

第三次全国土壤普查土壤属性 图与专题图编制技术规范

(修订版)

执笔人：徐爱国 卢昌艾 史舟 黄元仿 朱阿兴
李文西 张海涛 陈守伦 吴克宁 裴久渤
叶回春 李兆富 盛建东 赵小敏 张世文
刘京

国务院第三次全国土壤普查领导小组办公室

2023年2月

目 次

1 适用范围	371
2 制图对象和目的	371
3 制图原则与主要方法	371
3.1 数字土壤制图的原则	371
3.2 数字土壤制图的主要方法	372
4 制图数据准备及要求	372
4.1 土壤制图的数据、基础资料	372
4.2 样点数据整理及处理	372
4.3 环境变量制备及质量检测	373
5 制图思路	375
5.1 土壤属性制图	375
5.2 土壤专题制图	376
6 土壤属性图与专题图制图要求	376
6.1 土壤属性制图	376
6.2 土壤专题图	379
6.3 土壤属性制图相邻区处理	379
6.4 土壤制图结果的验证评价	379
6.5 土壤制图结果的面积统计	379
7 制图比例尺/分辨率与数据质量要求	379
7.1 制图比例尺（分辨率）	379
7.2 制图成果数据质量要求	380
8 专题图表达与质量要求	381
8.1 专题图编制方案设计	381
8.2 专题内容的表达	381
8.3 基础地理要素的选取和表达	381
9 制图设备与场所要求	382
附录 1 第三次全国土壤普查成果图模板	383
附录 2 数字土壤制图的理论基础与主要方法	383

1 适用范围

本规范适用于第三次全国土壤普查（以下简称“土壤三普”）中的土壤属性和专题图制图工作。统一规定了这两类制图的目的、原则、主要方法、制图思路、结果验证、成果图编制要求等。面向有一定数字土壤制图理论和实践基础的科研人员，以及经过数字土壤制图理论和实践学习的技术人员。

2 制图对象和目的

本规范包括土壤三普成果图中的两类图：一是土壤属性图，即土壤理化性状图，包括土壤表层质地、pH、盐碱度、有机质、土壤养分、中微量元素、重金属元素，以及有效土层厚度、0~100 cm 有机碳储量等；二是土壤专题图，包括耕地质量等级图、土壤障碍类型图、退化土壤（盐碱化、酸化等）分布图、黑土资源分布图等专题调查评价图。

通过数字土壤制图方法，采用统一的专题图评价指标，评价土壤质量和适宜性，掌握土壤理化性状空间分布状况和土壤质量底数；编制统一规范的普查成果图，以便指导农业生产和决策。

3 制图原则与主要方法

3.1 数字土壤制图的原则

数字土壤制图（digital soil mapping）方法作为一种新兴的、高效表达土壤及其性状空间分布的方法，较传统手工土壤制图更加高效。尤其在土壤属性制图方面，研究和应用也相对深入和广泛。鉴于数字土壤制图方法仍在不断发展完善，采用该方法进行土壤属性制图，需遵循以下原则：

3.1.1 土壤空间变异尺度效应原则

制图精确度要与制图空间尺度相对应。土壤的空间变化具有尺度效应，并以空间格局的形式呈现，即某一尺度只能揭示相应的变化规律，而某一空间结构只能在某一尺度下体现。在进行大尺度（大空间范围）土壤空间变异分析时，可得到整个区域土壤的空间分布规律，较小尺度下的空间分布特征往往被掩盖；而在小尺度分析时，大多体现的是土壤在微域环境内的变化。不同尺度下其主要影响因子也不尽相同。大尺度土壤空间分布主要与生物气候条件相适应；在较小范围内，土壤形成和发育主要受局部地形、母质等因素的影响。

3.1.2 因地制宜原则

选用相对成熟、区域较优的方法。现有方法均基于一定的数学假设，尚无单一方法或统一固定的环境辅助变量，可以适合不同地貌类型区域。因此，针对制图对象，选择适用的制图方法类别；针对具体土壤属性，根据制图区域特征和范围（尺度），结合样点的密度和均匀度，选用相对成熟的，精度检验较优的方法，且方法不宜过于繁杂。

3.1.3 精度保障原则

建立制图模型前，数据检验须符合制图模型的数学假设。制图方法多采用数学模型，基于统计均值的制图方法，要求样本符合相应的数学假设，例如符合正态分布。样本需验证并符合相关数学假设条件，方可进行模型制图。

数字土壤制图结果，需要进行预测样点验证，评估模型的制图精度。采用训练集和验证集验证的，按照4:1的比例随机选取20%的样点作为验证集，比较实测值与预测值，进行独立验证；也可以采取10%样点交叉验证，在保证精度的同时兼顾计算效率。通过相应的验证指标评估后，制图结果方可采用作为数据成果。对于争议比较大或与专家经验出现巨大差异的图斑区域，需进行实地勘察验证。

3.2 数字土壤制图的主要方法

数字土壤制图方法已广泛用于土壤属性制图。该方法是根据已知点的土壤信息通过数字手段推测其他点土壤特征的过程，以土壤—景观模型为理论基础，以空间分析和数学方法为技术手段，生成数字格式（栅格）的土壤属性空间分布图。比较常用的方法可分为五类：地统计方法、确定性插值、数理统计、机器学习和模糊推理方法。本规范主要采用地统计、机器学习和模糊推理方法，针对不同土壤属性，选择相应的模型制图方法。

地统计方法，包括克里格插值及其衍生方法，有普通克里格、泛克里格、回归克里格、地理加权回归克里格、协同克里格模型等，除普通克里格、泛克里格外，其余的克里格衍生模型是利用所预测土壤属性与环境辅助变量（成土因素）之间的相关性（要素相关性）来提高预测精度。普通克里格应用早而广泛，是本规范的推荐方法之一，主要利用变量空间自相关关系，适合较均一、土壤属性变化不强烈的环境。普通克里格会产生平滑效应，对于局部变异较大地区的预测可能会与实际情况不符。

