

基于知识规则的土地利用/土地覆被分类方法 ——以黑河流域为例

候玉婷¹, 王书功², 南卓铜³

- (1. 兰州大学资源环境学院干旱区水循环与水资源研究中心, 兰州 730000;
2. 匹兹堡大学土木与环境工程系, 美国 匹兹堡 15261;
3. 中国科学院寒区旱区环境与工程研究所, 兰州 730000)

摘要:本文提出了一种基于知识规则的土地利用/土地覆被分类的新方法。知识规则是基于专家经验建立起来的,反映研究区内不同分类系统下各类别的地理分布特性与地理分布交叉可能性。基于黑河流域 90 m 分辨率 DEM、2009 年逐月 1 km 分辨率 NDVI,参考美国地质调查局 (USGS) 1 km 分辨率土地利用/土地覆被数据在欧亚大陆上各类别的聚类中心,应用在上、中、下游分别建立的知识规则,以知识规则结合最近距离的 USGS 类别聚类的方法,制作了一套与 USGS 全球土地覆被分类标准一致的、可以用于大气模式以及陆面过程模式的黑河流域土地覆被类型分布数据。本方法分类结果与以往研究采用的类别映射方法的分类结果及实际地物影像进行对比,表明知识规则下的分类结果更能准确表达流域地表覆盖特征,对冰雪、冻土类别和沙地荒漠类别的表现更优。

关键词:黑河流域;土地利用/覆盖;知识规则;USGS 土地利用/土地覆被数据;最小距离分类法

1 引言

研究表明,土地利用/土地覆被 (Land use/land cover, LULC) 对区域的生态和环境有巨大影响^[1],LULC 数据及其特征参数为气候、水文及陆面过程模型提供下垫面信息,显著影响模拟结果^[2],是模拟的重要数据基础^[3-4]。在模型中 LULC 数据和陆面参数表组合表达陆面特征的空间分布^[5-8]。其中,模型参数表列出 LULC 类别相应的陆面参数值,例如地表粗糙度、反射率、土壤湿度、叶面积指数、气孔阻抗等。许多工作致力于探讨 LULC 数据对模拟结果的影响^[9-11],以及如何产生适宜模拟的 LULC 分类体系^[12]、数据^[9, 13]和参数表^[14-17]。这些研究表明,用于模拟的 LULC 数据除了要能保证精确性、时效性和足够的分辨率外,还必须配有可靠的参数表^[2]。陆面参数表的建立往往需要复杂的标定计算,不同方法标定的参数差异显著^[14]。此外,针对具体应用还需进一步调整,因而带有许多不确定性。所以,当前仅有若干广泛应用的土地覆被分类体系发展了较为成熟的陆面参数表。MM5 (Mesoscale Model 5) 中总结了 3 种流行分类体系—USGS (US Geological Survey, 美国地质调查局), SiB (简单生物圈模式) 和 PSU/NCAR (美国宾夕法尼亚大学/美国国家大气研究中心) 分类体系的参数,其中 USGS 分类体系^[18]还被 Noah LSM (Land Surface Model)^[6]和 WRF (Weather Research and Forecasting Model) 模式^[19]采用。此外 IGBP (The International Geosphere-Biosphere Programme)、Olson Global Ecosystems^[15-16]、BATS (Biosphere-Atmosphere Transfer Scheme)^[7] 分类系统也有对应的参数表。这些参数表被广泛认可,不断被各种模型采用 (例如,德国马普气象研究所大气环流模式 ECHAM5 和区域气候模式 RegCM^[7])。模型用户如使用当前无成熟参数表的其他 LULC 分类体系及数据,就必须自行发展陆面参数表^[12],但自行发展的参

收稿日期: 2010-07-22; 修订日期: 2010-12-20

基金项目: 科技部 863 项目(2008AA12Z205); 中科院重要方向项目 (KZCX2-YW-Q10-1)[Foundation: MOST 863 Project No.2008AA12Z205; Knowledge Innovation Project CAS, No.KZCX2-YW-Q10-1]

作者简介: 候玉婷 (1988-), 女, 硕士生, 主要从事黑河流域水文模拟研究。E-mail: yuting.hou88@gmail.com

通讯作者: 南卓铜 (1977-), 男, 博士, 研究员, 现从事空间决策支持系统、水文集成建模研究。E-mail: nztong@lzb.ac.cn

数表欠缺验证而没有普适性,故一般用户多选择上述分类体系下的LULC数据输入模型。

当前已有多套上述分类体系下的全球1 km分辨率LULC数据,其中USGS GLCC (Global Land Cover Characteristics Database) 数据被MM5和WRF指定为默认数据,是进行MM5和WRF模拟的首选。然而,这些全球数据在除北美外的很多区域精度过低^[20],在某些区域甚至完全错误^[21],在中国的精度远低于全球的平均精度^[22],数据存在很大的不确定性^[23],并且时效性不好^[24](USGS GLCC和马里兰大学全球数据集UMD为1992-1993年,MODIS LANDCOVER为2000-2001年,欧盟联合研究中心2000年全球土地覆盖计划GLC2000为1999-2000年)。为了给模式提供更加准确的下垫面信息,研究者或以遥感影像及衍生的植被指数等重新制作LULC数据^[9,13],或将其他不同分类体系的LULC数据、植被图等以映射或更复杂的算法转化到参数表匹配的分类体系下^[25-27]。但前者成本较高,分类方法、研究区域及分类人员的不同都对分类结果有影响,限制不同区域、不同研究间的比较;而后者成本较低,可以综合多源数据,制作过程可复制,因而近年来相关研究较多,如:刘伟等^[26]曾将中国土地利用分类映射到USGS分类体系下,再有华等^[25]曾应用推理决策方法集成多源数据产生USGS分类体系下的LULC数据,陈峰和谢正辉^[27]曾基于中国植被图采用映射并添加AVHRR森林覆盖数据、降水气温数据多资料决策的方法发展了一套用于NCAR/CLM (National Center for Atmospheric Research/Common Land Model) 模式的LULC数据。映射方法因其操作方便被用户广泛采用,但也存在不能充分考虑分类体系间的定义差别的不足,可见将复杂的对应关系简化为一对一的映射并不恰当,并且产生的结果无法增添原数据不含有类别。

当前公布的中国土地利用数据已有多期,其中国家自然科学基金委员会“中国西部环境与生态科学数据中心”(http://westdc.westgis.ac.cn) 提供全国范围内的1985年、1995年、2000年3期,中国科学院寒区与旱区环境与工程研究所制作有黑河流域的2005年和2007年两期,均采用分层的中国土地资源分类系统,基于Landsat MSS, TM和ETM遥感图像,由全国各地的有关专家根据野外调查采样,对图像光谱、纹理、色调等的认识并结合地形图目视解译而成,可以保证精度^[28-29],是国内模型研究的良好数据源。但是,由于中国土地利用图并没有相应的陆面参数表,气候、水文及陆面过程模式并不能直接应用,因而在中国土地利用数据的基础上产生流行分类体系下的LULC数据非常有意义。

