专题 海南资源环境遥感产品数据集

三亚市热异常遥感监测数据集

孟庆岩[1*]，谷艳春[1,2]，郝丽春[1,3,4]，胡蝶[1,3,5]，张颖[1,3,5]，张琳琳[1,3,5]

1. 中国科学院遥感与数字地球研究所，北京 100101
2. 福州大学福建省空间信息工程研究中心，福州 350000
3. 海南省地球观测重点实验室，海南三亚 572029
4. 河南理工大学测绘与国土信息工程学院，河南焦作 454000
5. 中国科学院大学资源与环境学院，北京 100049

摘要：针对三亚市热异常问题，本文以2008–2017年Landsat数据为主要数据源，基于辐射方程传输法反演地表温度和改进的箱线图法提取高温异常区，进行高温异常区频率统计，将频率大于60%的高温异常区判定为城市热异常区，得到10年间城市热污异常区。本数据集可应用在其他城市，进行多区域城市热异常区提取，从而发挥更高的应用价值。

关键词：热异常；辐射方程传输法；改进的箱线图法；高温异常区频率

数据库（集）基本信息简介

数据库（集）名称	三亚市热异常遥感监测数据集
数据作者	孟庆岩、谷艳春、郝丽春、胡蝶、张颖、张琳琳
数据通信作者	孟庆岩（mengqy@radi.ac.cn）
数据时间范围	2008–2017年
地理区域	三亚市（18°09′34″–18°37′27″N、108°56′30″–109°48′28″E）
空间分辨率	Landsat 5：30/120 m Landsat 7：30/60 m Landsat 8：30/100 m
数据量	4.05 MB
数据格式	*.vsdx,*.jpg,*.png, *.py, *.shp
数据服务系统网址	http://www.sciencedb.cn/dataSet/handle/700
基金项目	海南省自然科学基金（417219）；海南省重大科技计划（ZDKJ2016021）；四川省科技计划（2018JZ0054）；海南省重大科技计划（ZDKJ2017009）。
数据库（集）组成	本数据集主要包括7个文件夹，每个文件夹的内含文件格式和命名方式如下： 1. 封面图文件夹：内含命名为封面图*.jpg格式的数据； 2. 三亚市边界图文件夹示意图：内含格式为*.jpg的三亚市边界示意图数据；

www.csdata.org

ISSN 2096-2223

CN 11-6035/N

文献 DOI:

10.11922/csdata.2018.0077.zh

数据 DOI：

10.11922/sciencedb.700

文献分类：地球科学

收稿日期：2018-11-19

开放同评：2018-12-12

录用日期：2019-03-25

发表日期：2019-05-16

* 论文通信作者

孟庆岩：mengqy@radi.ac.cn

数据库（集）组成	3. 数据处理流程图文件夹：内含命名为数据处理流程图*.vsdx格式的数据； 4. 2008–2017年三亚市年度热异常提取结果图文件夹：内含2008–2017年*.png格式的逐年三亚市年度热异常提取结果图； 5. 2008–2017年三亚市年度热异常提取结果元数据文件夹：内含2008–2017年*.shp格式的逐年三亚市年度热异常提取结果元数据； 6. 代码文件夹：内含本数据集处理流程用到的*.py格式代码； 7. 三亚市矢量边界文件夹：内含*.shp格式的三亚市边界数据。

引　言

热异常是指由自然因素和人类活动中的热排放导致环境温度异常升高的现象[1]。城镇化进程的快速推进可带动经济发展，但同时也对水资源、局部气候、生存环境等产生了一系列负面效应。据统计，近 100 a 来，全球的年平均气温上升了 0.7–1℃，而大城市的平均气温则上升了 2–3℃，北京、上海、东京等世界性大都市的年均气温比郊区高出 2℃，夏季高出 6℃左右[2]。与物质污染带给环境的影响不同的是，物质污染是由污染物引起的，可以通过具体方法和措施得到控制，而热异常则不同，它主要来源于热源，有能源消耗则会产生热异常[3]。因此，热异常是不可避免的。针对此类现象，国内外学者已经做了大量探索和研究，例如 Wu 和 Chen 等通过监测不同地区的核电站热流排放情况评估热异常强度和对周边环境的影响[4-5]。Liu 和 Xia 等基于热异常产品，通过利用不同方式提取工业热源[6-7]；Zhang 等利用 Landsat 数据对首钢搬迁前后的热环境进行探索和分析[8]。还有一部分学者利用城市热岛强度与气温间的关系评估城市热异常[9-10]。可见，热异常已被人们广泛关注，并成为研究热点。但以往关于热异常的大量文献多针对工业城市，或有工业污染的地方，针对非工业城市的热异常研究还较少。而热异常不仅存在于工业城市，还存在于非工业城市。因此，针对非工业城市，如何有效快速提取热异常区，如何缓解改善热异常具有重要研究意义。