机器学习模型利用机器学习方法进行数据挖掘，提取土壤属性与环境变量之间的关系用来预测土壤属性的空间分布，解决土壤属性与环境变量的非线性问题，包括随机森林、人工神经网络、分类与回归树等。目前随机森林模型在土壤属性制图领域应用越来越广泛，作为本规范的推荐方法之一。

上述方法有两个条件：第一需要大量的土壤样点来提取统计关系；第二需要具有较好的空间代表性，除机器学习模型外，其他模型制图区域通常不宜过大。

模糊推理是将土壤与环境关系表达为隶属度值，利用单个土壤样点在空间上的代表性推测土壤目标变量的空间变化。该方法制图效果依赖于单个样点的可靠性，要求对样点的可靠性进行质量检查。

地统计和模糊推理方法在中小尺度取得了较高的精度，大尺度下机器学习方法的优势更明显。

4 制图数据准备及要求

4.1 土壤制图的数据、基础资料

土壤属性制图所需要的土壤目标变量、环境辅助变量等数据集。

调查数据：第三次全国土壤普查表层样点理化性状测试数据、剖面样点的土壤类型数据。

环境变量数据：全国第二次土壤普查的1:5万县域数字土壤图、1:1万土地利用现状图、1:5万地形图、1:25万地质图、气象资料以及高分辨率的遥感影像等。

其他数据：相应比例尺的行政区划图等，用于成果图的边界。

4.2 样点数据整理及处理

4.2.1 剖面样点数据整理

有效土层厚度等剖面调查数据，需从剖面点信息中提取，作为深层属性制图样点的基础数据层。

对于耕层点位不足的地区，可由剖面点数据补充。将剖面发生表层土壤属性数据，或者发生表层与亚顶层土壤属性数据经厚度加权平均，转换为耕层数值，加入到耕层点该属性基础数据中。

4.2.2 表层样点数据处理

第一，异常值检验。由于样点采集与化学分析过程的不确定性，需对土壤属性数值进行正态分布检验后做异常值剔除处理，结合数据的常规统计学特征和空间位置，将每个样点的属性值与总体及其邻近8个样点的均值和标准差进行比较，如果样点值在总体均值的5倍标准差之外，且大于或是小于邻近样点均值的3倍标准差，则需对异常值进行核验后剔除。

第二，测试方法分区标注，对不同地区采用不同测试方法的指标，标注其所在区域，用于分别成图。

第三，检查是否存在坐标异常情况，如点位飞出行政区、点位成直线等。

4.3 环境变量制备及质量检测

4.3.1 不同尺度的精度要求

环境变量提取栅格数据精度，要优于表1或表2的像素（像元）分辨率。其中，表1精度适用于大范围土地利用、种植结构比较单一区域，例如平原粮食作物区；表2精度适用于种植结构复杂的小范围地区或地块破碎区域。

由于所涉及的环境协变量种类较多，会出现不同环境协变量具有不同分辨率的情况，此时应根据制图尺度综合到统一的分辨率之下。统一分辨率操作的基本原则是：①尽量从高分辨率向低分辨率综合；②尽量避免从低分辨率到高分辨率内插。

表1 制图比例尺及对应的栅格数据像素（像元）分辨率
（适用于大范围土地利用、种植结构比较单一区域）

比例尺类型	成图比例尺	栅格数据 建议像素分辨率/m
大比例尺	1 : 1 万	5
	1 : 5 万	30
	1 : 10 万	30 或 50
中比例尺	1 : 25 万	90
	1 : 50 万	250
小比例尺	1 : 100 万	1 000

表2 制图比例尺及对应的栅格数据像素分辨率
（适用于小范围种植结构复杂或地块破碎区域）

比例尺类型	成图比例尺	栅格数据 建议像素分辨率/m
大比例尺	1 : 1 万	2.5
	1 : 5 万	10
中比例尺	1 : 25 万	30
	1 : 50 万	90
小比例尺	1 : 100 万	250
	1 : 400 万	1 000

4.3.2 环境变量的提取

利用土壤属性与环境辅助变量之间的相关性模型，需使用环境变量数据。目前主要利用除时间因素外的成土因素信息。特别是在地面有起伏的区域，因样点数量的局限，可采用此类模型提高制图精度。这类模型均需提取栅格格式图层数据参与模型制图。

目前常用的环境变量主要包括以下几点。

4.3.2.1 气候变量的表征与数据选取

气候因素在较大范围内主要考虑大气候，通常选择近5~10年的年降水、大于0℃或10℃积温（或太阳辐射量）等因子，并根据制图比例尺选用，或利用气象站点生成相应像素分辨率的气象因子栅格数据，相应方法参考气象数据的有关要求。

而在较小的空间范围内，气候因素对土壤的影响相对均一，可以忽略。相比之下，小范围内的地

形地貌信息可体现小气候对土壤的影响。

4.3.2.2 母质变量的表征与数据处理

土壤母质是土壤形成的物质基础，通常直接获得母质信息非常困难，实际制图中，常以地质图或地貌图来代替土壤母质分布图，这些地图上的信息通常为矢量化表达的地质类型。也可以从分级到土种的大比例尺土壤图中，通过土属或土种名称中的母质信息提取。

4.3.2.3 地形地貌变量的表征与数据处理

地形因素是最常用的环境变量，主要包括描述地形特征的地形属性和描述地貌部位信息的指标。地形属性可利用数字高程模型（digital elevation model, DEM）栅格数据提取：高程、坡度、坡向、平面曲率、剖面曲率、地形湿度指数、与河流的距离、与山脊的距离等，可通过 GIS 软件计算获得。地貌部位通常用坡位表达，可用于小流域土壤属性的空间分布推测。也可通过基于相似度的模糊推理方法，通过计算坡面上任一位置与各类坡位的典型位置在属性域与空间域上的相似度，对坡位在空间上的渐变信息进行定量描述。获得研究区中每一类坡位的空间渐变图，作为土壤制图的环境协变量。

