

基于 Web 的中国土种数据库^①

施建平, 宋 歌

(中国科学院南京土壤研究所, 南京 210008)

摘 要: 来源于第二次全国土壤普查汇总成果的中国土种数据库是目前最全面的全国性土壤数据库。它收录了 20 世纪 80 年代我国主要土种的分布、面积、性状、土地利用情况和生产性能等资料。自 2000 年以来, 中国科学院南京土壤研究所开始建设基于 Web 的中国土种数据库, 构造了一个包括地点与土壤类型的空间分布关系和依据中国土壤发生分类系统的分类层次关系的实体-关系(entity-relationship, E-R)数据模型, 形成了具有 2 473 个土种典型剖面、8 751 个发生层及物理化学性质的数据产品。该数据库可通过互联网广泛应用于土地退化评估、环境影响研究以及土壤碳储量研究, 亦可作为基础数据指导农业生产。

关键词: 土种数据库; 中国; 第二次全国土壤普查; 数据整合

中图分类号: S159.2

土壤是一个国家最重要的自然资源, 是农业发展的物质基础。受复杂的自然条件和深刻的人为因素的影响, 我国土壤类型繁多、千差万别。在土壤学中, 将影响土壤形成的各种自然条件, 归纳为地形、气候、成土母质、植被、成土年龄等五大因素, 称为成土因素。在不同的地区, 各因素具体产生作用的方式和特点不同, 不同因素间还以不同方式互相作用和配合, 从而形成各种各样的土壤。土种是土壤发生分类系统中的基层分类单元, 是在类似的水热条件下, 来自相同或相似景观部位, 具有相对一致的土壤剖面形态、发育层段、理化及生物特性、生产性能均相一致的一组土壤实体。土种的典型剖面是一个土种中心概念与边缘概念十分清晰的土壤个体, 体现了该土种的中心概念, 起到定位、定性、定量的作用, 是评比土壤性状差异与划分土种的依据。

中国土种数据库的数据来源于全国第二次土壤普查汇总成果《中国土种志》^[1]。《中国土种志》经过整理、选择、总结省(市、区)级有关土种资料基础上编写而成, 是迄今为止最全面的全国性的土壤数据资料, 它包括 2 473 个土种典型剖面, 覆盖了 20 世纪 80 年代我国主要土种在土壤发生分类系统中的归属、分布、面积、主要性状、典型剖面、生产性能和理化分析数据。

土壤数据具有时间和空间属性。在时间尺度上,

土壤性质随着自然条件和人类活动影响而改变。在空间尺度上, 土壤具有水平分布特征和垂直分布特征。同时土壤类型的划分与植物分类一样, 对不同类型土壤的定量和规范化分类描述是数据解译、模拟、对比的基础。第二次全国土壤普查数据采集跨度为 1978—1984 年, 可以不考虑时间尺度的变化。因此, 中国土种数据库结构设计需同时考虑空间分布关系和分类关系。

中国土种数据库以土种为单位, 提取了土种归属和分布、地形地貌、土地利用、土壤主要性状、典型剖面理化性质和统计剖面理化性质数据。典型剖面土壤理化性质包括不同发生层的土壤养分(有机质、全氮、全磷、全钾、碱解氮、有效磷、速效钾、pH)、土壤物理特征(颗粒组成、质地、容重、孔隙度)和土壤化学特征(交换性阳离子、阳离子交换量、有效阳离子交换量、碳酸钙)等 28 个参数。参考国外相关土壤数据库关系结构模型^[2-3], 建立地点与土种水平分布关系, 2 473 个土种分布在全国除台湾省外 30 个省 1 642 个县, 建立土种与发生层间垂直分布关系, 2 473 个土种典型剖面具有 8 751 条发生层数据, 平均每个剖面约 3.5 个发生层; 按照中国土壤发生分类(GB/T 17296)^[4], 建立土类、亚类、土种的分类关系。中国土种数据库可按地点和土壤分类进行查询检索, 收录的数据可用于土壤碳储量研究、土壤分类研究或