三亚市作为我国唯一一个热带滨海旅游城市，其生态环境保护尤显重要。恰逢海南省推进“国际旅游岛”建设，作为海南旅游发展龙头的三亚正以国际化标准致力于打造“一港两地”的国际性滨海旅游城市。但伴随着三亚市城镇化的不断发展，城市热异常对城市环境的影响日益突出。因此，针对三亚城市建设和区域、环境、植被等特点，开展具有针对性的城市热异常研究，对三亚城市建设、规划布局、环境保护均具有重要意义。

综上，针对如何监测三亚市热异常区，并提取热异常区等问题，本文基于长时间序列的 Landsat 数据，利用辐射方程传输法获取地表温度信息，结合改进的箱线图法提取相应高温异常区，通过统计分析高温异常区的年度出现频率，得到城市热异常区。该方法不仅能够准确大面积获取城市热异常区，且相较于其他方法，可以有效避免潜在热异常的影响。

1　数据采集和处理方法

1.1　区域范围

三亚市位于北纬 18°09′34″–18°37′27″，东经 108°56′30″–109°48′28″，地处海南省最南端，主要以旅游业和农业为主，是海南省第二大城市。作为海南省最重要的旅游城市之一，其城市环境一直是广大群众关注的热点。但随着城市的不断发展与扩张，城市热异常现象日益加剧，因此，选取三亚

市市辖区作为研究区，研究示意图如图 1 所示。

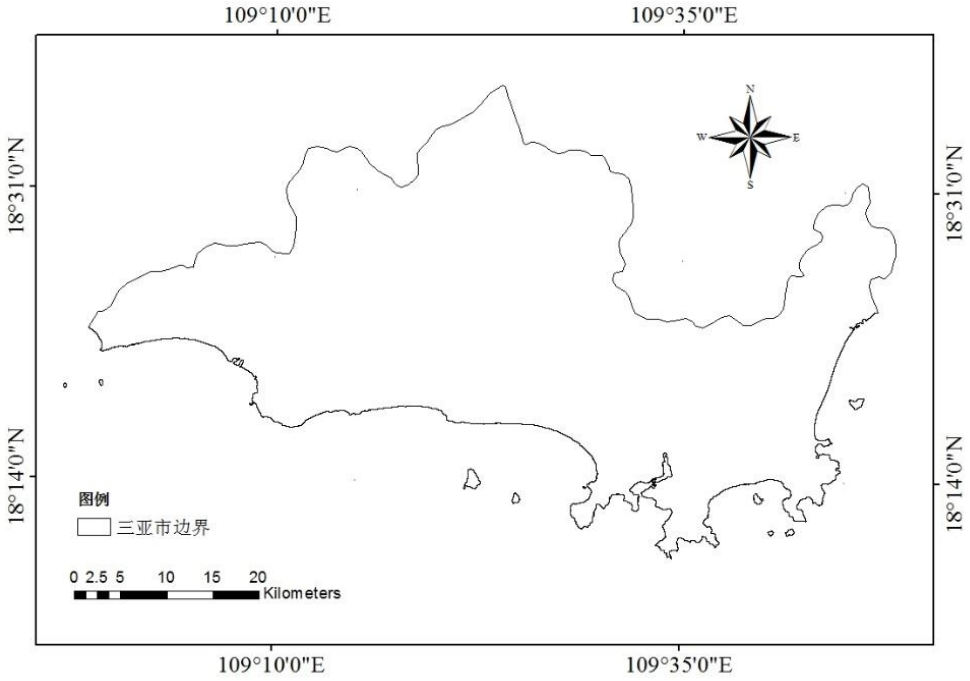


图 1　三亚市边界示意图

1.2　基础数据准备

本文以 2008–2017 年间所有云量小于 30%的 Landsat 5、Landsat 7 和 Landsat 8 数据为主要数据源（获取网址：https://glovis.usgs.gov）。本文温度反演所用到的波段及其空间分辨率如表 1 所示，其中，因 Band11 运行出现问题文中 Landsat 8 温度反演用到的是波段 10。表 2 为本文所用影像的具体时间信息。

表 1　温度反演所用波段信息

数据源	所用波段	空间分辨率/m
Landsat 5	Band3	30
	Band4	30
	Band6	120
Landsat 7	Band3	30
	Band4	30
	Band6	60
Landsat 8	Band4	30
	Band5	30
	Band10	100

表 2　数据集所用影像日期

年份	影像获取日期
2008 年	20080419、20081028、20081113、20081129、20081215
2009 年	20090820、20091210
2010 年	20110316、20100706、20100714、20101026、20101127、20101213
2011 年	20111021
2012 年	20120321、20120913、201209220121015
2013 年	20130519、20130620、20130924、20131010
2014 年	20140114、20140420、20140924、20141013、20141114
2015 年	20150407
2016 年	20160308、20160916、20161103、20161221
2017 年	20170106、20170122、20170311、20170701、20170802、20171021、20171208、20171224