其中，地形湿度指数的计算公式为：

$$TWI = \ln \left(\frac{\alpha}{\tan \beta} \right) \quad (1)$$

式中， α 为垂直于水流方向的汇流面积，单位为 m^2 ， β 为坡度（弧度）。

4.3.2.4 植被变量的表征与数据处理

定量的植被状况空间信息主要通过遥感影像数据的计算获取植被指数和生物物理参数，包括归一化植被指数（NDVI）、叶面积指数（LAI）、郁闭度（CC）等。其中，NDVI 是土地覆盖植被状况应用最广的一种遥感指标，能够检测植被生长状态、植被覆盖度和消除部分辐射误差等，定义为近红外通道与可见光通道反射率之差与之和的商。其计算公式为：

$$NDVI = (NIR - VIS) / (NIR + VIS) \quad (2)$$

式中， NIR 为近红外波段的反射值； VIS 为红光波段的反射值。

NDVI 的取值范围为 -1 和 +1 之间：若 $NDVI < 0$ ，表示地面覆盖着云、水、雪等，对太阳辐射中的可见光反射率较高；若 $NDVI = 0$ ，表示地表裸露的岩石或戈壁等处；若 $NDVI > 0$ ，则表示地表有植被覆盖，且植被覆盖密度越大，其值越高。

获取植被信息的遥感影像与调查时间同年同期为最佳。考虑气象条件对高质量影像获取的影响，可选最近 5 年与调查日期相近的影像。

4.3.2.5 土地利用变量的表征与数据处理

土地利用方式也是影响土壤养分分布的重要因素。但土地利用方式为类别变量，不能直接用于回归分析，可采用两种方法为其赋值引入回归方程。

(1) 哑变量方法。是应用比较普遍类别变量处理方式，以 0 和 1 进行赋值，表示不同的类别，方法如下：对 $n+1$ 个土地利用方式，定义 n 个哑变量（ $X_{81}, X_{82}, \dots, X_{8n}$ 。注：“8”为第 8 个环境变量），以哑变量的 0 和 1 组合表示 $n+1$ 个土地利用方式。

(2) 算术平均值变换。算术平均值变换是用类别自变量与定量因变量的关系建立起自变量各水平与定量结果变量之间的数量关系，以不同土地利用方式下定量因变量的算术平均值（如面积百分比）代替该土地利用方式。

4.3.2.6 其他变量的表征与数据处理

地表动态反馈：在平原或地形平缓的地区，采用地表动态反馈模式来解决基于土壤—景观关系的制图方法推测平缓区土壤空间分布。将太阳辐射作为对地表的输入，捕捉 1 天内地表热状态的动态反馈特征，利用时序遥感数据（如 MODIS，每日过境）获得陆地表面发生的动态变化作为平缓区土壤制图的环境变量。

在平原或地形平缓的地区，可以将温度植被干旱指数（TVDI）作为表示土壤湿度变化的环境变量，来获取土壤湿度变化与土壤质地等土壤属性相关性，进一步来推测土壤质地等属性。

$$TVDI = \frac{T_s - T_{\min}}{T_{\max} - T_{\min}} \quad (3)$$

式中, T_s 为地表温度, T_{\max} 为 NDVI 对应的最高地表温度, 即干边:

$$T_{\max} = a + b \times NDVI \quad (4)$$

T_{\min} 为 NDVI 对应的最小地表温度, 即湿边:

$$T_{\min} = c + d \times NDVI \quad (5)$$

式中, a 、 b 、 c 、 d 分别是干边和湿边线性拟合方程的系数。 T_s 越接近干边 TVDI 越大, 表示土壤干旱情况越严重; 相反 T_s 越接近湿边, TVDI 越小, 说明土壤含水量越高。因此, TVDI 与土壤含水量的相关性, 可以反映干旱情况。TVDI 的取值为 0~1, TVDI 越大, 表明该区干旱越严重。

4.3.2.7 已有土壤图数据处理与知识提取

通过两种方法从土壤图中提取隐含的土壤与环境关系, 主要用于: 一是在土壤分布范围内构建环境协变量的频率分布曲线, 以此来代表土壤与环境关系; 二是基于已有土壤图提取训练样点, 然后使用统计或机器学习算法归纳出样点所代表的土壤与环境关系。

(1) 从已有土壤图中提取土壤与环境频率曲线的方式。将已有土壤图与土壤形成具有协同变化关系的变量进行叠加, 针对每一个环境协变量, 为各制图单元构建环境频率曲线, 用于代表土壤与环境关系。即对每种土壤类型所对应的环境协变量, 对其像元数直方图拟合得到环境频率曲线。在空间推测时, 通过计算待推测像元的环境值在各制图单元所对应的土壤与环境关系曲线上的频率值, 来代表该像元在该环境协变量上隶属于各制图单元的程度(隶属度)。最后通过对所有环境协变量上的隶属度进行综合得到该像元对各制图单元的相似度, 选择最高相似度的类型作为最终的推测结果, 完成对已有土壤图的更新。

(2) 从已有土壤图中提取训练样点的方式。该类方法首先按照一定的方式从土壤图中选择训练样点, 利用统计或机器学习算法根据选择的训练样点和研究区的环境协变量获取土壤与环境关系, 通常采用线性或非线性的形式表达这一关系。从土壤图中选择训练样点的方式主要有 3 种: ①在各制图单元中随机选择相同数量的样点; ②在各土壤多边形中随机选择相同数量的样点; ③各制图单元中训练样点的数量按其研究区所占的面积比例选择。然后使用统计或机器学习模型算法归纳出样点所代表的土壤与环境关系。

5 制图思路

5.1 土壤属性制图

5.1.1 选择较优制图模型

划分典型地貌区, 每区对一类土壤指标, 推荐 2~3 个制图模型。操作时, 遵循方法相对成熟、精度较优的原则, 从推荐模型中筛选。经模型精度比较后, 也可采用其他精度更高、应用相对成熟的模型进行制图, 但必须考虑相邻地区的接边。

5.1.2 县级、省级、国家级成果图逐级汇总

县级成果图(大比例尺), 建议采取同地貌区、同一模型、多县统一制图的方式完成, 也可采用单县逐一制图。省级和国家级制图(分别为中、低比例尺), 建议主要以制图综合的方法完成。

原则上以典型地貌区为单位, 样点数量和密度达到模型要求, 应在该地貌区范围内进行大比例尺精度的制图。这既减少逐县制图的工作量, 也尽量减少了分县接边的不确定性。制图成果数据作为县级大比例尺空间图数据库。功能性评价图在县级精度上完成评价制图。