基于上述认识,本文在中国土地利用数据的基础上,提出了一种基于知识规则的土地利用/土地覆被分类的新方法,在黑河流域应用,得到一套USGS分类体系下的可用于气候、水文及陆面过程模式的LULC数据,并且做了精度验证。

2 研究区概况、数据及处理

黑河流域位于96°42'~102°00'E、37°41'~42°42'N之间的河西走廊中部,流域面积约13万km²(图1)。流域内景观分异鲜明,从南到北可划分出南部祁连山地、中部河西走廊和北部阿拉善高平原3大地貌单元,并部分与巴丹吉林沙漠和腾格里沙漠接壤。气候干燥,降水稀少而集中,多大风,日照充足,太阳辐射强烈,昼夜温差大。上游祁连山山区植被属温带山地森林草原,生长着呈片状、块状分布的灌丛和乔木林,垂直带谱极其明显;中下游地带性植被为温带小灌木、半灌木荒漠植被^[30]。

为建立准确的黑河流域LULC数据,本文使用了NDVI数据、DEM数据和多种、多期的土地利用数据,以下分述几组数据的来源及处理方式。

土地利用数据包括黑河流域的两期中国土地利用图(2000年1:10万全流域,2007年1:5万中上游部分)和USGS GLCC数据集USGS分类体系数据(欧亚大陆属性表数据库)。2000年黑河流域土地利用图数据来自“中国西部环境与生态科学数据中心”,截取自中国2000年

1:10万土地利用数据集^[28-29];2007年黑河中上游土地利用图由中科院寒区与旱区环境与工程研究所王建华等制作。两期土地利用图经过最大面积栅格化和叠加合成的处理,生成栅格分辨率为1 km,以和USGS数据分辨率保持一致,叠加是用2007年数据覆盖替换了2000年数据的中上游部分。2000年后黑河流域土地利用变化主要发生在上中游,而下游戈壁荒漠区变化较小,因而叠加可在一定程度上反应流域土地利用的现状,由此得到部分区域更新的黑河流域1 km栅格土地利用图。USGS GLCC数据集 USGS分类体系数据来自USGS数据下载中心 (<http://edc2.usgs.gov/glcc/glcc.php>),主要应用数据的欧亚大陆属性表,统计USGS分类体系下各类别涵盖的所有区域,计算了各类的属性平均值,包括NDVI月均值和高程平均值。

取2009年全年1 km分辨率月合成MODIS NDVI数据用以在分类中形成特征空间,使用MODIS NDVI是因为考虑到USGS GLCC的制作基于AVHRR NDVI,而MODIS NDVI是AVHRR NDVI的继承且两者相似^[31]。数据来自NASA MODIS数据分发中心(<https://wist.echo.nasa.gov/>),为二级植被指数产品,应用前仅对数据做了拼接和投影变换。取90 m分辨率SRTM (Shuttle Radar Topography Mission, 航天飞机雷达地形测绘任务) DEM数据作为提高分类精度的辅助信息,为配合其他数据,分辨率重采样到1 km。

3 分类方法

基于知识规则的土地覆被分类方法有以下两个环节,①可延续USGS数据类别定量属性的USGS类别聚类方法;②可将专家知识和先验信息加入分类的聚类知识规则。具体流程见图2。

3.1 USGS类别的最小距离聚类方法

USGS GLCC与多数基于人工解译定性表述地物的LULC数据在制作方法上有所区别,具有定量分类和依赖专家知识的特点:以季节性土地覆被区SLCR (Seasonal Land Cover Region) 为基础理念,同一SLCR的土地覆被植物区系相同,群落外貌相似,并具有相同的物候特征和第一性生产力,可以和其他SLCR相区别^[32];应用基于多时相NDVI数据的非监督分类方法,先以K-mean算法产生聚类,再参考DEM、气候区划图等其他资料

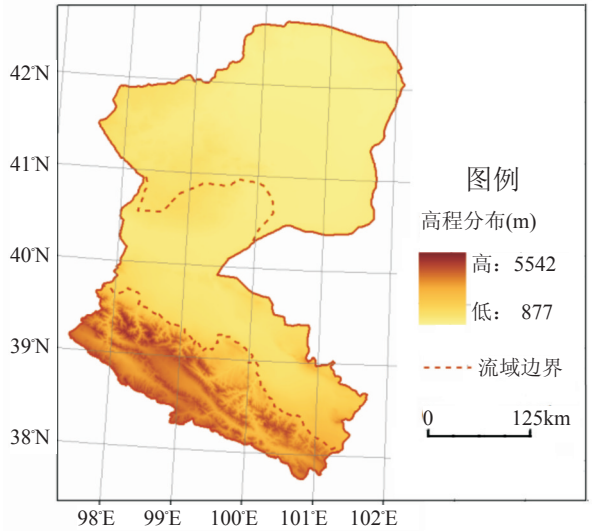


图1 黑河流域图

Fig. 1 The map and topography of the Heihe River Basin

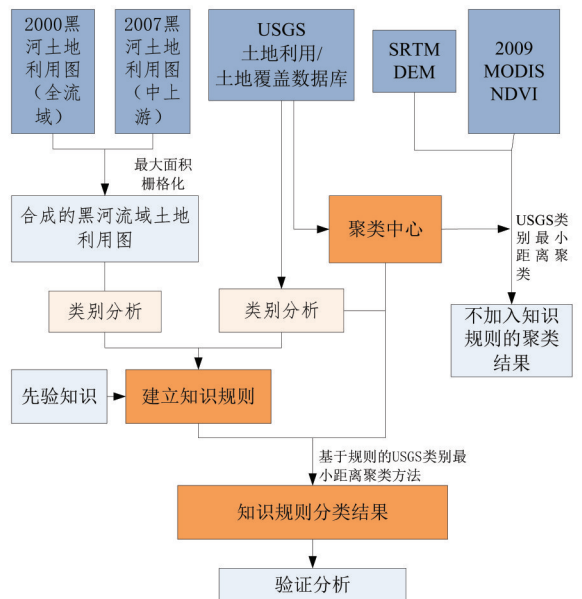


图2 研究方法流程图

Fig. 2 The flow chart of rule-based land cover classification method

由专家修订到不同SLCR下,最后确定SLCR所属的USGS类别^[32]。USGS的制作周期很长、工作强度大且涉及大量专家经验,重复USGS分类步骤更新数据难度相当大。但是,在USGS的季节性土地覆被区的理念和非监督分类的方法下,所分类别有可循的定量特征。因此,可以用USGS类别的定量特征区别类别,从而延续USGS数据的定量分类特点并沿用USGS陆面参数表。

