2017 年第 2 卷第 3 期

文献 DOI：
10.11922/csdata.170.2017.0131
数据 DOI：
10.11922/sciencedb.382
文献分类：地球科学

收稿日期：2017-03-17
开放同评：2017-04-28
发表日期：2017-08-26

2015 年“一带一路”25 个港口城市及其周边区域土地覆盖分类数据集

宋洋[1,2]，侯西勇[1*]

1. 中国科学院烟台海岸带研究所，烟台 264003；
2. 中国科学院大学，北京 100049

*通讯作者（Email: xyhou@yic.ac.cn）

摘要：利用 2015 年成像的 Landsat 8 OLI 多光谱影像，参考全球 30 m 土地覆盖数据（GlobeLand30）的分类标准、制图规范，建立“一带一路”港口城市土地覆盖分类系统，运用 ENVI 5.2 软件，基于支持向量机的监督分类方法，对“一带一路”海上丝绸之路沿线的 25 个重要港口城市及其周边区域的土地覆盖进行遥感分类，经过严格的修改与精度验证工作，得到土地覆盖分类数据。本数据集可用于土地利用变化、生态环境变化研究，能够清晰反映港口城市建成区及其周边地区的生态环境现状特征。

关键词：一带一路；港口城市；土地覆盖；生态环境

数据库（集）基本信息简介

数据库（集）中文名称	2015年“一带一路”港口城市及其周边区域土地覆盖分类数据集		
数据库（集）英文名称	A dataset of land cover classification in 25 port cities and their surrounding areas along “One Belt and One Road” (2015)		
数据作者	宋洋、侯西勇		
通讯作者	侯西勇（xyhou@yic.ac.cn）		
数据时间范围	2015年		
地理区域	25个港口城市及其周边区域。港口城市包括上海、釜山、新加坡、雅加达、皎漂、曼谷、关丹、孟买、瓜达尔、加尔各答、吉大港、科伦坡、多哈、阿巴斯、吉达、迪拜、里斯本、圣彼得堡、吉布提、苏丹港、亚历山大、比雷埃夫斯、伊斯坦布尔、悉尼、达尔文。		
空间分辨率	30 m	数据量	28.7 MB
数据格式	*.tif，*.tfw，*.dbf，*.ovr，*.xml		
数据服务系统网址	http://www.sciencedb.cn/dataSet/handle/382		
基金项目	国家自然科学基金项目（31461143032）		
数据集组成	本数据集主要包括25个港口城市及其周边区域的2015年土地覆盖分类数据，这些数据保存为1个压缩文件（“一带一路”港口城市及其周边区域土地覆盖分类数据集.rar），每个城市的数据单独存贮在以城市命名的文件夹里，总数据量28.7 MB。		

引　言

为响应“一带一路”战略构想，科技部国家遥感中心将“一带一路”生态环境状况遥感监测作为“全球生态环境遥感监测”2015 年度报告的主题。港口是“一带一路”基础设施建设的重点之一，对“海上丝绸之路经济带”沿线重要港口城市生态环境现状进行遥感监测意义突出。因此，以全球海运经济活动的地理分区特征、港口区位特征与发展现状为基础，重点考量港口在海上丝绸之路沿线城市经济发展中的巨大经济辐射与交通枢纽作用，同时，根据其在“一带一路”建设过程中发挥的作用大小及其与中国合作的密切程度，选择 25 个港口城市，对港口城市及其周边地区进行土地覆盖分类研究，得到“一带一路”港口城市及其周边区域土地覆盖分类数据集，并提供数据共享。相关研究成果作为“全球生态环境遥感监测”2015 年度报告的一部分，反映“一带一路”沿线港口城市的生态环境发展状况。

1　数据采集和处理方法

1.1　数据来源

本研究以美国地质勘探局（http://glovis.usgs.gov/）提供下载的 2015 年 Landsat 8 OLI 的 30 m 分辨率多光谱影像为数据源。考虑到港口城市近海的地理位置，主要下载夏秋季节无云或少云的数据。表 1 所示的是每个港口城市所用的遥感卫星影像序列号及成像时间信息。