省级、国家级成果图在县级成果图基础上, 通过 GIS 栅格数据精度转换的功能, 以省级、国家级比例尺对应像素精度进行转换。像素精度转换主要体现了制图综合中“图斑合并”“图斑取舍”“轮廓简化”的方法, 取像素面积最大值作为转换后像素属性值。逐级汇总保证了县级与省级成果图

面的趋势一致性。特殊区域的处理见 5.1.3。

对样点相对少的区域，经检验后采用适合的精度（比例尺）进行制图。

土壤属性及评价结果的统计，须以最高精度，即县级制图成果数值为基准。

5.1.3 特殊区域的制图综合

土壤属性和功能性评价图中，对面积小但有特殊指示作用的区域，如敏感元素属性极高区，或评价图中的极不适宜区、严重退化区等，如需在中、小比例尺图中予以体现，这类区域需根据原区域长度和面积，事先计算可显示的长度和面积大小，通过模型处理，单独进行像素转换，并将其替换到制图综合后数据图层中。制图综合后数据图层，不宜作为各类上报的数据统计基础图层。

5.2 土壤专题制图

5.2.1 建立评价指标，制作专题图

根据需评价的专题，依照国家或行业标准、规范或本领域普遍认可的研究方法，确定相关土壤属性因子及其权重，建立评价指标体系。以评价指标完成相关土壤属性制图，通过空间图层结合评价指标权重计算，确定评价单元等级，完成制图。

5.2.2 县级、省级、国家级成果图逐级汇总

同 5.1.2。

5.2.3 特殊区域的制图综合

同 5.1.3。

6 土壤属性图与专题图制图要求

6.1 土壤属性制图

数字土壤属性制图包括 4 个环节：样点数据的获取、环境变量的生成、制图模型或方法的建立、土壤制图结果的产生及验证。样点数据的获取、各种环境变量涉及的指标在“4 制图数据准备及要求”中已有介绍，本节从制图流程做说明。

6.1.1 数据制备

6.1.1.1 数据制备

GIS 软件可以完成样点数据（目标变量）和大部分环境变量栅格数据的制备；源自遥感影像的环境变量数据，需要采用专业遥感图像处理软件，全部数据最终统一到模型制图的 GIS 软件数据格式。

其中，环境变量的多种地形要素，可通过数字高程模型（DEM）栅格数据，在 GIS 软件相应功能模块中提取：高程、坡度、坡向、平面曲率、剖面曲率、地形湿度指数等分别作为一个环境变量数据参与相关模型的分析。其他非遥感影像数据也均可在 GIS 软件中进行栅格数据的制作和转换。利用专业遥感图像处理软件处理遥感影像，包括图幅镶嵌、计算，最终提取植被指数、温度植被干旱指数等遥感相关数据。

6.1.1.2 样点数据检验

相关性分析、正态分布检验采用常规统计软件，如 SPSS、SAS 及开源软件 R。普通商业 GIS 软件可进行半方差函数分析等数据检验，也可采用其他软件如 GS+。

6.1.1.3 数据坐标系

成果图统一采用 2000 国家大地坐标系，与国土三调成果一致。相关数据图层需以投影坐标系的方式进行运算和制图，不得以经纬度坐标进行制图。

制图前将各图层数据统一到 2000 国家大地坐标系后制图。也可先统一到一个坐标系（如 1954 年北京坐标系或 1980 年西安坐标系）进行制图，最后将成果图按要求转换到 2000 国家大地坐标系。

6.1.1.4 制图模型相关软件

一般商业 GIS 软件均具有多种地统计模型，具有自动构建训练集和验证集，以及计算均方根误差和决定系数的功能。可调用其相应模型的功能模块，按照操作说明进行预测制图，并检验其模型精度。

随机森林等机器学习模型，可采用开源软件 R 或自行编程完成模型预测，将预测结果值导入 GIS 软件，进行空间数据图层的制作而最终成图。

6.1.2 主要环境变量的选择

6.1.2.1 不同尺度的主要环境变量

基于不同制图尺度与地形分区的土壤属性图类型，选择不同的环境变量，见表 3。

表 3 不同尺度的备选环境变量

主导因素	制图尺度	
	大尺度（小比例尺）	小尺度（大比例尺）
气候	气候区、年均温、年降水、积温或太阳辐射量等	
生物	植被类型、植被物候期	归一化植被指数（NDVI）等植被指数、叶面积指数（LAI）、林冠郁闭度（CC）等
母质	母岩类型	母质、土壤类型、地表动态反馈、温度植被干旱指数（TVPI）等
地形	地形地貌	高程、坡度、坡向、曲率、地形湿度指数（TWI）、坡位等各种地形因子
人为因素	土地利用	土地利用

6.1.2.2 目标变量数据分析与环境变量筛选

(1) 符合数学假设检验。对样点数据（目标变量）进行正态分布检验。

(2) 半方差函数（空间自相关）分析。采用地统计模型方法（包括克里格插值及其衍生方法）之前，需通过半方差函数确定土壤属性具有空间自相关性。其表达式为：

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i + h) - Z(x_i)]^2 \quad (6)$$

式中， $\gamma(h)$ 表示间距为 h 的点对之间的平均半方差， $N(h)$ 表示距离为 h 时的所有点对数目， $Z(x_i)$ 则表示点 x_i 的观测值。比较常用的半方差拟合模型主要包括了高斯模型、指数模型、球状模型和线性模型。

块金效应反映了随机因素对变量空间自相关性的影响程度，其适用性见表 4。块金效应小于 75%，空间自相关距离（样点距离）在变程内，表明样点属性具有空间自相关性，可以进行地统计模型制图。

表 4 块金效应与空间自相关对应关系

块金效应（系数）	空间自相关
<25%	强烈
25%~75%	中等
≥75%	弱性

(3) 环境变量筛选—相关性分析。对利用土壤属性与环境变量关系的制图方法，进行土壤属性与环境变量之间的相关性分析，保证两者之间存在显著相关性，以判断哪些环境变量可以保留在模型