基金项目: 中国科学院信息化专项科技数据资源整合与共享工程重点数据库项目(XXH12504-1-02)资助。

作者简介: 施建平(1954—), 女, 江苏海门人, 学士, 高级工程师, 主要从事土壤信息管理研究。E-mail: jpshi@issas.ac.cn

作为土壤基础数据指导农业生产。

1 数据库结构设计

参照国内外有关土壤资源数据库的关系结构^[5-8]，建成一个具有空间分布关系和分类层次关系的关系型数据库，可以根据地点和土壤发生分类进行查询检索。

数据库由 9 张表组成：

1) 土种的基本信息表：SOIL_TYPE，包括土壤类型名称、一般性描述、母质、剖面构型、主要特征、生产性能等。

2) 典型剖面景观信息表：PROFILE_LANDSCAPE，一个典型剖面对应一个土种类型，记录典型剖面的采集地点、经纬度、地形地貌、高程、母质、年均温和年降水、自然植被、土地利用等。

3) 描述空间分布信息的 3 张表格：省名表为 LOCATION_NAME，县市名表为 sublocation，县市名与土种之间的关系用 LOCATION_USAGE 表描述。一个省有多个县市，省名表与县市名表是一对多关系。一个县市可能有多个土种，一个土种分布于多个县市，县市名与土种之间是多对多关系。例如：四川省县市名表中“马尔康”(sublocation_id 为 1518)分布两种土壤类型：二黄土(soil_id 为 60241)和马尔康黑毡泥土(soil_id 为 60765)，这需要在 LOCATION_USAGE 表格中分别填写 sublocation_id 和 soil_id，建立一个县市名与不同土种间分布关系(表 1)。

表 1 县市名(四川省马尔康县)与土种分布关系示例
Table 1 The example for the corresponding relationship between a city/county (Maerkang County, Sichuan Province) and several soil local types

location_usage_id	soil_id	sublocation_id
6881	60241	1518
7844	60765	1518

再如：西昌红泥土(soil_id 为 60035)分布于攀枝花(sublocation_id 为 1407)、雅安(sublocation_id 为 1408)和西昌(sublocation_id 为 1409)多个县市，这需要在 LOCATION_USAGE 表格中分别填写 soil_id 和 sublocation_id，建立一个土种分布于不同地点的分布关系(表 2)。

表 2 土种(西昌红泥土)与地点的分布关系示例
Table 2 The example for the corresponding relationship between a soil local type (Xichang-Hongnitu) and several locations

location_usage_id	soil_id	sublocation_id
6521	60035	1407
6522	60035	1408
6523	60035	1409

4) 描述土壤分类信息的 2 张表格：土类表为 SOIL_GROUP，用于记录该土种所属土类的土类名、土类英文名、土纲名、土纲英文名、土类描述、典型剖面图片等信息；亚类表为 SOIL_SUBGROUP，用于记录该土种所属亚类的亚类名、亚类英文名信息。一个土类有多个亚类，土类与亚类是一对多的关系。一个亚类有多个土种，亚类与土种也是一对多的关系。

5) 描述剖面理化性质的 2 张表格：典型剖面理化性质表为 VALUM1，统计剖面理化性质表为 VALUM_Stat，这两张表记录了土壤养分全量、速效养分、物理性质、交换性能等信息。一个土种的垂直剖面中含有多个不同深度的发生层，每个发生层具有不同的理化性质，所以土种与发生层是一对多的垂直分布关系。

中国土种数据库中各表之间的关系结构见图 1。

2 Web 数据发布及查询方式

数据库建成后，采用 VisualDB 可视化关系数据库管理与发布工具对数据库进行 Web 快速发布，并定制面向数据的查询应用^[9]。VisualDB 可视化关系数据库管理与发布工具是一款由中国科学院计算机网络信息中心开发的、可以帮助数据库建设人员简化数据库系统开发工作的软件。VisualDB 具有 Web 应用自动生成能力，在建立了数据库各表之间的关系结构基础上，开发人员只需要通过 VisualDB 配置发布模型，便可得到用于数据管理和数据查询访问的 Web 应用界面，发布后即可为用户所使用。