表 1　港口城市 Landsat 8 影像数据源

港口城市	影像序列号	港口城市	影像序列号
上海	LC81180382015215LGN00 LC81180392015215LGN00	苏丹	LC81710462015202LGN00 LC81710472015202LGN00
釜山	LC81140352015155LGN00 LC81140362015155LGN00 LC81150352015146LGN00 LC81150362015146LGN00	曼谷	LC81280512015141LGN00 LC81290502015308LGN00 LC81290512015308LGN00 LC81300502015075LGN00
迪拜	LC81590432015262LGN00 LC81600422015269LGN00 LC81600432015269LGN00 LC81610422015228LGN00	雅加达	LC81220642015243LGN00 LC81220652015227LGN00 LC81230642015266LGN00 LC81230652015218LGN00
亚历山大	LC81770382015244LGN00 LC81770392015244LGN00 LC81780382015267LGN00	圣彼得堡	LC81840182015229LGN00 LC81840192015229LGN00 LC81850182015236LGN00

港口城市	影像序列号	港口城市	影像序列号
	LC81780392015267LGN00		LC81850192015236LGN00
加尔各答	LC81380442015323LGN00 LC81380452015323LGN00 LC81390442015298LGN00 LC81390452015298LGN00	阿巴斯	LC81590412015262LGN00 LC81590422015262LGN00 LC81600412015269LGN00 LC81600422015269LGN00
吉达	LC81690442015268LGN00 LC81690452015268LGN00 LC81690462015268LGN00 LC81700452015259LGN00 LC81700462015259LGN00	伊斯坦布尔	LC81790312015194LGN00 LC81790322015210LGN00 LC81800312015249LGN00 LC81800322015249LGN00 LC81810312015240LGN00
里斯本	LC82040332015177LGN00	吉大港	LC81360442015309LGN00 LC81360452015293LGN00
多哈	LC81620422015235LGN00 LC81620432015235LGN00 LC81630422015274LGN00 LC81630432015258LGN00	孟买	LC81470462015274LGN00 LC81470472015290LGN00 LC81480462015281LGN00 LC81480472015281LGN00
新加坡	LC81250592015152LGN00	瓜达尔	LC81550432015266LGN00
悉尼	LC80890832015059LGN00 LC80890842015059LGN00 LC80900832015338LGN00 LC80900842015338LGN00	达尔文	LC81060682015131LGN00 LC81060692015147LGN00
		关丹	LC81260572014060LGN00 LC81260582015207LGN00
科伦坡	LC81410552015008LGN00 LC81410562015056LGN00 LC81420552015047LGN00	比雷埃夫斯	LC81820342015231LGN00 LC81830332015190LGN00 LC81830342015190LGN00
皎漂	LC81340472015327LGN00	吉布提	LC81660522015279LGN00

1.2 数据处理方法

遥感图像计算机分类是遥感技术应用的一个重要组成部分。传统的遥感图像计算机分类技术包括最大似然法、马氏距离法、最小距离法等，新兴的分类技术有神经网络、模糊聚类、决策树分类等算法[1-2]，这些算法在某种程度上对样本依赖性较高、算法稳定性较差、分类精度不高。而 SVM 方法是建立在统计学习理论的 VC 维（Vapnik-Chervonenkis Dimension）理论和结构风险最小原理基础上的一种非常有效的机器学习新方法，是近年来发展起来的一种新的人工智能算法，能从很大程度上解决上述分类方法的缺点和不足。其中，VC 维表征一个指数函数集，h 个样本将被函数以 2^h 种形式分开，VC 维越高，置信界限越大，机器学习的复杂性越高，经验风险也就相对较高[3]。支持向量机（Support Vector Machine，SVM）

可以使 VC 维尽量小，使结构风险最小化，它能根据有限的样本信息在模型的复杂性和学习能力之间寻求最佳的平衡，以期获得最好的推广能力，较好地解决了以往学习方法中存在的小样本、非线性、过学习、高维数、局部极小点等问题，具有很强的泛化能力，在降低样本依赖性、提高算法稳定性和分类精度等方面有更好的表现，目前被广泛地利用于遥感影像计算机分类等研究领域 [4]。因此，本文选择遥感图像解译领域比较先进的支持向量机方法。