中，并去除环境变量之间的共线性。

6.1.3 不同地形分区的推荐模型与备选模型

制图模型的选择基于一定的样点密度，当小尺度范围内样点密度较高时，相对简单的模型也可达到符合要求的精度。样点密度较低，特别是地形复杂地区，借助多种环境变量的模型则可以提高制图精度（表5）。

表5 土壤制图地形分区

序号	地形	分区规则	可选环境变量	推荐模型	备选模型 ^①
1	平原	海拔：≤200 m 地貌：宽广平坦，起伏很小	主因素：气候、植被、土地利用 次因素：母质、地形	普通克里格、地理加权回归克里格	随机森林，反距离加权克里格
2	丘陵	海拔：>200 m，且≤500 m 地貌：高低起伏，坡度较缓，由连绵不断的低矮山丘组成	主因素：地形、母质、植被 次因素：气候、土地利用	地理加权回归克里格、随机森林	回归克里格、其他机器学习方法
3	山地	海拔：>500 m 地貌：地表形态奇特多样，或相互重叠，犬牙交错，或彼此平行，绵延千里	主因素：地形、气候、植被 次因素：母质、土地利用	地理加权回归克里格、随机森林	回归克里格、其他机器学习方法
4	高原	海拔：>1 000 m 地貌：面积较大，顶面起伏较小，周围形成陡坡的高地	主因素：气候、植被、地形 次因素：母质、土地利用	普通克里格、地理加权回归克里格、随机森林	回归克里格、其他机器学习方法
5	盆地	地貌：四周高（山地或高原）、中部低（平原或丘陵）的盆状地形	主因素：气候、植被、土地利用、地形 次因素：母质	地理加权回归克里格、随机森林	其他机器学习方法

注：①其他机器学习方法：分类回归树、卷积神经网络模型。

模型选用原则如下。

(1) 研究区域平稳：推荐普通克里格模型。

(2) 环境变量少，主导因素确定（如平原和地势平缓区的土壤含盐量）：推荐随机森林、地理加权回归模型。

(3) 环境变量复杂，研究区域地形地貌复杂，推荐随机森林、地理加权回归克里格，备选其他机器学习方法，如径向基函数神经网络、分类回归树。

(4) 成分数据（如土壤机械/颗粒组成），可采用随机森林、成分克里格、普通克里格模型及模糊推理模型。

主要数字土壤制图方法介绍详见附录1。

6.1.4 制图模型的训练和评估

6.1.4.1 构建训练集和验证集

评估所得空间分布图的精度可使用独立验证点，验证样点的获取有两种途径。第一种途径是把已有样点集按照4:1的比例随机划分为训练集和验证集两个部分，其中训练集（80%）用于构建预测模型，验证集（20%）用于检验模型的预测精度。第二种途径是在野外采集验证样点，通常使用随机采样方法采集。独立验证样点数量与制图区域的面积有关，由于独立验证样点的生成或采集具有随机性，根据统计学的大数定律，独立验证样点数量应不少于50个。

6.1.4.2 模型评估

土壤属性数字制图模型精度的验证指标主要有均方根误差（RMSE）、决定系数 R^2 等。其中，对一般面积大小的农业区县，表层样点数平均达1 000个以上，一般决定系数 R^2 应大于0.5以上；均方根误差数值越小越好。

$$RMSE = \sqrt{\frac{\sum_{i=1}^n [\hat{Z}(x_i) - Z(x_i)]^2}{n}} \quad (7)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n [\hat{Z}(x_i) - \bar{Z}(x_i)]^2}{\sum_{i=1}^n [Z(x_i) - \bar{Z}(x_i)]^2} \quad (8)$$

式中， n 为验证集中的样点个数， \hat{Z}_{x_i} 为预测值， $Z(x_i)$ 为实测值， \bar{Z}_{x_i} 为实测值的平均值。

6.2 土壤专题图

第三次全国土壤普查形成的专题调查评价图主要有：耕地质量等级图、酸化土壤分布图、盐碱化土壤分布图、黑土资源退化分布图、土壤碳库与养分库贮量图、土壤利用适宜性分布图，特色农产品生产区域土壤专题调查图等列入本次普查成果清单中的专题图件。指标体系由相关规范确定，本规范主要规定制图过程和方法。

在完成土壤类型和土壤属性成果图基础上，依照评价指标体系，首先提取作为评价指标的土壤属性图层数据，参照 6.1 模型的方法，或者专题图规范规定的单一指标制图方法，完成其他相关指标的属性制图；然后根据各指标权重，通过 GIS 软件，进行各属性图层数据的空间计算，获得各评价单元（或像素）的评价指数；最后依照指标体系的评价标准，确定评价单元的评价等级，完成制图。

6.3 土壤属性制图相邻区处理

土壤属性多为数值，在空间上是连续的，并无界线截然分开。采用数字土壤模型制图方法进行土壤属性制图，输出结果为数值型的栅格地图。每个栅格像素为一个土壤属性值，像素之间不存在接边问题。但由于不同地貌区域最终采用的适宜模型可能不同，特别是在本区域制图而在区外无样点时，对区界处的预测属性值精确度会降低，可能会造成两区相邻处像素属性值出现相差较大的情况。为避免、减少此类情况，模型制图时，可增加域外样点，即区域制图时，将与本区相邻地区的区界样点补充进训练集进行模型制图，然后以区界裁切获得本区成果图。

6.4 土壤制图结果的验证评价

除模型评估外，对于依靠单个土壤样点对待推测点的代表性实现空间推测的制图方法，也可利用推测不确定性指标对土壤属性制图的结果进行评价，用于指示推测结果的可靠程度。该图为研究区的每个像元给出不确定性值，以体现推测不确定性的空间变化。由于不确定信息与精度具有相关关系，因此可以通过推测不确定性间接指示制图精度。可根据所能承受的误差水平（ $RMSE$ 等）来选择合适的的不确定性阈值，从而满足制图精度要求。