1.3　数据处理

首先对 Landsat 数据进行拼接、裁剪、重投影等数据预处理。其中 Landsat 7 ETM+影像因有条带，需在数据预处理基础上再进行去条带处理。其次用辐射方程传输法反演地表温度和改进的箱线图法提取高温异常区，得到高温异常区空间分布图。最后，为避免漏判现象，借鉴 Liu 等[6]提取工业热源的思路，以频率大于 60%为热异常区的标准统计分析热异常出现的频率，得到高温异常区频率空间分布图。同时，统计城市热异常区的总面积，具体数据处理流程如图 2 所示。

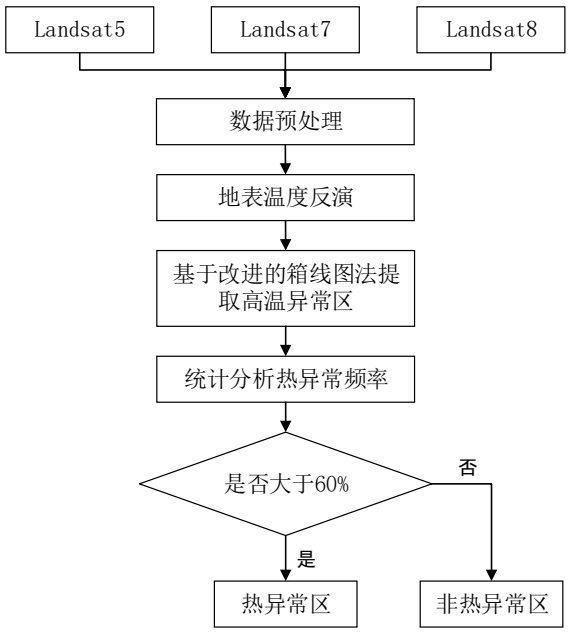


图 2　数据处理流程图

1.3.1　地表温度反演

基于 Landsat 卫星影像，根据 Landsat 数据热红外波段的增益值与偏置值计算相应表观辐射亮度

值；然后，利用影像植被覆盖度估算地表比辐射率；最后，根据 Planck 反函数和 Landsat 预设的定标系数计算地表温度。具体实现如下：

（1）表观辐亮度计算

表观辐亮度计算是将影像像元灰度值转换成相应热辐射强度的过程。利用 ENVI 软件中的辐射定标（Radiometric Calibration）工具将影像的像元灰度值转换为辐射亮度值。

（2）地表比辐射率估算

地表比辐射率表征地面物体的电磁波辐射的一种能力。基于影像的近红外波段和红光波段，根据公式（1）计算归一化植被指数 *NDVI*[11]。

$$NDVI = (TM_{\mathrm{NIR}} - TM_{\mathrm{Red}})/(TM_{\mathrm{NIR}} + TM_{\mathrm{Red}}) \quad (1)$$

式中：TM_{NIR}为影像的近红外波段反射率；TM_{Red}为影像的红光波段反射率。

其次，利用影像植被覆盖度估算地表比辐射率。由于地表物质结构具有差异性，因此针对不同地物类型，其相应的地表比辐射率计算方式也不同。大致可分为 3 种类型：水体、城镇和自然表面。而水体像元的地表比辐射率很高，与黑体的比辐射率较为相似，因此，在估算水体比辐射率时常被赋值为 0.995，自然表面和城镇像元的地表比辐射率估算分别根据公式（2）和公式（3）计算[12-14]：

$$e_{\mathrm{Surface}} = 0.9625 + 0.0614P_{\mathrm{Vege}} - 0.0461P_{\mathrm{Vege}}{}^{2} \quad (2)$$

$$e_{\mathrm{Building}} = 0.9589 + 0.086P_{\mathrm{Vege}} - 0.0671P_{\mathrm{Vege}}{}^{2} \quad (3)$$

式中：e_{Surface}代表自然表面像元的地表比辐射率；e_{Building}代表城镇像元的地表比辐射率；P_{Vege}代表植被覆盖度，计算方法如公式（4）所示[15-16]：

$$P_{\mathrm{Vege}} = \left[\frac{NDVI - NDVI_{\mathrm{Soil}}}{NDVI_{\mathrm{Vege}} - NDVI_{\mathrm{Soil}}}\right]^{2} \quad (4)$$

式中：$NDVI$是影像的归一化植被指数，$NDVI_{\mathrm{Vege}}$是影像中纯植被像元的 *NDVI* 值，$NDVI_{\mathrm{Soil}}$是影像中纯土壤像元的 *NDVI* 值。当某个像元的$NDVI$大于 0.70 时，P_{Vege}取值为 1；当$NDVI$小于 0.05 时，P_{Vege}取值为 0；当$NDVI$介于 0.05–0.7 之间时，则分别将$NDVI_{\mathrm{Vege}}$和$NDVI_{\mathrm{Soil}}$取值 0.70 和 0.05，并利用上式估算影像的植被覆盖度。