本文参考全球 30m 土地覆盖数据集（http://globallandcover.com/GLC30Download/index.aspx）的土地分类系统确立重要港口城市土地覆盖分类系统，如表 2 所示。利用 ENVI5.2 软件，采用基于支持向量机的监督分类方法[5-8]按照图 1 所示流程对每个城市及其周边地区的土地覆盖进行遥感分类，然后参考高分二号卫星和 Google Earth 高分辨率卫星影像对分类结果进行修正和结果验证，最终生成 2015 年“一带一路”重要港口城市土地覆盖数据集。本数据集制作所参考的 Globeland 30 有 2000 年和 2010 年的数据产品，但没有对建成区内部的城市绿地覆盖进行解译；本数据集为了能够更好地反映连续建成区内部的生态环境，对建成区内部的绿地覆盖也进行了遥感解译，并且遥感影像数据时相为 2015 年，能够反映 25 个港口城市近几年的生态环境状况。

表 2　土地覆盖类型及定义

代码	类型	定义
10	农田	用于种植农作物的土地，包括水田、灌溉旱地、雨养旱地、菜地、牧草种植地、大棚用地、以种植农作物为主间有果树及其他经济乔木的土地；以及茶园、咖啡园等灌木类经济作物种植地。
20	森林	乔木覆盖且树冠盖度超过 30%的土地，包括落叶阔叶林、常绿阔叶林、落叶针叶林、常绿针叶林、混交林；以及树冠盖度为 10%～30%的疏林地。
30	草地	天然草本植被覆盖，且盖度大于 10%的土地，包括草原、草甸、稀树草原、荒漠草原；以及城市人工草地等。
40	灌丛	灌木覆盖且灌丛覆盖度高于 30%的土地，包括山地灌丛、落叶和常绿灌丛；以及荒漠地区覆盖度高于 10%的荒漠灌丛。
50	湿地	位于陆地和水域的交界带，有浅层积水或土壤过湿的土地，多生长有沼生或湿生植物。包括内陆沼泽、湖泊沼泽、河流洪泛湿地、森林/灌木湿地、泥炭沼泽、红树林、盐沼等。
60	水体	液态水覆盖的区域，包括江河、湖泊、水库、坑塘、海洋等；
80	不透水层	指能够阻止水直接渗透到土壤中的任何自然和人工表面，主要由城市用地中的交通用地、建筑用地、工矿用地和建筑物屋顶等组成。
90	裸地	植被覆盖度低于 10%的自然覆盖土地，包括荒漠、沙地、砾石地、裸岩、盐碱地等。

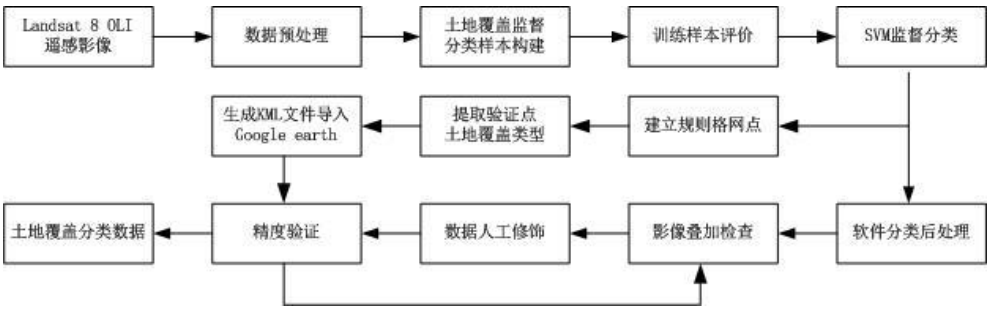

图 1　土地覆盖分类数据制作流程

2　数据样本描述

2.1　数据组成

本数据集包括 25 个港口城市及其周边区域的 2015 年土地覆盖分类数据，这些数据保存为 1 个压缩文件（“一带一路”港口城市及其周边区域土地覆盖分类数据集.rar），每个城市的数据单独存贮在以城市命名的文件夹里，总数据量 28.7 MB。每个文件夹中单独存储相应的 tif 栅格数据文件，为 8 Bit 栅格数据结构，以整型数据表达属性信息；采用的坐标系统为 UTM-WGS84。