6.5 土壤制图结果的面积统计

对成果图计算各类型的面。一般商业 GIS 软件均有此功能，即根据每个等级像素数量乘以每个像素代表的面积。

7 制图比例尺/分辨率与数据质量要求

7.1 制图比例尺（分辨率）

7.1.1 国家级、省级、县级的制图比例尺（分辨率）及上图面积

普查成果图，一般按照国家基本比例尺成图。国家级成果图比例尺 1:100 万~1:400 万，省级成果图比例尺一般为 1:25 万~1:50 万，县级成果图比例尺一般为 1:1 万~1:5 万。省级和县级也

可根据本行政区域范围大小,选择适当比例尺成图。以数字模型方法制作的成果图为栅格图,需达到相应的像素分辨率。与比例尺对应的栅格数据像素分辨率详见表 1。

不同比例尺上图面积可参照第二次土壤普查中关于各比例尺土壤图上图面积的规定,详见表 6。

表 6 各种比例尺土壤图的最小面积

制图比例尺	土壤图的最小面积			
	可达到的		适当的	
	在图上	在实地中	在图上	在实地中
1 : 2 000	所有比例尺当面积为长形,其长轴为 0.2 cm (2 mm×10 mm) 时或当面积为圆形,直径为 5 mm	80 m ²	1 cm ²	400 m ² (0.6 亩)
1 : 5 000		500 m ²	1 cm ²	2 500 m ² (3.75 亩)
1 : 1 万		0.2 hm ² (3 亩)	0.5 cm ²	0.5 hm ² (7.5 亩)
1 : 2 万		1.25 hm ² (18.75 亩)	0.5 cm ²	2 hm ² (30 亩)
1 : 5 万		5 hm ² (75 亩)	0.5 cm ²	12.5 hm ² (187.5 亩)

注:引自《中国土壤普查技术》。

7.1.2 环境变量栅格数据分辨率

数字模型制图过程中,需要制备地形参数、土地利用、植被等环境变量栅格数据,其像素分辨率应不低于成图比例尺对应的像素分辨率。应根据土壤采样点密度、运算速度、计算机容量,选择可满足精度要求的像素分辨率。

7.2 制图成果数据质量要求

7.2.1 数据精度、数据结构、元数据及拓扑要求

数字模型方法制作的成果图,交汇格式为栅格数据,应满足以下要求。

- (1) 栅格数据像素分辨率,需符合表 1 要求。
- (2) 数据格式,以普查统一要求的数据格式提交,中英文图层名与数据内容相符合。如中国土壤有效磷含量图(或 CN_AP)、河北省土壤有效磷含量图(或 Hebei_AP)等。
- (3) 数据字典,须包括图层名、字段名、字段释义、字段类型、字段长度等,属性字段名称、类型、长度、小数位数符合《第三次全国土壤普查数据库规范》要求。
- (4) 元数据,须包括土壤普查时间、制图时间、模型制图方法、模型精度(均方根预测误差 *RMSE*)、区域范围、元素形态、计量单位、大地坐标系、投影、分辨率、制图单位、制图负责人等信息。

元数据信息符合《土壤科学数据元数据》(GB/T 32739—2016)的相关规定。

矢量格式制作的面状成果图,除符合(2)、(3)、(4)外,还应满足以下要求。

- (1) 关键界线和面积的容差:如成果图含有土地利用界线时,对照已有土地利用界线,界线移位容差为 0.000 1 m。成果图以行政区域为单元的,其成果图面积与县域面积一致。
- (2) 无拓扑错误:①同一图层内不存在面与面重叠,包括完全重叠与部分重叠(即面相交),容差为 0.000 1 m,同一面层内不同面要素之间不存在缝隙,面裂隙容差为 0.000 1 m;②同一图层内不同要素间线要素不存在重叠或与自身重叠;③同一图层内线要素不存在自相交;④同一图层内线要素不存在悬挂点;⑤同一图层内线要素不存在伪节点;⑥面层内不存在不规则图斑。⑦面层内不存在碎片多边形;⑧面层内要素不允许存在组合图斑;⑨同一线层内不存在碎线(长度小于 0.2 m);⑩图形节点密度符合规范要求,不能过于稀疏、稠密(平均节点密度大于 70 m,或小于 1 m);⑪图形不存在面自相交、环方向错误等不符合入库要求的错误。

7.2.2 区域不合理性专业检查

对成果图,在一定区域范围内,某些土壤属性如出现明显不同于周边的情况,应说明原因。

7.2.3 投影与坐标系

平面坐标系：采用 2000 国家大地坐标系。

高程系统：采用 1985 国家高程基准。

投影方式：大于 1:50 万比例尺（不含 1:50 万），采用高斯—克吕格投影，大于 1:1 万比例尺按 3°分带，1:2.5 万~1:50 万比例尺按 6°分带。小于 1:100 万比例尺，采用正轴等角割圆锥投影。1:50 万比例尺，一般采用 6°分带的高斯—克吕格投影，内蒙古等经度分带跨度大的西部省，可采用正轴等角割圆锥投影。

7.2.4 图分幅

图的分幅应符合《第三次全国土壤普查数据库规范》要求。

8 专题图表达与质量要求

8.1 专题图编制方案设计

图件历来是土壤普查的重要成果。在编制单位、图名、普查时间等制图内容，文字内容、位置、字体大小等，各地方必须采用全国统一方案。

编制内容主要包括：图名、编制单位、制图单位及制图人员、制图时间、土壤调查时间、坐标系、地图投影、比例尺。其他说明包括地理要素所采用的地形图比例尺和时间。这些内容在图廓外的位置应平衡美观。成果图模板见附录 1。

8.2 专题内容的表达

8.2.1 专题制图表达

土壤专题图采用质底法。

土壤属性图，图面表达包括属性配色；如为属性分级图，图面表达还包括分级编号。原则上一个土壤属性对应一个色调，从颜色上区分土壤属性类别。此外，以颜色深浅表示含量大小。

其他专题图，颜色的选择应避免已有标准指定的土壤属性颜色，选用新的色调及符号。

8.2.2 图例要求

对于土壤属性分级图，专题图例由计量单位、分级代码、色块、分级的养分含量范围和测试分析方法 5 部分组成。对于土壤属性栅格渐变图图例，由计量单位、养分含量上下限、渐变色带和测试分析方法 4 部分组成。各属性含量分级由全国土壤普查办统一制定，省级分级可在国家分级范围内，做更细划分，但不得跨级。