以按地点查询和按土壤分类查询为例，本文列出福建省漳州市赤红壤土类泥赤土土种数据集的典型样本，以说明数据库的具体查询方式。

按地点查询土种信息时首先选择首页(<http://www.soil.csdb.cn>)的数据集导航，若欲详细了解福建省漳州市土壤类型及养分状况，可查询中国土种数据库→按地点查询。选择福建省漳州市，可以获知全市分布有潴育水稻土和赤红壤两个亚类，包括青底灰泥田、泥赤土、赤黏土、黄赤土和霞潭赤黏土 5 个土种(表 3)。点击土种名“泥赤土”，可获得该土种的分布情况、典型剖面景观。

一个土种的典型剖面有多个发生层，每个发生层有不同的深度和理化性质。土种与发生层及发生层理化性质是一对多的关系。按土壤分类查询得到的泥赤土典型剖面全部发生层的深度和理化性质数据，见表 4。

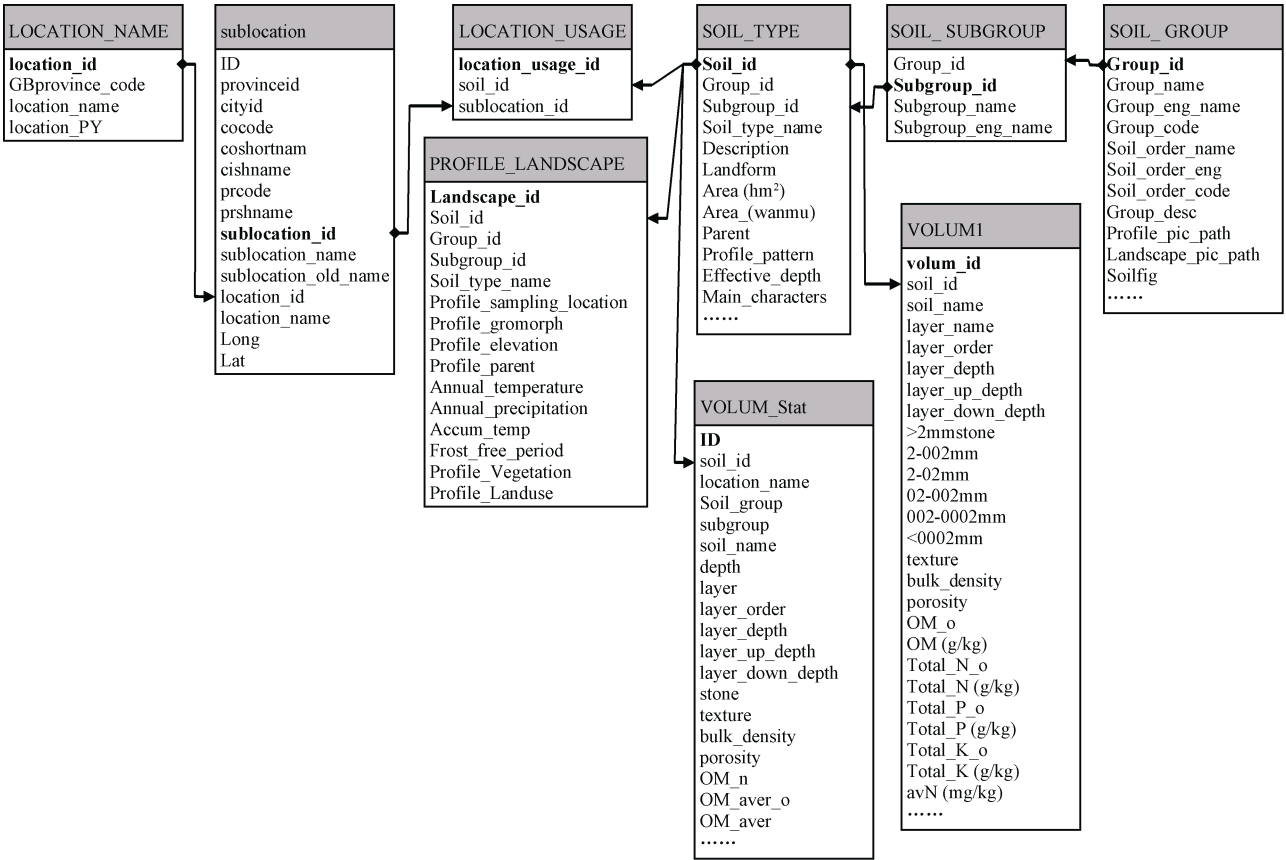


图 1 中国土种数据库关系结构的实体-关系图(E-R 图)

Fig. 1 The entity-relationship diagram (E-R diagram) to describe the data relationship structure in Soil Type Database of China

表 3 中国土种数据库中按地点查询福建省漳州市的结果

Table 3 The results of a location query to Zhangzhou, Fujian Province in Soil Species Database of China

省名	县市名	土类	亚类	土种	分布与地形地貌	面积 (hm ²)
福建	漳州	水稻土	潴育水稻土	青底灰泥田	福建的漳州、三明、建阳三地市的冲积平原，河谷盆地低洼处，或山间垄口地段	40 680
福建	漳州	赤红壤	典型赤红壤	泥赤土	福建省漳州和泉州两市所属县内的 300 m 以下低丘陵地带	28 467
