

# Research on Earthquake Disaster Background Data in Spatial Information Diffusion Model

Rundong Wang<sup>1,2</sup> Chongfu Huang<sup>1,2,\*</sup>

<sup>1</sup> Key Laboratory of Environmental Change and Natural Disaster, Ministry of Education, Beijing Normal University, Beijing 100875, China;

<sup>2</sup> Academy of Disaster Reduction and Emergency Management, Ministry of Emergency Management & Ministry of Education, the Peoples' Republic of China, Beijing 100875, China

\*Corresponding author. Email: hcongfu@bnu.edu.cn

## ABSTRACT

When a major natural disaster occurs, the data in the disaster area is usually incomplete and information islands may appear. Considering the time required by the on-site survey statistical methods, geospatial information diffusion models can be used to infer the rescue needs of information islands. The basic principle is to spread the disaster information around the information island to the information island. Data that is related to rescue needs and can be collected before a disaster occurs is called background data. The background data in each geographic unit in the disaster area played a role as a bridge for information diffusion. The rapid assessment of earthquake disasters is an important basis for earthquake emergency response and government decision-making, and reliable rapid assessment of earthquake disasters is of great significance to improving earthquake emergency response capabilities. The rapid earthquake disaster assessment is an emergency assessment based on the preparation of the preliminary data and the difficulty of obtaining the post-disaster data, starting from the time of the earthquake. Under the requirements of rapid disaster assessment, what needs to be sought is a simple and efficient relationship that meets certain accuracy requirements, so that the rapid assessment of the disaster situation can be carried out based on easily accessible or available indicators in advance. This paper uses the literature analysis method and uses CiteSpace to study the high-frequency co-occurrence keywords in the field of earthquake disaster loss assessment from 1996 to 2020 and the evolution of research hotspots in the field from 2010 to 2020, and obtain parameters that can be used as background data.

**Keywords:** Background data, Geospatial Information Diffusion, CiteSpace, Rapid Disaster Assessment

## 空间信息扩散模型中地震灾情背景数据的研究

王润东<sup>1,2</sup> 黄崇福<sup>1,2,\*</sup>

<sup>1</sup> 北京师范大学, 环境演变与自然灾害教育部重点实验室, 北京 100875, 中国

<sup>2</sup> 应急管理部/教育部减灾与应急管理研究院, 北京 100875, 中国

\*通讯作者. 电子邮箱: hcongfu@bnu.edu.cn

## 摘要

当重大自然灾害发生后, 发生灾害区域的数据通常是不完整的且可能会出现信息孤岛, 考虑到现场调查统计方法所需要的时间, 可以采用地理空间信息扩散模型来推测信息孤岛的救助需求。基本原理是将信息孤岛周边的灾情信息, 扩散到信息孤岛中去。与救助需求有关且能在灾害发生前就能收集备用的数据, 称为背景数据。而灾区各地理单元中的背景数据, 为信息扩散起到桥梁作用。地震灾害快速评估是地震应急响应与政府

决策的重要依据,可靠的地震灾害快速评估对提高地震应急能力有重要意义,地震灾害快速评估是根据前期数据的准备以及灾后数据获取的难易程度,以地震发生时刻为起点进行的应急评估。在灾情快速评估的要求下,需要寻求的是满足一定精度要求下的简单、高效的关系,从而基于易获取的或提前可获得的指标进行灾情的快速评估。本文通过文献分析法,使用 CiteSpace,研究地震灾害损失评估领域 1996 - 2020 年的高频共现关键词以及 2010 - 2020 该领域研究热点演变,获得可以作为背景数据的参数。

**关键字:** 背景数据,地理空间信息扩散, CiteSpace, 灾情快速评估

## 1. 引言

“空间信息扩散模型中地震灾情背景数据的研究”属于国家重点研发计划中研究任务“基于地理空间信息扩散的救助需求与能力在线评估”中的一项工作。传统的快速灾情评估方法是一种通常是粗放型的“隔空判灾”式方法,是以县级行政区域为单位进行评估,且在时效性和精度上均和地震应急的实际需求存在着一定的差距。而本研究考虑用精细的“采点判灾”代替“隔空判灾”式方法,为重大自然灾害发生后,由零星的灾区采点信息,线上推演灾区的全部灾情,提供技术支持。

在过去,地震工作者苦于数据的匮乏而无法迅速开展地震应急和灾评工作,例如在 2008 年汶川特大地震中,强地震动使得生命线系统丧失功能,通讯系统破坏惨重,灾区多地形成了“信息孤岛”,灾情信息无法及时传递,给应急指挥和部署救援工作带来了相当大的困难。而地理空间上信息扩散技术,依托物联网,依靠灾害发生前收集整理的与灾情有关的背景数据,将信息孤岛地理单元上的灾情推测出来,有望增强实效性、提高精度,并且将快速评估由县级精度提高到乡镇精度。而只有借助灾区各个地理单元中与灾情有关的背景数据,才能够实现已知灾情的信息扩散,所以,背景数据是地理空间信息扩散的基础,物联网中存有什么样的背景数据,能否用地理空间信息扩散技术从背景数据和灾区采点数据中学习可靠的因果关系,将决定着推演灾区的全部灾情是否符合实际。