2.2　数据样本

基于支持向量机的监督分类方法，经过修正与精度验证工作，完成 2015 年“一带一路”25 个港口城市及周边地区的土地覆盖分类数据。图 2 为釜山及其周边地区的土地覆盖分类数据样本。土地覆盖遥感分类结果表明，釜山连续建成区及其边界做 50km 缓冲区范围的周边陆地区域主要土地覆盖类型是森林，面积为 3875.08 km^2，占 69.01％；其次是不透水层，面积为 710.7 km^2，主要分布在城市建成区内，占 12.66%；农田面积为 694.09 km^2，主要分布在下游河流入海口和山谷区域，面积为 694.09 km^2，占 12.36％。连续建成区的面积为 253.39 km^2，以不透水层和植被覆盖为主，面积分别为 154.16 km^2 和 96.11 km^2，占连续建成区的比例分别为 60.84％和 37.93％。其他类型的土地覆盖面积相对较小，且空间分布较为零散。

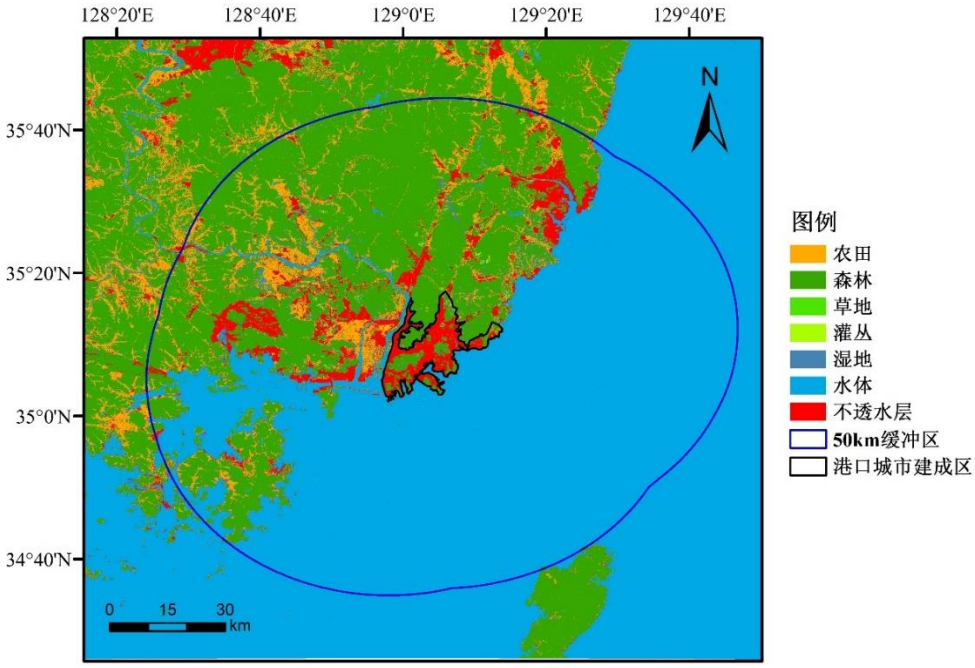

图 2　2015 年釜山及其周边土地覆盖分类图

3　数据结果与验证

3.1　数据结果与精度验证技术路线

基于 Google earth 的高分辨率影像进行目视解译，所得数据用于分析 2015 年“一带一路”港口城市及其周边区域土地覆盖遥感制图结果的精度特征，技术路线如图 3 所示。具体过程为：（1）利用 ArcGIS 中的 Generate 工具，在各港口城市建成区以及缓冲区内生成等间距采样点;（2）将点数据与解译结果图进行叠加，在点属性表中添加新的属性字段记录样点落入的多边形的属性信息；（3）根据样点的经纬度坐标，基于 Google Earth 高空间分辨率影像判读样点的实际土地覆盖类型；（4）基于混淆矩阵计算总精度和 Kappa 系数，分别对 25 个港口城市的土地覆盖解译结果进行精度评价。