多时相的NDVI数据组可以反映地表物候和生产力的变化^[33],表达分类的关键信息^[34-35],并且,USGS GLCC数据正是基于多时相的NDVI建立起来的,因此可以用各类别的NDVI物候变化来表征USGS类别的定量特征,故本文建立的LULC聚类方法主要是基于多时相的NDVI数据。但是,考虑到存在部分LULC类别的物候特性非常接近,如果增加数据源信息^[34],特别是DEM信息可以显著提高分类精度^[36-37],所以为了减少错分,本方法引入高程信息作为辅助,以提高类别间的可分度。USGS各类别的NDVI物候变化特征和高程分布特征来自对USGS GLCC数据库属性表的统计,所形成的特征向量值为各别逐月NDVI和平均高程值。特征向量作为类别的聚类中心,在更新的逐月NDVI数据和DEM数据形成的13维特征空间中,计算栅格点样本特征向量到聚类中心的欧氏距离,以欧氏距离最小作为判定类别的标准。USGS GLCC数据库欧亚大陆属性表中仅含有24个USGS类别中的21个,不含有的是城市与建设用地类(USGS用掩模确定此类)、旱地/水浇地混合和牧场、草本苔原,故总共得到21个类别的聚类中心。

3.2 聚类知识规则

单纯使用上述聚类方法并不能精准的区分类别,为提高分类精度,需要引入更多信息和专家知识^[38]。建立知识规则,可以为USGS类别聚类方法补充其他来源的数据信息和专家知识(后文统称“知识”),如此也符合了USGS数据制作依赖专家知识的特点。

以知识规则方式引入分类的“知识”包括:中国土地利用数据对地表类型的准确描述,以及中国土地资源分类系统和USGS分类系统间的对应关系。后者基于研究人员野外工作积累的地学知识和对分类体系差异(表1、表2)的认识,与“一对一”的映射不同,提出“一对多”的对应可能性,是研究人员对各类别的地理分布特性与地理分布交叉可能性的判断。在知识规则下,首先流域内的栅格点依据中国土地利用数据分类,利用专家知识为不同种类的栅格点赋以不同的类别空间,类别空间涵盖该栅格种类可能归属的多种USGS类别。在映射方法下,所有栅格点的类别空间类别唯一,但在知识规则下,类别空间取决于“知识”,所含类别数介于1-24之间。形成的知识规则表达为中国土地利用类型和USGS类型的对应矩阵,如认为中国土地资源分类系统下某类别覆盖的区域可能对应USGS的某类别,在此两类别行列交叉的位置上以“√”标记,标记的USGS类别的总和即为该中国土地资源分类类别下栅格点对应的类别空间。

在类别空间中采取USGS类别的聚类方法判断所属类别,如此就是知识规则下的USGS类别聚类方法。本方法集合了专家知识和定量的分类特点,因而可产生符合USGS数据特点的LULC数据。

4 分类结果和精度评价

4.1 建立的知识规则

黑河流域景观格局分异明显,上游和中下游的地物种类和群落组成差别大,中国土地资源分类系统和USGS分类系统的对应关系并不相同,故对上游和中下游流域分别建立知识规则(表3)。为了使分类结果简洁清晰,规则中排除了USGS分类体系下的3类混合类别(农田/林地混合、农田/草原混合、灌木/草原混合)。考虑到黑河流域的实际状况,排除了热带或亚热带稀树草原、落叶针叶林、常绿阔叶林3类。USGS和中国土地资源分类系统下的农田

表1 中国土地资源分类系统和USGS分类系统

Tab. 1 China land cover classification system (a), and USGS land cover classification system (b)

(a)	A1	耕地	a44	永久性冰川雪地		
	a11	水田	a45	滩涂		
	a12	旱地	a46	滩地		
	A2	林地	A5	城乡、工矿、居民用地		
	a21	有林地	a51	城镇用地		
	a22	灌木林	a52	农村居民点		
	a23	疏林地	a53	其它建设用地		
	a24	其它林地	A6	未利用土地		
	A3	草地	a61	沙地		
	a31	高覆盖度草地	a62	戈壁		
	a32	中覆盖度草地	a63	盐碱地		
	a33	低覆盖度草地	a64	沼泽地		
	A4	水域	a65	裸土地		
	a41	河渠	a66	裸岩石砾地		
a42	湖泊	a67	其它			
a43	水库坑塘					
(b)	B1	Urban and Built-Up Land	城市和建筑用地	B13	Evergreen Broadleaf Forest	常绿阔叶林
	B2	Dryland Cropland and Pasture	旱地和牧场	B14	Evergreen Needleleaf Forest	常绿针叶林
	B3	Irrigated Cropland and Pasture	水浇地和牧场	B15	Mixed Forest	混合林
	B4	Mixed Dryland/Irrigated Cropland and Pasture	旱地/水浇地和牧场	B16	Water Bodies	水体
	B5	Cropland/Grassland Mosaic	农田/草原混合	B17	Herbaceous Wetland	草本湿地
	B6	Cropland/Woodland Mosaic	农田/林地混合	B18	Wooded Wetland	多树木湿地
	B7	Grassland	草原	B19	Barren or Sparsely Vegetated	荒地或者稀疏植被下垫面
	B8	Shrubland	灌木	B20	Herbaceous Tundra	草本苔原
	B9	Mixed Shrubland/Grassland	灌木/草原混合	B21	Wooded Tundra	多树木苔原
	B10	Savanna	热带或亚热带稀树草原	B22	Mixed Tundra	混合苔原
	B11	Deciduous Broadleaf Forest	落叶阔叶林	B23	Bare Ground Tundra	裸露的苔原

表2 中国分类系统和USGS分类系统对比

Tab. 2 The comparison of the China classification system and USGS classification system

	中国土地利用分类系统	USGS土地利用/土地覆被分类系统
类别总数	25类	24类
耕地	区分旱地和水田	区分旱地和水田
林地	林地的分类指标是郁闭度和高度,分四类:有林地 (>30%),灌木林 (>40%, <2 m),疏林地 (10%~30%),其它林地(果园等园地)	划分详细,根据常绿/落叶和阔叶林/针叶林组合出四类,还包括灌木、稀树草原和混合林,共7类
草地	草地划分详细,分高、中、低三种覆盖度草地,覆盖度范围分别是>50%、20%~50%、5%~20%	只有一类
城镇用地	城镇建设用地划分详细,分三类	只有一类
裸地类	详细区分出沙地、戈壁、盐碱地、裸土地、裸岩石砾地和其它(高寒荒漠和苔原),共6类	只有一类,且不是单纯的裸地,也含稀疏植被下垫面
湿地类	只有一类	分草本湿地和多树木湿地
苔原类	无苔原类	分四类:草本苔原、多树木苔原、混合苔原和裸露苔原
混合类	无混合类	包含很多混合像元类
总结	对低植被覆盖的区域分类详细,没有根据植被类型来分类,高寒植被区分不佳,林地区分简单,草地分类笼统	植被类型区分详细,但没有按照覆盖度或郁闭度来分类

类别、建设用地、水域、冰雪和部分未利用土地可以良好对应,故规则中还是采用了“一对一”的映射。中国土地资源分类体系下的低覆盖度草地以5%为覆盖度下线,USGS分类下此覆盖度应属于荒地或稀疏植被下垫面,故设定此两类存在对应关系。中国土地资源分类体系中灌木类别的覆盖度下线为40%,USGS分类体系中则为10%,之间相差的部分可能为草地