福建	漳州	赤红壤	典型赤红壤	赤黏土	福建省漳州和泉州两市所属县内的 300 m 以下低丘陵地带	57 000
福建	漳州	赤红壤	典型赤红壤	黄赤土	福建省漳州市各县，在 200 m 以下的低丘坡麓地带，或滨海二级阶地上	28 293
福建	漳州	赤红壤	典型赤红壤	霞潭赤黏土	300 m 以下低丘陵地带	164 400

按照中国土壤发生分类查询土种信息，一个土类有多个亚类，一个亚类有多个土种。由此，可按土类→亚类→土种方式逐级查询检索(图 2)。顺次查询赤红壤→典型赤红壤→泥赤土，点击土种名“泥赤土”同样可以查询到该土种的分布情况(表 5)、典型剖面景观以及典型剖面理化性质(表 4)。

3 数据质量控制

中国土种数据库(version 1.0)的建设由中国科学院南京土壤研究所于 2006 年基本完成，数据完整性和一致性经人工抽查基本无误。2011—2013 年对县市名进行了修正，并依据原始资料，重新复核各数据

字段、单位和数值，依据法定计量单位进行了校正。截止目前，中国土种数据库已经建立了数据库设计文档，具备数据字典和关系结构图等完整的文档资料。

中国土种数据库在运行过程中，每年根据用户使用意见，检查数据库内容，修正错误，建立土种数据库维护日志。

3.1 法定计量单位转换

典型剖面理化性质表和统计剖面理化性质表中有机质、全氮、全磷、全钾的原始数据单位为 %。数据录入检查无误后，保留原始数据字段，同时按照法定计量单位增添新的数据字段，将数值乘以 10 以后单位转换为 g/kg。

表 4 中国土种数据库中泥赤土典型剖面主要理化性质
Table 4 The physico-chemical properties of a typical profile of Nichitu in Soil Species Database of China