（3）地表温度计算

基于热红外波段的表观辐亮度和地表比辐射率计算结果，根据辐射传输方程反推地表热辐射亮度，如公式（5）所示：

$$B(T_{\mathrm{Sens}}) = \frac{L - L^{\uparrow} - \tau(1-\varepsilon)L_{\downarrow}}{\tau\varepsilon} \quad (5)$$

式中：$B(T_{\mathrm{Sens}})$是地表热辐射亮度；L 是行星亮度温度值；$L^{\uparrow}$和$L_{\downarrow}$分别是大气上行辐射和大气下行辐射，可根据 Landsat 卫星过境时间、兴趣区中心经纬度、大气模式及传感器类型等获得，具体参数可通过 NASA 相关网页直接获得（http://atmcorr.gsfc.nasa.gov/）；τ为大气路径透过率，可由 NASA 相关网页查询获取；ε为地表比辐射率。

利用 Plank 反函数结合 Landsat 预设的定标常数 K_1 和 K_2 计算地表真实温度[17-18]，如公式（6）所示：

$$T_{\mathrm{Surface}} = \frac{K_2}{\ln\left(\frac{K_1}{\mathrm{B(T_{Sens})}} + 1\right)} - 273.15 \quad (6)$$

式中：T_{Surface}是地表真实温度；K_1和K_2为 Landsat 预设的定标常数，可从影像头文件中获取。

为避免多时相数据带来的影响，将多时相地表温度数据逐像元进行归一化处理，具体实现如公式（7）所示：

$$Image = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (7)$$

式中：$Image$代表归一化处理后的影像像元值，像元值处于 0 到 1 之间；X代表影像地表温度值；$X_{\max}$和$X_{\min}$分别代表影像地表温度的最大值与最小值。

1.3.2 基于改进的箱线图法提取热异常

改进的箱线图法共有 5 个统计量，分别为上下四分位数、上下非异常值截距线及中位数。其中，上下非异常值截距线用于区分正常值与异常值，处于非异常值截距线之内的数据为正常值，否则为异常值。该方法与统计学中最常见的箱线图法不同之处主要有两点：一是引入鲍利系数，将鲍利系数以“1”的形式作为样本数据的偏度调控乘数。当数据集是正态分布时，鲍利系数集中在 0 附近，当数据是右偏态分布时，鲍利系数取值为 1；当数据是左偏态时，鲍利系数取值为−1。二是利用半四分位距计算非异常值截距线，可以更好地适用于具有偏态的数据[19]。该方法的具体实现如下：

（1）获取上下四分位数以及半四分位距

通过 IDL 程序逐像元获取地表温度数据，然后利用 SPSS 软件获取相应数据的上下四分位数和中位数，即将数据从小到大排序后，分别处于 25%、50%和 75%位置的数值。

之后，基于根据四分位数相应的数值，利用 EXCEL 软件计算上下半四分位距，如公式（8）和公式（9）所示[20]：

$$SIQR_{\text{down}} = Q_2 - Q_1 \quad (8)$$

$$SIQR_{\text{up}} = Q_3 - Q_2 \quad (9)$$

式中：$SIQR_{\text{down}}$和$SIQR_{\text{up}}$分别为箱体长度的上下界限，即改进的箱线图的上下半四分位距；Q_1、Q_2和Q_3分别为样本数据的上四分位数、中位数和下四分位数。

（2）计算非异常值截距线

非异常值截距线包括非异常最大值截距线和非异常最小值截距线，用来判定样本数据是否为异常值。用非异常最大值截距线作为是否是热异常值的判断依据，若样本数据超出此值，则判定其为热异常值，即可能存在热异常现象。样本数据偏离极端异常值截距线外侧越远，说明其热异常现象就越严重。其实现过程如公式（10）、公式（11）和公式（12）所示：

$$B_{\text{c}} = \frac{SIQR_{\text{up}} - SIQR_{\text{down}}}{SIQR_{\text{up}} + SIQR_{\text{dowm}}} \quad (10)$$

式中：B_{c}是鲍利系数，介于−1 到 1 之间；$SIQR_{\text{up}}$和$SIQR_{\text{dowm}}$分别为箱体的上下半四分位距。

$$f_{\text{up}} = Q_3 + 1.5 \times IQR \times (\frac{1 + B_{\text{c}}}{1 - B_{\text{c}}}) \quad (11)$$

$$f_{\text{down}} = Q_1 - 1.5 \times IQR \times (\frac{1 - B_{\text{c}}}{1 + B_{\text{c}}}) \quad (12)$$