表 3 知识规则表
 Tab. 3 The classification tables for the upstream mountainous area
 a. 上游

	A1			A2			A3			A4			A5			A6						
	a11	a12	a21	a22	a23	a24	a31	a32	a33	a41	a42	a43	a44	a46	A5	a61	a62	a63	a64	a65	a66	a67
B1															✓							
B2		✓																				
B3	✓																					
B4																						
B5																						
B6																						
B7							✓	✓	✓													
B8				✓	✓																	
B9																						
B10																						
B11			✓		✓	✓																
B12																						
B13																						
B14			✓		✓																	
B15			✓		✓	✓																
B16										✓	✓	✓			✓							
B17							✓	✓														
B18																				✓		
B19									✓							✓	✓	✓			✓	✓
B20																					✓	✓
B21			✓		✓																	
B22				✓			✓	✓	✓													
B23									✓												✓	✓
B24													✓									

b. 中下游

	A1			A2			A3			A4			A5			A6						
	a11	a12	a21	a22	a23	a24	a31	a32	a33	a41	a42	a43	a44	a46	A5	a61	a62	a63	a64	a65	a66	a67
B1															✓							
B2		✓																				
B3	✓																					
B4																						
B5																						
B6																						
B7							✓	✓	✓													
B8				✓	✓		✓	✓	✓													
B9																						
B10																						
B11			✓		✓	✓																
B12																						
B13																						
B14																						
B15			✓		✓	✓																
B16										✓	✓	✓			✓							
B17							✓	✓													✓	
B18																					✓	
B19									✓							✓	✓	✓			✓	✓
B20																						
B21																						
B22																						
B23																						
B24													✓									

注:分类编号参见表 1

或疏林,故中国土地资源分类下的草地、疏林和灌木都与USGS灌木类别对应。中国土地利用图中下游居延海大小明显不符合当前状况,水域面积过大,当前仅有依靠调水维持的小片

水域,所以做了特别调整,设定居延海面积缩减部分的类别空间涵盖全部类别。

上述是适用于全流域的知识规则,上游和中下游各自独立的知识规则分述如下:

(1) 上游。黑河上游是黑河流域的径流形成区,水资源受冰川、冻土、积雪和水源涵养林的共同作用^[39],且冻土对黑河水资源形成和稳定有重要影响^[40]。因此在规则中,尽量详细和准确的确定森林、湿地、苔原、裸土类别非常重要。林地类型的不同可影响长期的水文效益,因而林地必须细分开。上游乔木林以针叶林居多,也有阔叶林和混合林,但中国分类系统下均归为有林地一类,故将有林地对应USGS落叶阔叶林、常绿针叶林、混合林以及多树木苔原4类。中国土地利用分类体系下的灌木林可能涵盖高山植被带上的灌丛草甸,故将该类对应于USGS灌木和混和冻土两类。疏林地则与USGS下的灌木和所有乔木林都建立对应关系。中国土地资源分类中的其他林地多指种植园地,USGS中没有种植园地一类,故设定其可能对应于所有林地类别。上游中国土地资源分类中各草地类别所指地表类型基本一致,都包含了草原和高山草甸植被类型,对应USGS的草原和草本苔原,苔原类别为上游特有。另外设定裸土地和其他土地都与裸露的苔原这一类对应。

(2) 中下游。黑河流域中游山前冲积扇下部和河流冲积平原上分布有灌溉绿洲栽培农作物和林木,呈现以人工植被为主的绿洲景观。下游两岸三角洲与冲积扇缘的湖盆洼地里,生长有荒漠地区特有的荒漠河岸林,灌木林和草甸植被,荒漠植被的植物种类和中游差别不大,景观也主要为荒漠和天然绿洲景观。林木基本为阔叶林,少有针叶林,故在规则中去除了中国土地资源分类下有林地类型与USGS针叶林的对应。和上游规则相比,另一则改动是去除了所有的苔原类别。

4.2 基于知识规则的分类结果

图3展示的是基于知识规则的分类结果,共含有16个类别。① 上游草原面积最大,草本湿地、荒地或稀疏植被下垫面和灌木林次之,三者为上游最主要类别;存在成片分布的常绿针叶林、阔叶林和混合林,三者连接成片并有混合交错的现象;冰雪和冻土类别层次明显,冰雪类别的外围被多树木苔原、混合苔原和裸露冻土包围;此外在整个上游范围内,零星分布许多多树木苔原。② 中游部分以灌木林、水浇地、旱地和荒地或稀疏植被下垫面占优势,且存在成片的草本湿地和落叶阔叶林。③ 下游大面积的荒地或稀疏植被下垫面的背景之上,成片的灌木是最主要的覆盖类型,其次还存在落叶阔叶林和草原。

4.3 知识规则分类结果精度评价

在黑河流域范围内,挑选6块典型区域(图3、图4),比较知识规则分类结果、映射方法结果和不加入知识规则的聚类结果的差别,本文采用Google Map遥感影像作为实际地物目视对比的基准。

图5a是不加入知识规则的聚类结果,全流域中不设定聚类规则,按聚类方法分类到有聚类中心数据的21个类别中,未加城市和建设用地掩模,也不含旱地/水浇地混合和牧场以及草本苔原类别,故此3类并未分析。图5b是映射方法的分类结果。本文所用的映射方法和刘伟等^[29]的做法相似:除乔木林外的所有类别一对一映射到USGS分类系统下,但乔木林

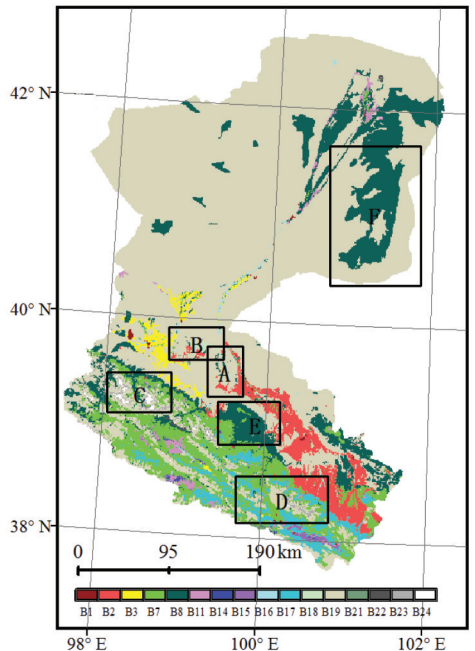


图3 知识规则的分类结果及选择的6个典型区域图

Fig. 3 Schematics of classified LULC by the rule-based method and selected 6 areas for comparison

并未参考植被分类图做二级划分,图例中全部标注为混合林。这是因为土地利用图与植被分类图并不能良好嵌套,土地利用图中的有林地类往往并没有直接落入植被分类图下的乔木林中,而是分散于草甸植被、草原植被和荒漠植被区中,并不能直接确定乔木林种类。