土种 ID	10003			
土种名称	泥赤土			
发生层次名称 ⁽¹⁾	A	AB	B	BC
层次相对厚度(cm)	8	15	19	58
层最上深度(cm)	0	8	23	42
层最下深度(cm)	8	23	42	100
>2 mm 石砾(%)	— ⁽²⁾	—	—	—
颗粒组成 2~0.02 mm(g/kg)	—	—	—	—
颗粒组成 2~0.2 mm(g/kg)	132.1	69.7	78.1	92.6
颗粒组成 0.2~0.02 mm(g/kg)	263.9	191.7	206.4	235.6
颗粒组成 0.02~0.002 mm(g/kg)	194.0	236.4	242.2	290.7
颗粒组成<0.002 mm(g/kg)	410.1	502.0	473.0	381.8
质地 ⁽³⁾	LC	C	C	LC
容重(g/cm ³)	—	—	—	—
孔隙度(%)	—	—	—	—
有机质(g/kg)	39.8	18.9	15.3	9.1
全氮(g/kg)	1.78	0.97	0.89	0.44
全磷(g/kg)	0.25	0.19	0.25	0.17
全钾(g/kg)	2.6	2.9	2.9	3.9
碱解氮(mg/kg)	—	—	—	—
有效磷(mg/kg)	—	—	—	—
速效钾(mg/kg)	—	—	—	—
缓效钾(mg/kg)	—	—	—	—
交换性氢(cmol/kg(+))	0.35	0.38	0.37	0.48
交换性铝(cmol/kg(+))	4.95	4.42	4.02	3.98
交换性酸总量(cmol/kg(+))	5.70	4.80	4.39	4.46
交换性钙(cmol/kg(+))	—	—	0.82	—
交换性镁(cmol/kg(+))	1.75	0.49	0.00	0.58
交换性钾(cmol/kg(+))	0.13	0.13	0.13	0.12
交换性钠(cmol/kg(+))	—	0.06	—	0.19
交换性盐基总量(cmol/kg(+))	1.70	0.68	0.95	0.89
阳离子交换量(cmol/kg(+))	—	—	—	—
有效阳离子交换量(cmol/kg(+))	7.40	5.48	5.34	5.35

注：(1) 发生层次名称中 A 表示表层，AB 表示亚表层，B 层表示母质特征消失的土层或淀积层，BC 层为从 B 层向母质层(C 层)过渡的土层。(2) “—”表示该土壤理化性质在该土种典型剖面的对应发生层次上没有数据。(3) 质地中 LC 表示壤质黏土，C 表示黏土。



图 2 中国土种数据库中按土壤类型查询的界面
Fig. 2 The soil type query interface in Soil Type Database of China

典型剖面理化性质表中碱解氮、有效磷和速效钾原始数据单位为 ppm,按照法定计量单位增添新的数据字段，数值保持不变，单位转换为 mg/kg。

典型剖面理化性质表中阳离子交换量和交换性阳离子原始数据单位为 me/100g 土,按照法定计量单位增添新的数据字段，数值保持不变，单位转换为 cmol(+)/kg。

土壤颗粒组成按国际制分级，原始数据单位为%，数据录入检查无误后，保留原始数据字段，同时按照法定计量单位增添新的数据字段，将数值乘以 10 以后单位转换为 g/kg；土壤容重单位为 g/cm³，为法定计量单位不需要进行转换。

3.2 采样的深度表示

原始数据发生层深度用相对厚度表示(例如 20)，

表 5 中国土种数据库中按中国土壤发生分类(GSCC)查询泥赤土分布情况的结果
Table 5 The results on Nichitu distribution by a Genetic Soil Classification of China (GSCC) query in Soil Species Database of China

土类	亚类	土种	分布与地形地貌	面积(hm ²)	县市名
赤红壤	典型赤红壤	泥赤土	福建省漳州和泉州两市所属县内的 300 m 以下低丘陵地带	28 466	漳州泉州华安

使用者不易区分其绝对深度或深度范围。参照 ISO 28258^[8],在数据录入检查无误后,增添发生层最上深度和发生层最下深度两个字段(例如 0、20),分别计算录入。