式中：f_{up}和f_{down}分别为箱体的非异常最大值截距线和非异常最小值截距线，当样本数据集中某一数据大于f_{up}时，则归为热异常值；当样本数据集中某一数据小于f_{down}时，则归为冷异常值。

最后，利用 ArcGIS 软件逐像元筛选大于非异常最大值截距线的值。

1.3.3 基于热异常频率提取热异常区

热异常区的地表温度往往高于周边地表温度，形成局部热岛现象，且其出现的频率一般大于由其他因素引起的热异常出现的频率。因此，基于热异常提取结果，统计分析年度热异常区出现的频率。为避免漏判现象，当某一热异常区出现的频率大于 60%时，则判定为城市热异常区，具体实现如公式（13）所示：

$$f_{\text{local}} = \frac{i_{\text{local}}}{n_{\text{local}}} \quad (13)$$

式中：f_{local}代表热异常出现的频率；i_{local}代表局部某一热异常区一年出现的次数；n_{local}代表一年中整体热异常区出现的总次数。

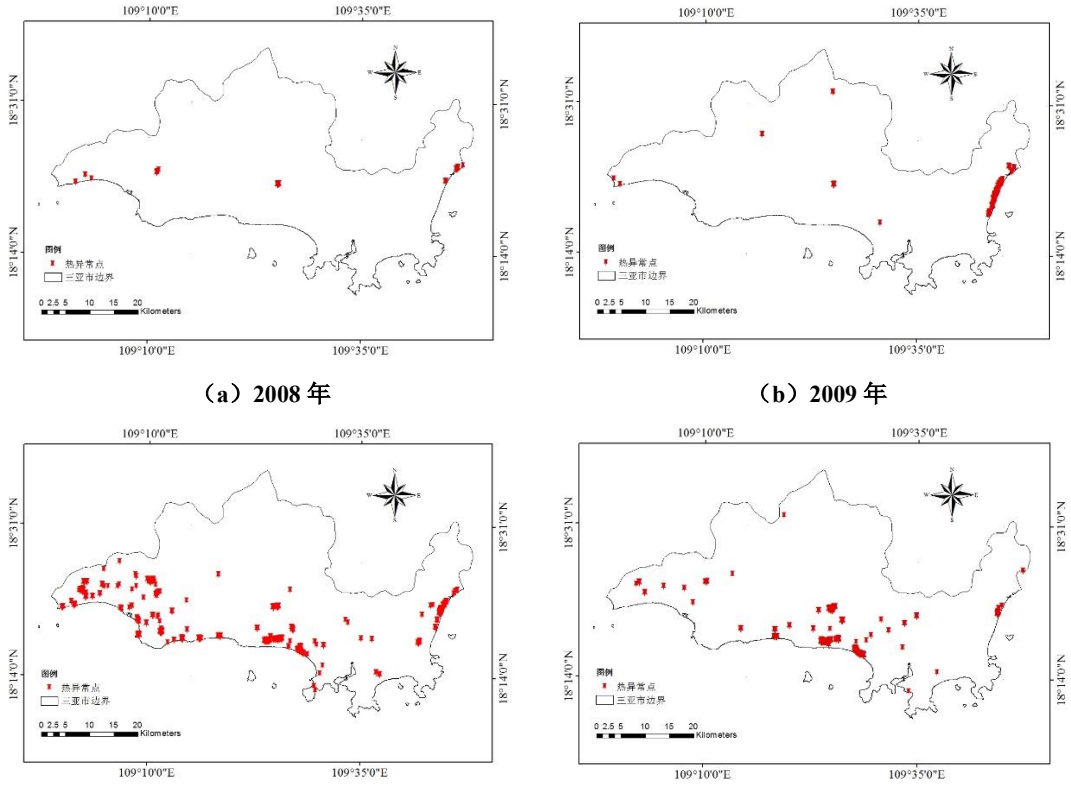
为进一步探究城市热异常区的分布范围，统计分析 2008–2017 年热异常出现的频率，将提取频率大于 60%的区域作为这 10 年的城市热异常区，具体实现如公式（14）所示：

$$f_{\text{global}} = \frac{i_{\text{global}}}{n_{\text{global}}} \quad (14)$$

式中：f_{global}代表 10 年中热异常出现的频率；i_{global}代表局部某一热异常区 10 年中出现的总次数；n_{global}代表 10 年中整体热异常区出现的总次数，即取值为 10。