15 种土壤养分图图例要求细则依照《1:25 000~1:500 000 土壤养分图用色与图例规范》(GB/T41475—2022) 制图，其他比例尺图可参照执行。对耕地质量等级、特色农产品生产区域等其他专题图，专题图例可参考上述国标自行确定，其他图面信息参照附图。

8.3 基础地理要素的选取和表达

地理底图是专业地图的骨架，根据专题图特点，对地理要素进行必要的选取，保留具有体现土壤类型或属性特征的要素，舍去干扰专题特征的地理内容，有利于突出土壤专业内容。

8.3.1 要素选取

根据成图比例尺，选择相对应或更小比例尺的地理要素。

水系：适当选取以反映河网密度和结构。包括河流（常年河、时令河、消失河等）、湖泊、水库、坎儿井、水渠、运河、咸水湖。

居民地（点）：根据比例尺，选择相应行政级别的居民地或居民点上图。小比例尺图，原则上选取县以上级别居民地/点，中等比例尺原则上可选择到乡镇级，大比例尺，可选择到村级，并根据居

民地密度适当取舍。

交通：原则上铁路均可上图；公路则依据比例尺大小，相应地选取国家级、省级和县乡级公路。

境界：小比例尺显示国界和省界；大中比例尺显示省界和县界，地处边疆省区，需显示国界。

8.3.2 制图表达要求

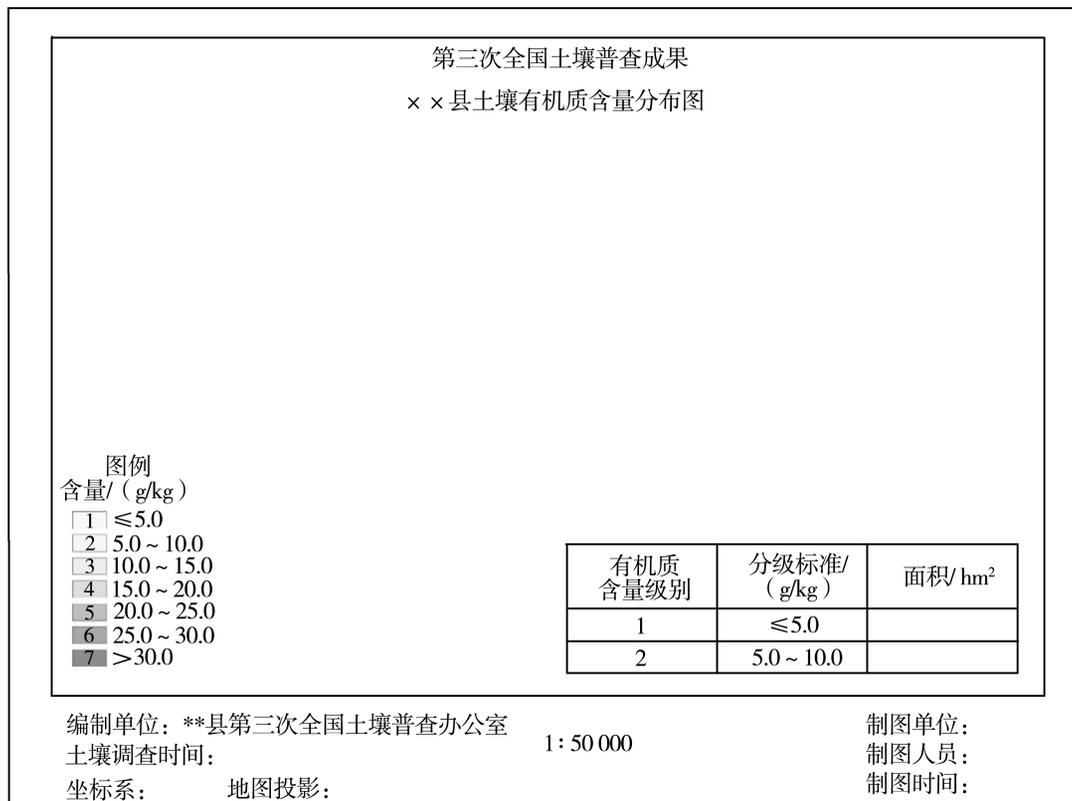
公共地理信息通用地图符号采用相应国家标准制作。

9 制图设备与场所要求

鉴于制图所需大量高精度图层数据，制图单位需设定专用场所，准备相应级别配置的专用计算机，仅内网相连，单机设备不得与互联网链接；拆除其他拷贝接口，仅留一个接口用于数据拷贝，并做好使用数据的登记管理。制图单位制定保密规定，相关人员签订保密协议。

附录 1 第三次全国土壤普查成果图模板

第三次全国土壤普查成果图模板见图附录 1-1。



图附录 1-1 第三次土壤普查成果图模板

附录 2 数字土壤制图的理论基础与主要方法

1 数字土壤制图的理论基础

数字土壤制图反映的是土壤的空间分布特征和规律，主要基于两个理论基础，第一是土壤成土因子学说，即某一地区的土壤，是土壤母质在一定水热条件和生物因素、人为因素作用下，经过一系列物理、化学和生物化学过程所形成，其土壤的空间分布与环境因子的空间分布具有协同变化的关系；特定的环境条件组合形成特定的土壤，相似的环境因子组合下分布着相似的土壤，环境组合越相似，其对应的土壤越相似。第二个理论基础是地理学第一定律，即环境因子的连续性造成土壤空间分布呈现连续渐变的特征，表现为空间上距离越近的点位土壤属性越相似；相邻两种土壤类型间在空间上往

往没有明显的界线，而是呈现出一个过渡区。

2 数字土壤制图的主要方法

目前，数字土壤制图领域比较常用的方法可分为五类：地统计学方法、确定性插值、数理统计、机器学习和模糊推理方法。

2.1 地统计学方法

地统计学方法的核心是区域化变量理论，基本工具为变异函数，利用样本信息进行空间插值，最后获得变量的空间分布。该方法于 20 世纪 70 年代末至 80 年代初被引入到土壤学的研究工作中，其中以克里格插值及其衍生方法最具代表性。