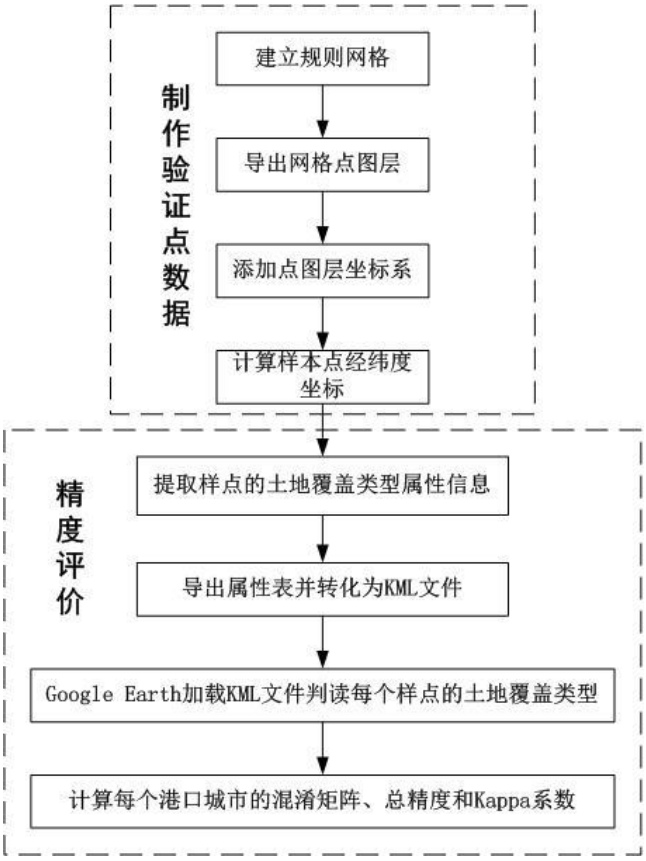

图 3　土地覆盖分类数据验证技术流程图

3.2　数据结果验证方法

混淆矩阵为最常用的土地利用/覆盖遥感制图精度评价方法，根据混淆矩阵能够计算出总体精度和 Kappa 系数。混淆矩阵是由 n 行 n 列组成的矩阵，n 为类别数。混淆矩阵的列方向依次为实际类别的第 1 列，第 2 列，…，第 n 类的代码或者名称；矩阵的行方向依次为分类结果各类别的第 1 列，第 2 列，…，第 n 类的代码或者名称；矩阵中的元素是指各类的样点数。

总体分类精度等于被正确分类的样点总和除以总样点数，被正确分类的样点沿着混淆矩阵的对角线分布，它显示出被分类到正确地表真实分类中的样点数。而 Kappa 系数能够更全面、更客观地反映图像分类的总体精度。计算公式为：

$$P = \frac{\sum_{i=1}^{n} x_{ii}}{N} \tag{1}$$

$$K = \frac{N\sum_{i=1}^{n} x_{ii} - \sum_{i=1}^{n} x_{i+}x_{+i}}{N^2 - \sum_{i=1}^{n} x_{i+}x_{+i}} \tag{2}$$

式中，P 为总体精度，N 是用于精度评估的总样点数；n 为混淆矩阵中的总列数；

x_{ii} 为混淆矩阵中第 i 行、第 i 列上样点数；x_{i+} 和 x_{+i} 分别是第 i 行和第 i 列的总像素数量，K 为 Kappa 系数。

3.3 数据结果验证精度

“一带一路”港口城市及其周边区域 2015 年土地覆盖遥感分类数据的精度评价结果如表 3 所示。可以看出，25 个港口城市的土地覆盖遥感分类结果的总精度都较高，除了加尔各答为 88.94%，其他均在 90%以上；Kappa 系数普遍在 0.8 以上，多数城市的 Kappa 系数在 0.85 以上。综上，25 个港口城市及其周边区域的土地覆盖分类数据精度总体较好。

表 3 “一带一路”港口城市土地覆盖分类数据集精度验证结果

城市	总精度（%）	Kappa系数	样本数
科伦坡	97.32	0.91	149
阿巴斯	96.10	0.88	154
比雷埃夫斯	90.50	0.90	180
达尔文	96.00	0.94	142
釜山	94.19	0.88	196
迪拜	95.83	0.90	210
瓜达尔	97.92	0.96	192
多哈	96.63	0.82	196
关丹	97.41	0.83	206
吉布提	93.00	0.88	100
吉达	96.23	0.85	210
吉大港	93.15	0.89	146
皎漂	92.11	0.89	114
曼谷	93.59	0.87	156
上海	93.25	0.90	164
孟买	90.77	0.87	130
加尔各答	88.94	0.81	208
圣彼得堡	95.23	0.93	210
苏丹港	97.63	0.79	127
里斯本	93.13	0.91	133
新加坡	97.06	0.95	102
悉尼	97.04	0.94	135
亚历山大	96.94	0.95	98
雅加达	91.52	0.88	165
伊斯坦布尔	97.37	0.96	190