区域 A 位于黑河中游,除深绿色的耕地外,以荒地或稀疏植被下垫面为主,零星分布有未连接成片的荒漠草原和灌木。在映射结果(图 4A-2)中,草原明显过多;在知识规则分类结果(图 4A-1)中则表现灌

木和荒地或稀疏植被下垫面;不加入知识规则的聚类结果(图 4A-3)中耕作区存在严重误判,其他区域与知识规则分类结果相似。在区域 A,知识规则分类结果最接近影像。

区域 B 位于黑河中游,这里不存在明显水体,但植被状况较好,存在成片耕地。映射结果(图 4B-2)中的水体明显与当前状况不符。不加入知识规则的聚类结果(图 4B-3)中,对应区域主要被判为了荒地或稀疏植被下垫面,耕地全部误判。而知识规则分类结果(图 4B-1)最接近真实地物,对应区域表现为多树木湿地。

区域 C 选取上游西侧高海拔地段。这里山顶覆盖积雪冰川,随海拔变化植被表现明显的垂直地带性。知识规则(图 4C-1)与映射结果(图 4C-2)的冰雪面积和影像对应较好,而不加入知识规则的聚类结果(图 4C-3)明显高估了冰雪的覆盖面积。知识规则分类结果包含的类别比映射结果更加丰富,冰雪被裸露的冻土包围,具有明显的地带性特征,此外山谷和山前出现成片的灌木。三者相较知识规则下的分类结果最好。

区域 D 选取上游东侧的高海拔地段。山顶覆盖积雪冰川,山麓分布较多林地。但冰雪和林地只在不加入知识规则的聚类结果(图 4D-3)中体现,在知识规则分类结果(图 4D-1)和映射结果(图 4D-2)中则基本误判为荒地或稀疏植被下垫面类,并且两者林地都很少。知识规则分类结果中存在大片的冻土苔原类别,有一定的高山植被特点,映射结果则基本没有明显的高山植被特征。不加入知识规则的聚类方法总体表现最佳,知识规则方法次之。

区域 E 代表的是中游的荒漠植被覆盖区。除东侧的耕地外此区域内以荒漠草原和荒漠灌丛占优势,映射结果(图 4E-2)中几乎全被判作草原,而在知识规则分类结果(图 4E-1)和不加入知识规则的聚类结果(图 4E-3)中则有一半被判为灌木,在 USGS 分类体系的定义下,后者更接近实际地物。

区域 F 代表的是下游沙地植被覆盖区,生长着大片的沙地灌木和小半灌木,在中国土地利用图中并没有被划归灌丛,而是被判为草原,故在映射结果中,这一区域全显示为草原。而在不加入知识规则的聚类方法下,这里被划为荒地或稀疏植被下垫面,低估了实际的植被

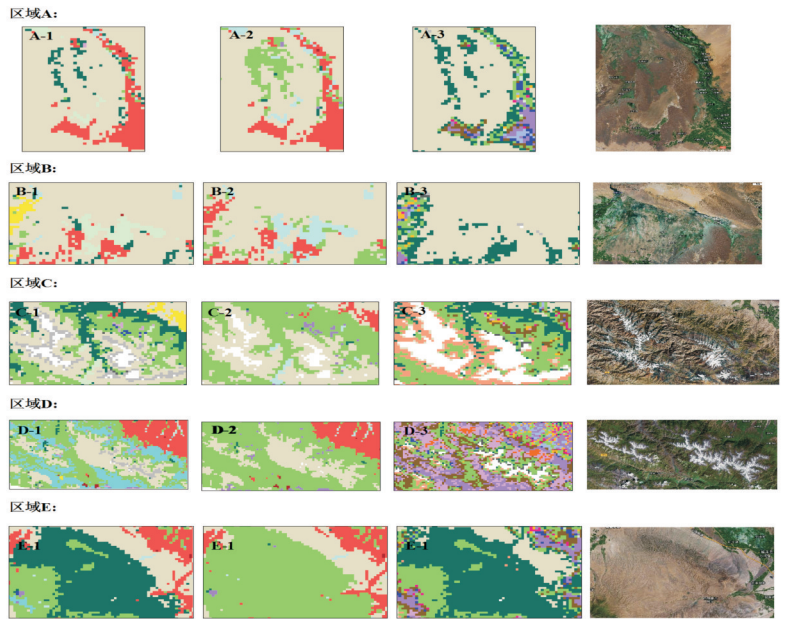


图4 各典型区上的分类结果和影像对比(图例参见图3)

Fig. 4 LULC classification in selected areas by minimum distance method, direct mapping method and rule based method, and images of Google Map (from left to right). (The same legends as in Fig. 3 were applied)

量。知识规则的分类方法将此区域判为灌丛,显示其原本的地物类型,因而表现最好。

简单的典型区目视对比发现,加入和未加入知识规则的分类方法均表现出比映射方法结果更合理的分类结果。下面讨论知识规则对分类结果的总体上的改善。

5 讨论

5.1 聚类中心对聚类结果的影响

基于聚类中心的USGS类别聚类法是知识规则分类的重要环节,有必要探讨此方法是否存在弊端、是否可靠、对聚类结果有何影响。由图5a可见,不加入知识规则的聚类方法能较准确的划分多数类别,能够很好的反映出黑河流域的地表覆盖特征:上游多林地、草地且垂直地带性明显,中游多灌木草原并且绿洲区集中,下游区分出了沿黑河两岸的绿洲植被;但在部分类别上却存在明显的错判:农田和林地错分严重,下游植被覆盖面积明显过少。造成分类精度不佳的原因有二:①部分类别的NDVI变化特征非常相像,图6为USGS各类别的NDVI特征变化图,其中农田和林地这些被明显错判的类别恰是特征最相近的几个类别;②参考的信息量过少,仅用NDVI和高程信息不足以区分类别。因而为了提高分类精度,需要引入更多的知识和数据作为分类参考的依据。

5.2 高程特征对聚类结果的影响

图7是仅采用聚类方法且不考虑高程特征的分类结果,分类时未使用规则,聚类中心特征向量中未加入高程值。和图5b对比可见,高程对聚类结果的影响集中在上游的高海拔地区。加入高程特征后类别间交叠混杂的程度降低,同时冰雪和苔原判定的更为准确,但出现的林地和湿地类别改变,农田、林地的错判没有改善。

