate Change*. Boston: MIT Press, 1998. Ch. 6. P. 309–350.
45. *Lutz J.A., van Wagtenonk J.W., Franklin J.F.* Climatic water deficit, tree species ranges, and climate change in Yosemite National Park // *J. Biogeography*. 2010. V. 37. P. 936–950.
46. *MacMillan R.A., Shary P.A.* Landforms and landform elements in geomorphometry // *Geomorphometry: Concepts, Software, Applications. Developments in Soil Science*. V. 33. Amsterdam: Elsevier, 2009. Ch. 9. P. 227–254.
47. *Martz L.W., de Jong E.* CATCH: a Fortran program for measuring catchment area from digital elevation models // *Computers and Geosciences*. 1988. V. 14. № 5. P. 627–640.
48. *McBratney A.B., Odeh I.O.A., Bishop T.F.A., Dunbar M.S., Shatar T.M.* An overview of pedometric techniques for use in soil survey // *Geoderma*. 2000. V. 97. № 3–4. P. 293–327.
49. *McBratney A.B., Mendonça Santos M.L., Minasny B.* On digital soil mapping // *Geoderma*. 2003. V. 117. № 1–2. P. 3–52.
50. *McKenzie N.J., Ryan P.J.* Spatial prediction of soil properties using environmental correlation // *Geoderma*. 1999. V. 89. № 1–2. P. 67–94.
51. *Mitašová H., Hofierka J.* Interpolation by regularized spline with tension. II. Application to terrain modeling and surface geometry analysis // *Mathematical Geology*. 1993. V. 25. № 6. P. 657–669.
52. *Montgomery D.C., Peck E.A.* *Introduction to Linear Regression Analysis*. N.Y.: John Wiley & Sons, 1982. 504 p.
53. *Moser D., Dullinger S., Englisch T., Niklfeld H., Plutzer C., Sauberer N., Zechmeister H.G., Grabherr G.* Environmental determinants of vascular plant species richness in the Austrian Alps // *J. Biogeography*. 2005. V. 32. P. 1117–1127.
54. *Pike R.J., Evans I.S., Hengl T.* *Geomorphometry: a brief guide // Geomorphometry: Concepts, Software, Applications. Developments in Soil Science*. V. 33. Amsterdam: Elsevier, 2009. Ch. 1. P. 3–30.
55. *Pierce K.B. Jr., Lookingbill T., Urban D.* A simple method for estimating potential relative radiation (PRR) for landscape-scale vegetation analysis // *Landscape Ecology*. 2005. V. 20. № 2. P. 137–147.
56. *Scull P., Franklin J., Chadwick O.A., McArthur D.* Predictive soil mapping: a review // *Progress in Physical Geography*. 2003. V. 27. № 2. P. 171–197.
57. *Shary P.A.* Land surface in gravity points classification by a complete system of curvatures // *Mathematical Geology*. 1995. V. 27. № 3. P. 373–390.
58. *Shary P.A., Sharaya L.S., Mitusov A.V.* Fundamental quantitative methods of land surface analysis // *Geoderma*. 2002. V. 107. № 1–2. P. 1–32.
59. *Shary P.A., Sharaya L.S., Rukhovich O.V.* Model-based forecasting winter wheat yields using landscape and climate data // *Landscape Modelling and Decision Support*. Switzerland: Springer Nature Switzerland AG, 2020. Ch. 20. P. 383–396.
60. *Stephenson N.L.* Actual evapotranspiration and deficit: biologically meaningful correlates of vegetation distribution across spatial scales // *J. Biogeography*. 1998. V. 25. P. 855–870.
61. *Thomas A.L., King D., Dambrine E., Couturier A., Roque J.* Predicting soil classes with parameters derived from relief and geologic materials in a sandstone region of the Vosges mountains (Northeastern France) // *Geoderma*. 1999. V. 90. № 3–4. P. 291–305.
62. *Thompson J.A., Bell J.C., Butler C.A.* Digital elevation model resolution: effects on terrain attribute calculation and quantitative soil-landscape modeling // *Geoderma*. 2001. V. 100. P. 67–89.
63. *Thorntwaite C.W.* An approach toward a rational classification of climate // *Geographical Review*. 1948. V. 38. № 1. P. 55–94.
64. *Troeh F.R.* Landform parameters correlated to soil drainage // *Soil Science Society of America Proceedings*. 1964. V. 28. № 6. P. 808–812.
65. *Viscarra Rossel R.A., Bui E.N.* A new detailed map of total phosphorus stocks in Australian soil // *Science of the Total Environment*. 2016. V. 542. P. 1040–1049.
66. *Viscarra Rossel R.A., Brus D.J., Lobsey C., Shi Z., McLachlan G.* Baseline estimates of soil organic carbon by proximal sensing: Comparing design-based, model-assisted and model-based inference // *Geoderma*. 2016. V. 265. P. 152–163.

## Environmental Variables in Predictive Soil Mapping

P. A. Shary\*

*Institute of Physicochemical and Biological Problems in Soil Science, Russian Academy of Sciences,  
Pushchino, Moscow region, 142292 Russia*

*\*e-mail: p\_shary@mail.ru*

In the well-known conceptual model SCORPAN, given soil property is considered as dependent on the following environmental factors: soil, climate, organisms, topography, time and space. Predictive mapping of soils in digital soil mapping is based on similar ideas, but environmental factors may include not only factors of soils formation, but also, for example, remote sensing data, and have found a wide distribution not only in soil science, but also in ecology, agriculture and geomorphology. This paper provides a review of environmental factors that are used in predictive mapping with a special attention to situations when wide sets of environmental factors may be used and a part of them is not quantitative, such as vegetation types. Most developed are systems of quantitative variables for topography and climate description, so that a special attention is paid to them. Land surface description is performed using both local and non-local variables that need integration. In climate description variables are essential that estimate dry or wet terrain features, such as moisture ratio or water deficit. They need evaluation of potential evapotranspiration that is not measured by meteo-stations, but may be calculated. Possibilities of accounting these and other environmental factors including non-quantitative ones in quantitative statistical models of predictive mapping are described together with principles of their construction, verification, comparison, choice of appropriate models. Examples of soil predictive mapping applications are given for various scales, their specifics for different scales is outlined. Some aspects of remote sensing data usage are discussed.

*Keywords:* Model SCORPAN, digital soil mapping, geomorphometry, climate, spatial relations