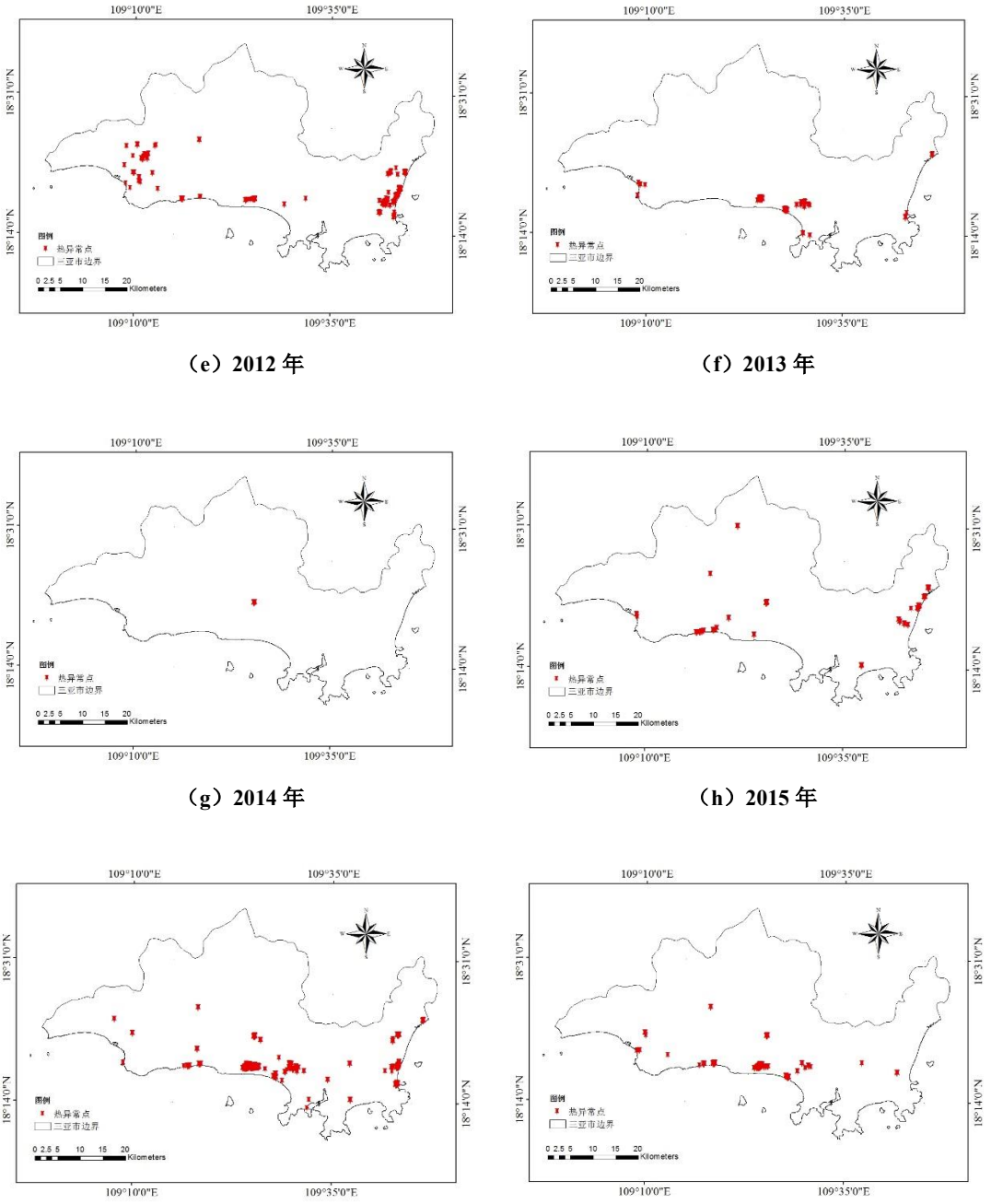
2 数据样本描述

使用上述步骤完成 2008–2017 年三亚热异常空间分布产品数据集，图 3 是三亚市 10 年热异常区空间分布产品示例图。



（a）2008 年　　（b）2009 年

（c）2010 年　　（d）2011 年



（e）2012 年　　（f）2013 年

（g）2014 年　　（h）2015 年

（i）2016 年　　（j）2017 年

图 3　2008－2017 年三亚市热异常区空间分布结果图

3　数据质量控制

本数据集通过以下几方面进行质量控制：

为验证方法有效性，利用 Google 地球可以查看往年影像的优势进行统计 2008–2017 年逐年工厂区总个数和改进的箱线图法识别的工厂区个数及工厂区个数，以此间接验证方法的可靠性。具体方法为：首先，在 Google 地球上逐年圈出工厂区并统计个数；其次，把改进的箱线图法识别的热异常转为 kmz 格式数据，加载到 Google 地球里并逐年统计其在工厂区的个数；最后，一个工厂区包含一个或多个工厂，将改进的箱线图法识别出的工厂统计到所属工厂区，统计工厂区个数，该工厂区个数与 Google 地球里圈出的工厂区个数之比为精度。由于有些工厂停产或排放的热量不足以达到热异