3.3 土壤分类的规范化

对已有数据资源中土壤分类名称进行规范化,是土壤学科领域数据整合的重要内容之一。参照 GB/T 17296^[4],对土类、亚类的命名进行规范化。当 GB/T 17296 中出现土类与亚类命名相同,且该土类包含不止一种亚类时,与土类命名相同的亚类名称前增加“典型”两字,以示与土类命名的区别。例如,原土类为“红壤”,亚类也为“红壤”的亚类名称,改为“典型红壤”。这类亚类名称的更改有 41 处,数据库在亚类表 SOIL_SUBGROUP 中增添了字段 subgroup_GB_name 来记录;同时原有的老亚类名继续保留,以便与历史资料对应。土壤分类和归属按照历史资料分类暂时不变。

在元数据中标识数据资源的土壤类型名称及其在中国土壤分类系统(CSCC)、中国土壤系统分类(CST)之间的参比信息。

3.4 行政地点的规范化和更新

20 世纪 80 年代以后,行政地点名称和归属有较大变化。例如,原上海川沙县,现已归属为浦东新区;重庆市为 1997 年成立,现在重庆市具有土壤普查资

料的区县,历史上归属于四川省。在数据库中共更新了 38 个县市的名称,并增添字段保留原始名称;同时更新了 8 个县的归属。此外,为了查询方便,数据库中增添了与土种分布相关的省、市、县行政区划代码^[10]和市、县近似经纬度。

3.5 数据类型约定

数据的基本类型为文本、数字、备注等。依据原始数据,对数字型数据的精度进行了设置。由于原始的数字型数据有缺失,空值在数据传输中易转换为“0”与实际的数据 0 难以区分,故对于数据缺失较多的数字型字段,改用文本型表示。

4 数据应用

4.1 数据的分布

在应用之前,使用者应了解数据的适用范围、尺度和数据的分布情况。中国土种数据库的数据是在《中国土种志》的基础上构建的,其原始数据没有覆盖全国所有的县市。2 473 个土种典型剖面分布于 1 642 个县市(区)或地理标识区(如太行山、大兴安岭),其中土种数最多是辽宁省,有 159 个土种;最低是北京市,有 11 个土种(图 3)。典型剖面分布密度最大的是上海市,为 36 个/10⁴ km²;分布密度最小的是新疆维吾尔自治区,为 0.6 个/10⁴ km²,全国平均 2.6 个/10⁴ km²(图 4)。

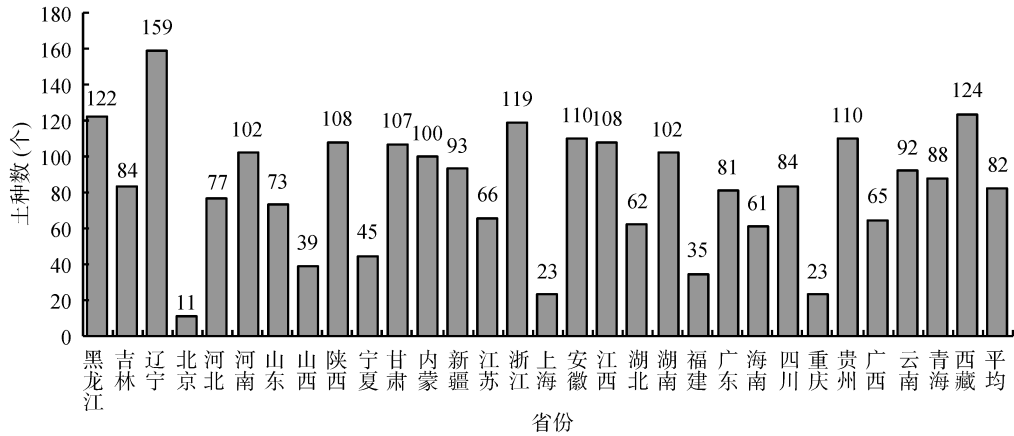


图 3 中国土种数据库收录的土种在各省的分布
Fig. 3 The soil local type distribution of Soil Type Database of China in 30 provinces