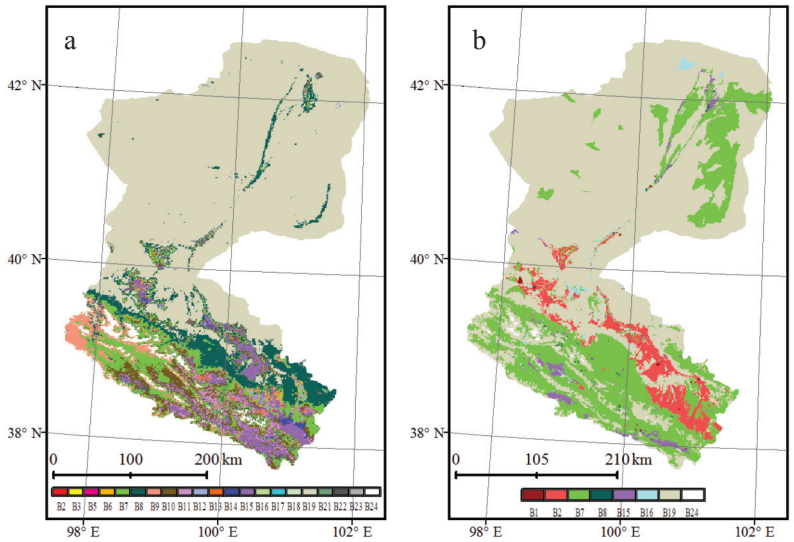


图5 不加入知识规则的(a)聚类结果;(b)映射结果
Fig. 5 LULC classification by minimum distance method (a), and LULC classification by direct mapping method (b)

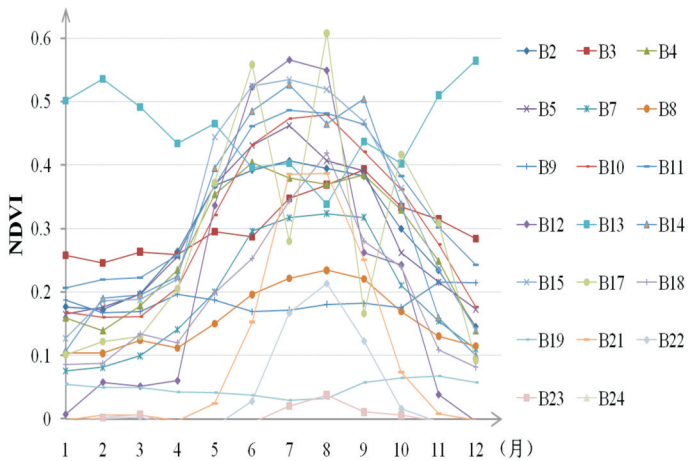


图6 USGS类别的NDVI变化特征图
Fig. 6 Monthly NDVI of USGS categories (Category B16-water bodies not shown, with a NDVI of approx. -1)

(未标注B16水体类,其NDVI值接近于-1)

未标注B16水体类,其NDVI值接近于-1

5.3 知识规则对聚类结果的影响

图3是知识规则的分类结果。其中特征相近类别的错判消失,说明知识规则能够有效的规避USGS类别聚类方法的缺点。

建立转移矩阵比较了知识规则分类结果与不加入知识规则的聚类结果间的差异。由此可见,不加入知识规则的聚类结果与知识规则输出结果的相似性并不高,知识规则的加入明显改变了加入知识规则的聚类结果,从表4可以看出,这种改变主要体现在草原、灌木、荒地或稀疏植被下垫面类别上,同时不加入知识规则的聚类结果中冰雪、冻土和苔原类别的面积要比规则分类结果中多出许多,其中冰雪面积增加了6倍,多出部分的栅格在规则结果中多属于裸露的冻土和荒地或稀疏植被下垫面。

建立转移矩阵比较了知识规则分类结果和映射结果间的差异。从表5中可以看出,知识规则的分类结果与映射结果间草原类别差异最大,其次是林地:归并方法下的草原被剖分到了草原、灌木、草本湿地、荒地或稀疏植被下垫面、混合苔原和裸露的冻土类共6类中;林地除包含3种林地类别外,还有部分属于灌木林、多树木苔原;另外在荒地或稀疏植被下垫面中也有部分归属于了裸露的冻土类。可见知识规则的分类方法大大增加了映射结果的类别丰富程度。

6 结论

本文提出了一种基于知识规则的土地利用/土地覆被分类的新方法。基于黑河流域90

表4 知识规则的分类结果和不加入知识规则的聚类结果间的转移矩阵

Tab. 4 The transfer matrix for the LULC classifications by minimum distance method and rule based method

土地利用类型	B3	B5	B6	B7	B8	B9	B10	B11	B12	B13	B14	B15	B18	B19	B21	B22	B23	B24
B1		5	5	25	66	0	13	3	1	0	0	17	1	34	0	0	0	0
B2		181	324	1118	432		586	25	344		456	1805	33	85		2		
B3		39	101	416	146		296		72		192	270	55	108	1	2	3	3
B7	5	108	95	6431		286	3521	863	59	15	229	335	63		10			
B8		4	4	84	5424	996	32	13		3	4	17	7	9287	6	11	4	1
B11		7	7	186	180		319	205		14	1		50	130				
B14			1	67			69		1		17							
B15							28				2	278						
B16		10	8	113	118	11	14	1	6		6	4	8	99	2	33	59	14
B17							167	1083	2	149	11	2120						
B18		3		89	28	26	13				1		1	178			4	2
B19		19	38	2492	3449	2054	399	91	22		62	112	62	74950	86	206	134	1038
B21				147	5	32									6	8	3	6
B22															19	2	30	52
B23																8	5	539
B24				4		7									2			312

注:行对应知识规则分类结果中的类别,列对应不加入知识规则的聚类结果中的类别

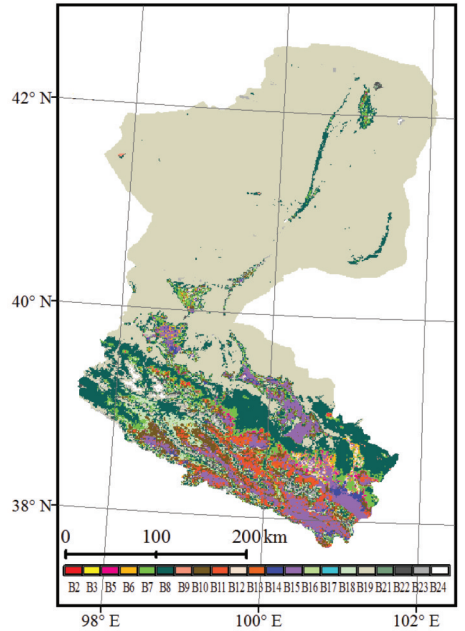


图7 未考虑高程特征、不加入知识规则的聚类结果

Fig. 7 LULC classification by minimum distance method without considering elevation

m分辨率DEM、2009年的逐月1 km分辨率NDVI,参考USGS 1 km分辨率土地利用/土地覆被数据在欧亚大陆上各类别的聚类中心,应用在上、中、下游分别建立的知识规则,以知识规则和USGS类别聚类的方法,制作了一套与USGS分类标准一致的、延续USGS类别定量特点的的1 km分辨率黑河流域土地覆被数据,可与USGS陆面参数表对应,可以直接用于大气及陆面过程模拟,在一定程度上解决了黑河流域缺少陆面过程模型、水文模型所需的土地覆被/利用类型生物、物理参数的问题。该数据与不加入知识规则的聚类结果、映射结果对比,发现知识规则的分