常的范围，或某些人为原因致使小领域温度达到热异常范围，以及人工判读工厂的误差，都会影响最终精度，其具体逐年精度如表 3。

表 3　2008–2017 年三亚市热异常区提取精度表

年份	识别个数	识别区数	工厂区个数	精度/（%）
2008 年	3	5	6	83.33
2009 年	5	5	6	83.33
2010 年	14	6	6	100
2011 年	8	6	6	100
2012 年	5	5	6	83.33
2013 年	7	5	6	83.33
2014 年	1	1	6	16.67
2015 年	7	6	6	100
2016 年	13	6	6	100
2017 年	8	6	6	100

4　数据价值

本数据集以三亚市为研究区，基于长时间序列的 Landsat 数据，结合改进的箱线图法进行热异常提取的成果可为城市规划与建设和生态质量评价等提供数据产品服务，为三亚市的居住环境舒适度提供参考标准。

此外，通过利用遥感技术对城市热异常进行监测和评价，可有助于减轻城市发展对环境的压力，促进环境保护和治理，为三亚国际岛建设提供技术参考。该数据集的长时间序列城市热异常产品和技术也可在其他城市中进行推广和应用。

数据作者分工职责

孟庆岩（1971—），男，黑龙江肇东人，博士，研究员，博士生导师，主要从事城市陆表环境遥感与地震红外遥感研究。主要承担数据集设计及技术指导工作。

谷艳春（1993—），女，河南周口人，硕士生，主要研究方向为城市热环境。主要承担数据处理和分析工作。

郝丽春（1993—），女，山西阳泉人，硕士生，主要研究方向为城市热环境。主要承担数据处理和分析工作。

胡蝶（1994—），女，天津人，博士生，主要研究方向为城市热环境。主要承担数据下载与分析工作。

张颖（1994—），男，湖北武汉人，博士生，主要研究方向为城市热环境。主要承担数据集设计及技术指导工作。

张琳琳（1994—），女，河北衡水人，博士生，主要研究方向为城市热环境。主要承担数据集设计及技术指导工作。

致　谢

感谢城市陆表环境团队提供技术支撑服务，感谢 USGS 和中国空气质量在线监测分析平台提供数据支撑服务。

参考文献

[1] 杨新兴, 李世莲, 尉鹏, 等. 环境中的热污染及其危害[J]. 前沿科学, 2014, 8(3): 14-26.

[2] 曲辰. 酷暑的反思——解读城市热污染[J]. 环境导报, 2003(19): 18-18.

[3] 郭桂香, 陈东威. 认识热污染问题[J]. 环境保护, 1994(5): 35-37.

[4] WU C , WANG Q , YANG Z , et al. Monitoring heated water pollution of the DaYaWan nuclear power plant using TM images[J]. International Journal of Remote Sensing, 2007, 28(5): 885-890.

[5] Chen C Q, Shi P, Mao Q W. Application of Remote Sensing Techniques for Monitoring the Thermal Pollution of Cooling-Water Discharge from Nuclear Power Plant[J]. Environmental Letters, 2003, 38(8): 10.

[6] LIU Y, HU C, ZHAN W, et al. Identifying industrial heat sources using time-series of the VIIRS Nightfire product with an object-oriented approach[J]. Remote Sensing of Environment, 2017: S0034425717304820.

[7] XIA H, CHEN Y, QUAN J. A simple meathod based on the thermal anomaly index to detect industrial heat sources[J]. International Journal of Applied Earth Observation & Geoinformation 2018 (73): 627-637.

[8] ZHANG L , MENG Q , SUN Z , et al. Spatial and temporal analysis of the mitigating effects of industrial relocation on the surface urban heat island over China[J]. ISPRS International Journal of Geo-Information, 2017, 6(121): 121.

[9] MURATA A , SASAKI H , HANAFUSA M , et al. Estimation of urban heat island intensity using biases in surface air temperature simulated by a nonhydrostatic regional climate model[J]. Theoretical & Applied Climatology, 2013, 112(1-2): 351-361.

[10] YANG X , LEUNG L R , ZHAO N , et al. Contribution of urbanization to the increase of extreme heat events in an urban agglomeration in east China[J]. Geophysical Research Letters, 2017, 44. DOI: 10.1002/2017GL074084.

[11] 赵英时. 遥感应用分析原理与方法[M]. 北京: 科学出版社, 2003.

[12] 覃志豪, 李文娟, 徐斌, 等. 陆地卫星 TM6 波段范围内地表比辐射率的估计[J]. 国土资源遥感, 2004, 16(3): 28-32.

[13] 丁凤, 徐涵秋. TM 热波段图像的地表温度反演算法与实验分析[J]. 地球信息科学学报, 2006, 8(3): 125-130.

[14] 邓书斌. ENVI 遥感图像处理方法. 第 2 版[M]. 北京: 高等教育出版社, 2014.