在中国土种数据库收录的 2 473 个土种典型剖面 和 8 751 个发生层中,含有 28 项土壤养分数据,包括:有机质、全氮、全磷、全钾、pH、速效养分、颗粒组

成、容重、阳离子交换量和交换性阳离子数据。全部 典型剖面各发生层的土壤养分数据量为 51 229 条,其 中数据量最多的为 pH、有机质、全氮、全磷和全钾。

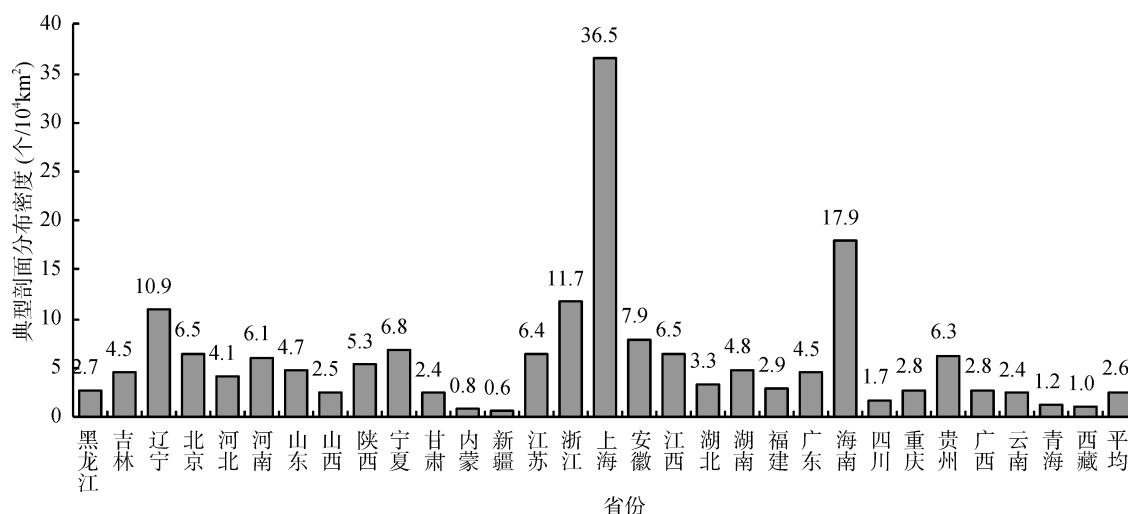


图 4 中国土种数据库收录的土种在各省的典型剖面分布密度

Fig. 4 The typical profile distribution densities of Soil Type Database of China in 30 provinces

4.2 数据应用

匿名用户可在在线查询本数据库中所有内容。协议用户正式填写数据使用协议,经数据产生者同意并签署数据使用协议后,由中国科学院南京土壤研究所土壤分中心离线分发用户所需的具体应用数据。使用者应在产生成果的显著位置注明中国土种数据库版权所有,在公开发表的成果中必须标注“数据来源于《中国土壤科学数据库》<http://www.soil.csdb.cn>”。数据 DOI 引用方式:施建平,宋歌,等.中国土种数据库.中国土壤数据中心.2014. doi:10.11666/geo20140001.ver1.db.

基于 Web 的中国土种数据库自 2006 年建成以来,为科学研究、教育、规划管理、施肥决策和社会咨询提供了大量的服务,例如为中国 1:100 万 SOTER 项目提供了同时具有地点属性和土壤理化性质属性的全国土种剖面数据,由此进行全国土地退化和耕地后备资源的评价^[11],作为基础数据用于农田土壤固碳潜力估算^[12-15]等工作中。

致谢:周慧珍、孙波、潘贤章、徐江兵、王昌昆参加了中国土种数据库建设的部分工作,特此致谢!