表5 知识规则的分类结果和映射结果间的转移矩阵

Tab. 5 The transfer matrix for the LULC classifications rule based method and direct mapping method. The rows and columns of the matrix are corresponding to LULC classifications by rule based method and direct mapping method, respectively

土地利用类型	B1	B2	B7	B8	B15	B16	B19	B24
B1	159	6	2	1				2
B2	2	5299	52			14		24
B3	1	1663	3	1		1		35
B7		34	11929	9	21	16		11
B8		7	15639	159	34	6		52
B11			34		1063			2
B14			5		150			
B15			7		301			
B16		26	51	1		393		35
B17		3	3512		3	13		1
B18		2	10			327		6
B19		71	777		3	239	84122	2
B21			3		197	7		
B22			81			22		
B23			2			5	543	2
B24								325

类结果与两者的区别主要表现在:① 冰雪和高山植被带;② 沙地荒漠植被覆盖。这两类都是黑河流域及至寒区旱区的重要地表覆盖类型,对于寒区旱区的陆表过程模拟研究有重要意义。将分类结果与实际地物影像进行对比,探讨了不同分类方法对分类结果精度的影响,主要结论如下:① 不加入知识规则的聚类方法能较准确的反映黑河流域的地表覆盖特征,但也会高估冰雪的覆盖面积,并且低估下游的植被覆盖度,有一定的准确性,但也存在由信息不足导致的明显错判。② 映射结果可分出的类别较少,对冰雪和冻土类别、湿地类别表现性不佳。③ 建立知识规则下的聚类方法,可以应用其他来源的数据信息和专家知识有效地避免不加入知识规则的聚类方法对冰雪的过高估计、对下游植被的过低估计和明显错判,且在映射结果的基础上,合理的丰富了类别种类。④ 与加入知识规则的聚类结果及映射结果相较,知识规则的分类结果总体上表现最好,精度优势最明显的区域是冰雪、高山植被和沙地荒漠植被覆盖区。本文提出的方法可以应用于产生符合USGS分类系统的全国土地利用/土地覆被数据集,其中知识规则按自然景观分区结合专家知识生成,分类结果可望改善现有全国尺度的陆面过程模型的模拟结果。

参考文献 (References)

- [1] Vorosmarty C, Lettenmaier D, Leveque C et al. Humans transforming the global water system. *Eos*, 2004, 85(48): 509-513.
- [2] Zhu Deqin, Gao Xiaoqing, Chen Wen. Validation of SSiB model over gobi in Dunhuang and its sensitivity to vegetation parameters. *Journal of Desert Research*, 2006, 26(3): 466-472. [朱德琴, 高晓清, 陈文. 陆面模式 (SSiB) 对敦煌荒漠戈壁下垫面陆面过程的模拟及敏感性试验. *中国沙漠*, 2006, 26(3): 466-472.]
- [3] Pielke R A, Lee T J, Dalu G A et al. Nonlinear influence of mesoscale land use on weather and climate. *Journal of Climate*, 1991, 4(11): 1053-1069.
- [4] Pielke S, Roger A. Influence of the spatial distribution of vegetation and soils on the prediction of cumulus convective rainfall. *Rev. Geophys.*, 2001, 39(2): 151-177.
- [5] Xue Y, Zeng F J, Mitchell K E et al. The impact of land surface processes on simulations of the US hydrological cycle: A case study of the 1993 flood using the SSiB land surface model in the NCEP Eta regional model. *Monthly Weather Review*, 2001, 129(12): 2833-2860.
- [6] Chen F, Dudhia J. Coupling an advanced land surface-hydrology model with the Penn State--NCAR MM5 modeling system. Part I: Model implementation and sensitivity. *Monthly Weather Review*, 2001, 129(4): 569-585.
- [7] Elguindi N, Bi X, Giorgi F et al. RegCM Version 3.1 User's Guide [EB/OL]. PWCG Abdus Salam ICTP, 2006.

- [8] Xiu A, Pleim J E. Development of a land surface model (Part I): Application in a mesoscale meteorological model. *Journal of Applied Meteorology*, 2001, 40(2): 192-209.
- [9] Sertel E, Robock A, Ormeci C. Impacts of land cover data quality on regional climate simulations. *International Journal of Climatology*, 2009, doi: 10.1002/joc.2036.
- [10] Byun D, Kim S, Cheng F Y et al. Utilization of satellite-derived high resolution land use/land cover data for the meteorological, emissions, and air quality modeling. Fifth Conference on Urban Environment. Amer. Meteor. Soc., 2004.
- [11] Civerolo K L. Simulating the effects of urban-scale land use change on surface meteorology and ozone concentrations in the New York City metropolitan region. Seventh Conference on Atmospheric Chemistry, 2005.
- [12] Cheng F Y, Byun D W. Application of high resolution land use and land cover data for atmospheric modeling in the Houston-Galveston metropolitan area (Part I): Meteorological simulation results. *Atmospheric Environment*, 2008, 42(33): 7795-7811.
- [13] Lam J S L, Lau A K H, Fung J C H. Application of refined land-use categories for high resolution mesoscale atmospheric modelling. *Boundary-Layer Meteorology*, 2006, 119(2): 263-288.
- [14] Burian S J, Stetson S W, Han W et al. High-resolution dataset of urban canopy parameters for Houston, Texas. Fifth Conference on Urban Environment, 2004: 23-26.
- [15] Hagemann S. An improved land surface parameter dataset for global and regional climate models. Max-Planck-Institut für Meteorologie, 2002: 1-21.
- [16] Hagemann S, Botzet M, Dümenil L et al. Derivation of global GCM boundary conditions from 1 km land use satellite data. Max-Planck-Institut für Meteorologie, 1999: 1-20.
- [17] Sellers P J, Los S O, Tucker C J et al. A revised land surface parameterization (SiB2) for atmospheric GCMs (Part II): The generation of global fields of terrestrial biophysical parameters from satellite data. *Journal of Climate*, 1996, 9(4): 706-737.
- [18] Anderson J R, Hardy E E, Roach J T et al. A land use and land cover classification system for use with remote sensor data. Washington, DC US Geological Survey: Professional Paper 964, 1976.
- [19] Wang W, Barker D, Bruyere C et al. WRF Version 2 Modeling System User's Guide, 2004.
- [20] Frey K E, Smith L C. How well do we know northern land cover? Comparison of four global vegetation and wetland products with a new ground-truth database for West Siberia. *Global Biogeochem. Cycles*, 2007, 21: B1016.
- [21] Sedano F, Gong P, Ferrao M. Land cover assessment with MODIS imagery in southern African Miombo ecosystems. *Remote Sensing of Environment*, 2005, 98(4): 429-441.
- [22] Gong Peng. Accuracy assessment of global land cover datasets based on global field stations. *Progress in Natural Science*, 2009, 19(7): 754-759. [宫鹏. 基于全球通量观测站的全球土地覆盖图精度检验. *自然科学进展*, 2009, 19(7): 754-759.]
- [23] Herold M, Mayaux P, Woodcock C E et al. Some challenges in global land cover mapping: An assessment of agreement and accuracy in existing 1 km datasets. *Remote Sensing of Environment*, 2008, 112(5): 2538-2556.
- [24] Ran Youhua, Li Xin, Lu Ling. Accuracy evaluation of the four remote sensing based land cover products over China. *Journal of Glaciology and Geocryology*, 2009, 31(3): 490-500. [冉有华, 李新, 卢玲. 四种常用的全球 1 km 土地覆盖数据中国区域的精度评价. *冰川冻土*, 2009, 31(3): 490-500.]
- [25] Ran Youhua, Li Xin, Lu Ling. China land cover classification at 1 km spatial resolution based on a multi-source data fusion approach. *Advances in Earth Science*, 2009, 24(2): 192-203. [冉有华, 李新, 卢玲. 基于多源数据融合方法的中国 1 km 土地覆盖分类制图. *地球科学进展*, 2009, 24(2): 192-203.]
- [26] Liu Wei, Gao Yanhong, Li Haiying et al. Landuse patterns of Heihe River Basin and its impact modeling. *Plateau Meteorology*, 2007, 26(2): 278-285. [刘伟, 高艳红, 李海英等. 黑河流域土地覆盖分类数据的建立及其影响的模拟. *高原气象*, 2007, 26(2): 278-285.]
- [27] Chen Feng, Xie Zhenghui. A land cover dataset based on Chinese vegetation data and its impact on land surface simulations. *Chinese Journal of Atmospheric Sciences*, 2009, 33(4): 681-697. [陈锋, 谢正辉. 基于中国植被数据的陆面覆盖及其对陆面过程模拟的影响. *大气科学*, 2009, 33(4): 681-697.]
- [28] Liu J, Liu M, Tian H et al. Spatial and temporal patterns of China's cropland during 1990-2000: An analysis based on Landsat TM data. *Remote Sensing of Environment*, 2005, 98(4): 442-456.
- [29] Liu J, Liu M, Deng X et al. The land use and land cover change database and its relative studies in China. *Journal of Geographical Sciences*, 2002, 12(3): 275-282.
- [30] Yellow River Conservancy Committee. Overview of the Heihe River Basin. <http://www.yellowriver.gov.cn/vh/heihe/lykk.php>, 2010-06-22. [黄河水利委员会. 黑河概览. <http://www.yellowriver.gov.cn/vh/heihe/lykk.php>, 2010-06-22.]
- [31] Gallo K, Ji L, Reed B et al. Comparison of MODIS and AVHRR 16-day normalized difference vegetation index composite data. *Geophys. Res. Lett.*, 2004, 31: L7502.
- [32] Assessment F R. Global land cover characteristics data base version 2.0. 2000.
- [33] Brown J F, Reed B C, Huewe L. Advanced strategy for multi-source analysis and visualization in land cover characterization. Proceedings, Pecora 13, Human Interactions with the Environment: Perspectives from Space, 1998:

367-382.

- [34] Wang Rongjing. SVM-based MODIS land cover classification feature select and application [D]. Beijing: China Agricultural University, 2005. [王荣静. 基于SVM的MODIS数据土地覆盖分类方法研究[D]. 北京: 中国农业大学, 2005.]
- [35] Pan Yaozhong, Li Xiaobing, He Chunyang. Research on comprehensive land cover classification in China: Based on NOAA/AVHRR and Holdridge PE index. *Quaternary Sciences*, 2000, 20(3): 270-281. [潘耀忠, 李晓兵, 何春阳. 中国土地覆盖综合分类研究: 基于NOAA/AVHRR和Holdridge PE. *第四纪研究*, 2000, 20(3): 270-281.]
- [36] Du Mingyi, Wu Wenbo, Guo Dazhi. Research on multi-source geographic information based classification of desertification. *Journal of Image and Graphics*, 2002, 7(7): 740-743. [杜明义, 武文波, 郭达志. 多源地学信息在土地荒漠化遥感分类中的应用研究. *中国图象图形学报: A辑*, 2002, 7(7): 740-743.]
- [37] Li Daofeng, Tian Yuying, Hao Fanghua. NDVI data based study on complex classification of vegetation cover of Yellow River Basin. *Research of Soil and Water Conservation*, 2003, 10(4): 88-91. [李道峰, 田英, 郝芳华. 基于NDVI数据的黄河流域地表植被覆盖综合分类研究. *水土保持研究*, 2003, 10(4): 88-91.]
- [38] Yu Q, Gong P, Tian Y Q et al. Factors affecting spatial variation of classification uncertainty in an image object-based vegetation mapping. *Photogrammetric Engineering & Remote Sensing*, 2008, 74(8): 1007-1018.
- [39] Wang Jinye, Che Kejun, Yan Kelin et al. Analysis of the runoff components in the forestry areas of the Qilian Mountains and temporal and spatial variation. *Journal of Glaciology and Geocryology*, 1999, 21(1): 59-63. [王金叶, 车克钧, 闰克林等. 祁连山森林覆盖区河川径流组成与时空变化分析. *冰川冻土*, 1999, 21(1): 59-63.]
- [40] Chen Rensheng, Kang Ersi, Ji Xibin et al. Preliminary study of the hydrological processes in the alpine meadow and permafrost regions at the Headwaters of Heihe River. *Journal of Glaciology and Geocryology*, 2007, 29(3): 387-396. [陈仁升, 康尔泗, 吉喜斌等. 黑河源区高山草甸的冻土及水文过程初步研究. *冰川冻土*, 2007, 29(3): 387-396.]

A Rule-based Land Cover Classification Method for the Heihe River Basin

HOU Yuting¹, WANG Shugong², NAN Zhuotong³

(1. Center for Hydrologic Cycle and Water Resources Research in Arid Region, College of Earth and Environmental Science, Lanzhou University, Lanzhou 730000, China;

2. Department of Civil and Environmental Engineering, University of Pittsburgh, Pittsburgh, PA, 15261, USA;

3. Cold and Arid Regions Environmental and Engineering Research Institute, CAS, Lanzhou 730000, China)

Abstract: A novel rule-based land use/land cover classification approach is presented in this study. Rule tables were generated based on geographic characteristics of each class of the China land use classification schema and its possible transferability into other classes of the USGS schema. The USGS land use/land cover (LULC) data product, with a 1-km spatial resolution, was used to locate clustering centers, referred as NDVI fingerprints, of each land use class. A minimum distance approach was then applied to the 1 km NDVI of the year 2009 and 90 m DEM of the Heihe River Basin (HRB), with rule tables considered, to produce a land use/land cover map with schema and attributes consistent with USGS's. The produced map can be used in atmospheric models and land surface models. A comparison to the previous work and satellite images indicates that our rule-based approach is better in distinguishing land cover characteristics, especially for snow-cover, frozen soil and desert types.

Key words: Heihe River Basin; land use/land cover; rule-based classification approach; USGS LULC dataset; the minimum distance classifier