[15] SOBRINO J A , JIMENEZ-MUOZ J C , SORIA G , et al. Land surface emissivity retrieval from different VNIR and TIR sensors[J]. IEEE Transactions on Geoscience and Remote Sensing, 2008, 46(2): 316-327.

[16] CARLSON T N , RIPLEY D A . On the relation between NDVI, fractional vegetation cover, and leaf area index[J]. Remote Sensing of Environment, 1997, 62(3): 241-252.

[17] JIMENEZMUNOZ J C , CRISTOBAL J , SOBRINO J A, et al. Revision of the single-channel algorithm for land surface temperature retrieval from Landsat thermal-infrared data[J]. IEEE Transactions on Geoscience & Remote Sensing, 2009 47(1): 339-349.

[18] CHANDER G, MARKHAM B. Revised landsat-5 tm radiometric calibration procedures and postcalibration dynamic ranges[J]. IEEE Transactions on Geoscience and Remote Sensing, 2003, 41(11): 2674-2677.

[19] WALKER M L , DOVOEDO Y H , CHAKRABORTI S , et al. An improved boxplot for univariate data[J]. American Statistician, 2018: 1-13.

[20] HUBERT M, VANDERVIEREN E. An adjusted boxplot for skewed distributions[J]. Computational Statistics & Data Analysis, 2004, 52(12): 5186-5201.

论文引用格式

孟庆岩，谷艳春，郝丽春，等. 三亚市热异常遥感监测数据集[J/OL]. 中国科学数据, 2019, 4(2). (2019-04-23). DOI: 10.11922/csdata.2018.0077.zh.

数据引用格式

孟庆岩，谷艳春，郝丽春，等. 三亚市热异常遥感监测数据集[DB/OL]. Science Data Bank, 2018. (2019-03-23). DOI: 10.11922/sciencedb.700.

Remote sensing data for urban thermal anomaly in Sanya City

Meng Qingyan[1*], Gu Yanchun[1,2], Hao Lichun[1,3,4], Hu Die[1,3,5], Zhang Ying[1,3,5], Zhang Linlin[1,2,5]

1. Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100101, P. R. China;
2. Spatial Information Engineering Research Centre of Fujian Province, Fuzhou 350000, P. R. China
3. Hainan Provincial Key Laboratory of Earth Observation, Sanya 572029, P. R. China;
4. School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo 454000, P. R. China;
5. College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, P. R. China;

* Email: mengqy@radi.ac.cn

Abstract: Aiming at the problem of thermal anomaly in Sanya, this paper takes Landsat data from 2008 to 2017 as the main data source, extracts the high temperature anomaly area based on the radiation equation transfer method and the improved box-plot method, carries out frequency statistics of the high temperature anomaly area, determines the high temperature anomaly area whose frequency is more than 60% as the urban thermal anomaly area, and obtains 10-year urban thermal pollution anomaly area. This data set can be applied

to other cities to extract multi-regional urban thermal anomaly areas, so as to play a higher application value.
Keywords: thermal anomaly; radiation transfer equation; modified boxplot method; frequency of the high-temperature anomaly area

Dataset Profile

Title	Remote sensing data for urban thermal anomaly in Sanya City
Data corresponding author	Meng Qingyan (mengqy@radi.ac.cn)
Data authors	Meng Qingyan, Gu Yanchun, Hao Lichun, Hu Die, Zhang Ying, Zhang Linlin
Time range	2008–2017
Geographical scope	Sanya (18°09′34″–18°37′27″N, 108°56′30″–109°48′28″E)
Spatial resolution	Landsat 5：30/120 m Landsat 7：30/60 m Landsat 8：30/100 m
Data volume	4.05 MB
Data format	*.vsdx,*.jpg,*.png, *.py, *.shp
Data service system	<http://www.sciencedb.cn/dataSet/handle/700>
Sources of funding	Natural Science Foundation of Hainan Province (417219); Major Science and Technology Program of Hainan Province (ZDKJ2016021); Science and Technology Program of Sichuan Province (2018JZ0054); Major Science and Technology Program of Hainan Province (ZDKJ2017009).
Dataset composition	This data set mainly includes seven folders comprised of the following: 1. Cover Map Folder: It contains data named Cover Map*.jpg format; 2. Sanya City Boundary Map Folder: Contains the boundary map data of Sanya City with the format *.jpg; 3. Data Processing Flow Chart Folder: It contains data named Data Processing Flow Chart*. vsdx format; 4. The folder of annual thermal anomaly extraction results of Sanya City from 2008 to 2017: annual thermal anomaly extraction results of Sanya City in 2008-2017*.png format; 5. Metadata folder of annual thermal anomaly extraction results in Sanya City from 2008 to 2017: metadata of annual thermal anomaly extraction results in Sanya City in 2008-2017*.shp format; 6. Code folder: contains the *.py format code used in the data processing process; 7. Sanya Vector Boundary Folder: Sanya City Boundary Data with *. SHP format.