普通克里格 (ordinary Kriging, OK) 由于简单易操作被广泛应用于早期的土壤物理和化学性质以及土壤养分的空间制图中。OK 对变量的空间自相关进行了假设，因此适用于较均一、土壤属性变化不强烈的环境，在小尺度和均质景观区域取得了较好的效果，对区域变异大、面积覆盖广的区域，OK 的制图精度不太理想。泛克里格 (universal Kriging, UK) 在 OK 的基础上，引入趋势面方程分离土壤属性与空间位置的趋势项来消除不平稳性，在一定程度上可以减小 OK 的局限性。但这两种方法均忽视了土壤属性与环境要素之间的关系。协同克里格 (co-Kriging, CK) 和回归克里格 (regression Kriging, RK) 均利用所预测土壤属性与环境辅助变量之间的相关性来提高预测精度，不同的是，CK 将预测变量的空间自相关性及其与辅助变量间的交互相关性结合起来用于无偏最优估计；RK 则将常规统计学与地统计相结合，先用回归模型拟合土壤属性与辅助变量之间的关系，然后对回归残差应用克里格法插值，最后将回归预测结果与残差插值结果结合起来得到最终预测结果。成分克里格 (compositional Kriging) 是一种专门针对成分数据 (如土壤机械组成) 的空间插值方法，需满足非负、定和、预测误差最小和无偏估计；成分克里格可将成分数据作为一个整体考虑，而非对单个成分插值制图，避免插值后出现同一点位各个成分之和超过定和或负值，例如土壤机械组成砂、粉、黏含量之和超过 100%。

关于 CK 和 RK 两种方法的精度对比，有研究表明，结合地形属性的 RK 预测精度高于 CK。当辅助变量较多时，RK 的预测精度高于 CK。除此之外，较新近的经验贝叶斯克里格 (empirical Bayesian Kriging, EBK) 也被证明是一种可对非均质景观区域进行空间预测的插值方法，它在 OK 的基础上发展而来，可以通过构造子集、建立局部模型对非平稳数据进行空间插值。EBK 由于考虑了空间异质性，可以更好地揭示土壤属性的空间变异结构。但目前该方法应用较少。

尽管地统计学方法被证明是一种易操作且结果较为可靠的数字土壤制图的方法，但其要求数据满足地统计学相关假设，给实际应用带来一定困难。此外，插值过程需要利用预测变量的空间自相关性，而地学现象的复杂性和独特性使得一个地区的模型很难直接应用到其他地区。

2.2 机器学习模型

机器学习模型利用机器学习与数据挖掘方法，提取土壤属性与环境变量之间的关系用来预测土壤属性的空间分布。与地统计和梳理统计方法相比，机器学习模型可以解决土壤属性与环境变量的非线性问题，且对数据分布没有要求，因此被越来越多地应用于数字制图领域。常用的机器学习模型包括人工神经网络 (artificial neural networks, ANN)、分类与回归树 (classification and regression tree, CART)、随机森林 (random forest, RF) 等。

ANN 可以模拟人脑神经网络对信息进行处理工作，建立某种简单模型，按不同的连接方式组成不同的网络，对连续型变量和类别变量均有很好的预测效果。研究表明，BP-ANN 比 OK 法更适用于大区域小样本下的土壤属性预测。虽然机器学习模型有易过度拟合、不易解释等不足，但该类方法能够有效地解决土壤属性与环境因子之间的非线性问题，且在大范围区域表现良好，已经逐渐成为土

壤数字制图的主流方法。

2.3 专家知识模型

专家知识模型是基于地理相似性原理的土壤—景观推理模型。一般首先将土壤与环境条件关系的知识表达为隶属度函数，然后根据多个因子的隶属度函数来综合评价某点的土壤属于某种土壤类型的隶属度值，因此某点的土壤可与多个土壤类型具有隶属度（相似度），根据这些隶属度可确定该点的土壤的类型和属性，隶属度的利用可以使土壤空间变化的连续性得到较好的体现。即利用样点与待推测点之间的地理相似性来刻画单个样点对待推测点的代表程度，然后将代表程度作为权重参与推测待推测点的土壤目标变量的值。同时，这些代表程度还体现了样点对推测区域的推理能力，代表性高则表示现有样点能较好地推测待推测点的地理变量值，代表性低则说明现有样点对待推测点的推测具有较高的不确定性，利用代表程度度量的推测不确定性可有效指示推测结果的精度。

2.4 确定性插值法

包括反距离加权法、最邻近法和样条插值法等，是以区域内部的相似性或以平滑度为基础，由已知样点来创建表面。其使用环境与普通克里格相近。

2.5 数理统计方法

数理统计方法通过探索已知样点的土壤属性与环境辅助变量之间的统计关系并建立函数表达式，用来预测土壤属性的空间分布，完成空间制图。常用于数字土壤制图的数理统计方法包括多元线性回归（multiple linear regression, MLR）、广义多元线性回归（generalized multiple linear regression, GMLR）、判别分析（discriminant analysis, DA）等。数理统计方法简单直观，且能表达土壤属性与环境因子之间的相关关系。但此类方法均假设土壤属性与辅助变量是线性相关，且需要较大的样本量来提取这种线性关系，因此在小尺度区域预测精度较高。大尺度区域，土壤属性与环境变量不一定是简单的线性关系。

综上所述，基于土壤属性与环境变量要素相关性的方法是现有数字土壤制图方法中应用最广泛的方法，其中随机森林方法在数据挖掘方法中应用越来越普遍。基于空间自相关推测土壤属性空间分布的方法应用也很广泛，克里格插值法应用最多，这类方法要求样本密度高，且需要样本能很好捕捉土壤属性的空间自相关特征。由于基于要素相关性和空间自相关的方法，需同时考虑空间自相关性和环境变量相关性，一定程度上能提高土壤推测的精度，缺点是对样本数量与分布要求较高，需满足二阶平稳的假设并要求要素相关性稳定，往往在实践应用中很难得达到。基于地理相似性的土壤—景观推理模型是基于知识的制图方法的代表方法，但制图效果依赖于单个样点的可靠性。