参考文献:

- [1] 全国土壤普查办公室.中国土种志(第一卷至第六卷)[M].北京:中国农业出版社,1996
- [2] Soil Data Subcommittee, Federal Geographic Data Committee. Soil Geographic Data Standard (FGDC-STD-006) [R/OL]. Reston, Virginia. 1997[2015-12-01]. <http://www.fgdc.gov/standards/projects/FGDC-standards-projects/soils/soil997.PDF>
- [3] United States Department of Agriculture, Natural Resources Conservation Service. SSURGO Data Packaging and

Use[R/OL]. 2012[2015-12-01]. http://www.nrcs.usda.gov/wps/PA_NRCSCConsumption/download?cid=stelprdb1237815&ext=pdf

- [4] 中国农业科学院农业资源与农业区划研究所,全国农业技术推广服务中心,中国标准化研究院.中国土壤分类与代码(GB/T 17296-2009)[S].北京:中国标准出版社,2009
- [5] Shi X Z, Yu D S, Warner E D, et al. Soil database of 1:1 000 000 digital soil survey and reference system of the Chinese genetic soil classification system[J]. Soil Survey Horizons, 2004, 45(4): 129-136
- [6] Nachtergaele F, Velthuisen H, Verelst L. Harmonized World Soil Database (version 1.2)[R/OL]. 2009[2015-12-01]. http://www.fao.org/fileadmin/templates/nr/documents/HWSD/HWSD_Documentation.pdf
- [7] GlobalSoilMap.net Science Committee. Specifications Version 2 GlobalSoilMap.net products[R/OL]. 2011[2015-12-01]. http://www.globalsoilmap.net/system/files/GlobalSoilMap_net_specifications_v2_0_edited_draft_Sept_2011_RAM_V12.pdf
- [8] Soil quality-Digital exchange of soil-related data (ISO 28258:2013(E))[S]. Switzerland, 2013
- [9] 中国科学院科学数据中心支撑软件组.VisualDB2.0 用户使用手册(内部资料)[R].北京:中国科学院计算机网络信息中心,2010
- [10] 中国标准化研究院,民政部区划地名司.中华人民共和国行政区划代码(GB/T 2260-2007) [S].北京:中国标准出版社,2007
- [11] 张甘霖,吴运金,赵玉国.基于 SOTER 的中国耕地后备资源自然质量适宜性评价[J].农业工程学报,2010, 26(4): 1-8
- [12] Yan X Y, Cai Z C. Number of soil profiles needed to give a reliable overall estimate[J]. Soil Science and Plant Nutrition, 2008, 54(5): 819-825
- [13] 覃章才,黄耀.基于模型的农田土壤固碳潜力估算[J].中国科学:生命科学,2010, 40(7): 658-676

- [14] 马渝欣, 李德成, 李徐生, 等. 利用农田表层(0~20 cm) 有机碳估算 1 m 土体有机碳的可靠性[J]. 土壤学报, 2014, 51(1): 189–193
- [15] 李晓迪, 王淑民, 张黎明, 等. 土壤数据源和制图比例尺对旱地土壤有机碳储量估算的影响[J]. 土壤学报, 2016, 53(1): 58–71

Web Based Soil Type Database of China

SHI Jianping, SONG Ge

(*Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, China*)

Abstract: Soil type database of China based on the second national soil survey is the most comprehensive soil data resource on the national scale. It embodies most soil types throughout the country in the 1980's, including distribution, area, characteristics, land-use and product performance of these soil types. Since 2000, the task of building the web based soil type database of China has been undertaken by Institute of Soil Science, Chinese Academy of Sciences. An entity-relationship (E-R) data model was constructed which involved the spatial relationship between location and soil type, and the classification relationship according to Genetic Soil Classification of China. The database contains 2 473 typical profiles, 8 751 soil horizons and the corresponding physical and chemical properties. The database can be widely used for land degradation evaluations, environmental impact studies, and soil carbon sequestration studies. It also can be used as a basic data for agricultural practices.

Key words: Soil Type Database; China; The second national soil survey; Data integration