4 数据使用方法和建议

可利用 ArcGIS 等地理信息系统软件对本数据集进行编辑、查询和分析。本数据集代表性强，可以作为“一带一路”沿线重点港口城市生态环境监测的基础数据集，应用于土地利用/覆盖变化、生态环境变化等研究。

5 数据作者分工职责

宋洋（1988—），男，山东乳山人，博士研究生，研究方向为海岸带土地利用变化与海岸线变化遥感监测。主要承担工作：Landsat 8 OLI 影像数据采集、基于向量机的监督分类、土地覆盖分类数据精度验证与修改、分类后数据的分析等。

侯西勇（1975—），男，山东泰安人，博士，研究员，博士生导师，研究方向为海岸带土地利用变化与海岸线变化遥感监测、脆弱性评价及综合管理等方面的研究。主要承担工作：本数据集生产和质量控制的整体思路设计。

致　谢

感谢科学技术部国家遥感中心邀请参与“全球生态环境遥感监测”2015 年度报告的 “21 世纪海上丝绸之路”港口城市部分工作；同时也感谢国家遥感中心的工作人员在研究过程中所做的大量协调性工作和具体帮助；感谢王远东、刘静、王俊惠、魏辽生、王晓利、侯婉等在技术方案讨论、基础数据下载与预处理等方面所做的工作和贡献。

参考文献

[1] 王一达, 沈熙玲, 谢炯. 遥感图像分类方法综述[J]. 遥感信息, 2006 (5): 67–71.

[2] 李石华, 王金亮, 毕艳, 等. 遥感图像分类方法研究综述[J]. 国土资源遥感, 2005 (2): 1–6.

[3] 袁玉萍, 陈庆华, 汪洪艳. 关于支持向量机 VC 维问题证明的研究 [J].农业与技术, 2006, 26(3): 210–211.

[4] 王小明, 毛梦祺, 张昌景, 等. 基于支持向量机的遥感影像分类比较研究[J]. 测绘与空间地理信息, 2013, 36(4): 17–20, 23.

[5] 刘炜. 土地利用/覆被变化信息遥感图像自动分类识别与提取方法研究[D]. 咸阳: 西北农林科技大学, 2012.

[6] 傅文杰, 洪金益, 林明森. 基于光谱相似尺度的支持向量机遥感土地利用分类[J]. 遥感技术与应用, 2006, 21(1): 25–30.

[7] 荆鑫, 舒清态, 刘庆生. 基于支持向量机的 TM 遥感影像土地利用/覆盖类型监测[J]. 安徽农业科学, 2014, 42(22): 7631–7632.

[8] 杨青生, 黎夏. 基于支持向量机的元胞自动机及土地利用变化模拟[J]. 遥感学报, 2006, 10(6): 836–846.

论文引用格式

宋洋, 侯西勇. 2015 年“一带一路”25 个港口城市及其周边区域土地覆盖分类数据集[J/OL]. 中国科学数据, 2017, 2(3). DOI: 10.11922/csdata.170.2017.0131.

数据引用格式

宋洋, 侯西勇. 2015 年“一带一路”25 个港口城市及其周边区域土地覆盖分类数据集[DB/OL]. Science Data Bank, 2017. DOI: 10.11922/sciencedb.382.

A dataset of land cover classification in 25 port cities and their surrounding areas along “One Belt and One Road” (2015)

Song Yang[1,2], Hou Xiyong[1*]

1. Yantai Institute of Coastal Zone Research, Chinese Academy of Sciences, Yantai 264003, P. R. China;

2. University of Chinese Academy of Sciences, Beijing 100049, P. R. China

*Email: xyhou@yic.ac.cn

Abstract: Based on Landsat 8 OLI satellite images captured in 2015, and referring to the classification criteria and technical specifications of GlobeLand30 dataset, Support Vector Machine (SVM) provided by ENVI 5.2 has been used to obtain the dataset of land cover classification in 25 port cities and their surrounding areas along “One Belt and One Road”, furthermore, strict precision verification and data modification have been carried out by collecting verification samples from high resolution remote sensing images such as Google maps. The status quo of ecological environment of 25 port cities and their surrounding areas could be reflected clearly based on this dataset, and this dataset could also be applied to researches such as changing characters of land use and environment.

Keywords: One Belt and One Road; port cities; land cover; ecological environment