近年来,伴随着移动互联网技术的快速发展和普及,互联网时代已经到来。利益相关者比以往更加关注灾害事件,人们更希望参与到信息的发布和传播中去,而不仅仅只做信息的接受者。以风险沟通的方式来确定地震灾情背景数据中相对暴露度的指标类型及标准,让利益相关者参与到防灾减灾工作中来,为救灾物联网的地理空间信息扩散技术提供灾情背景数据支持,从而提高地震灾情快速评估的实效性与精度,辅助地震灾情的快速研判,为政府

部门实施赈灾救灾方案、做出左右决策提供有力依据。

## 2. 地理空间信息扩散技术研究现状

在研究中,我们经常遇到观测样本较少的小样本问题。为了有效提高小样本的利用率,黄崇福教授在 1997 年给出了信息扩散的严格定义,其核心原理可以简述为:令  $X$  是一个给定的样本,假定用它估计一个在论域  $U$  上的关系。当且仅当  $X$  不完备时,必定存在一个适当的扩散函数和相应的算子,使得扩散估计比非扩散估计更靠近真实关系。此处,论域  $U$  是指集合  $X$  中元素(样本点)的定义域。

概率空间中的信息扩散方法采用模糊关系来表达,其基础比较稳固,应用涉及面较广。在灾情快速评估时,用周围地理单元的灾情数据推测空白单元的未知灾情数据同样属于小样本问题,但它属于地理空间上信息扩散,传统的概率空间上的信息扩散将不再适用;由于乡镇级别的地理单元不满足连续性假设,常用的内插模型也将不再适用。只有将概率空间中的信息扩散技术,发展为地理空间上的信息扩散技术,才能将信息孤岛地理单元上的灾情,由周边地理单元上的灾情推测出来。黄崇福教授给出了地理空间信息扩散的形式化定义:设  $W$  是用于识别区域  $G$  上的现象  $F$  的不完整数据集, $Z_G$  是背景数据。如果模型  $\gamma$  能用  $Z_G$  使  $W$  变成完整数据,则称  $\gamma$  用  $Z_G$  对  $W$  的信息在  $G$  上进行了扩散。

图 1 展示了地理空间信息扩散技术的计算过程。包括两个部分。先通过有完整数据集的地理单元上的背景数据和已观测数据,构成监控空间,然后进行正态扩散公式,将每个控制点上的信息进行累加,对所得矩阵进行归一化,得到关系矩阵。第二部分是用空白单元的背景数据,正态扩散后,进行输入模糊化,结合第一部分得到的关系矩阵,使用近似推理公式得到模糊输出,而后运用重心法,最后得到空白单元的观测数据。

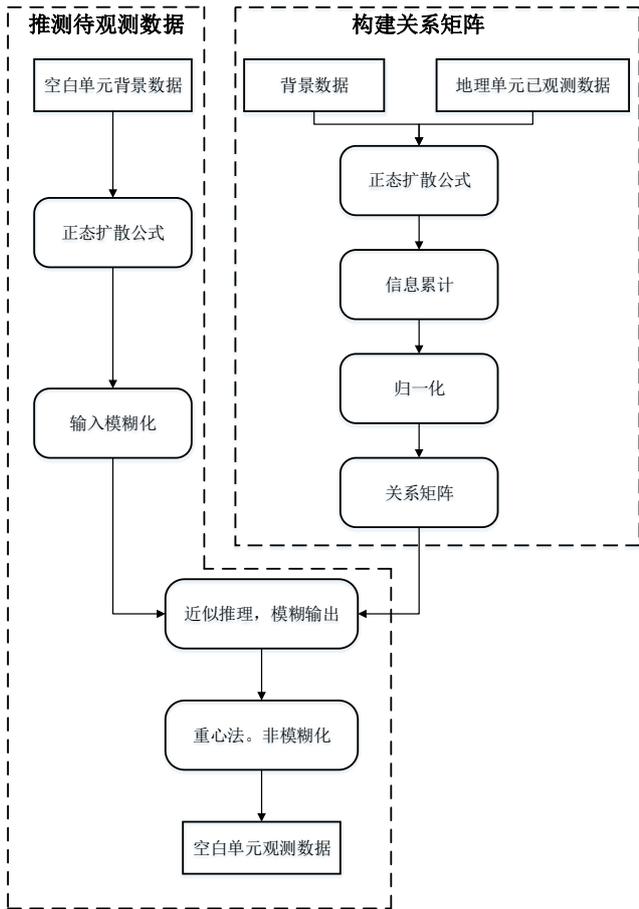


图 1 地理空间信息扩散技术的计算过程

灾情背景数据是地理空间信息扩散的基础。背景数据的定义：设  $g_1, g_2, \dots, g_n$  是  $n$  个被观测单元， $o$  是一个空白单元，记作  $G = \{g_1, g_2, \dots, g_n, o\}$ 。设  $z$  是多个特征的属性值向量， $Z_G = \{zg_1, zg_2, \dots, zg_n, zo\}$  是地理特征属性向量的集合。如果  $Z_G$  中的数据可以作为媒介来对空白单元  $o$  进行赋值，称  $Z_G$  为背景数据集，简称背景数据。一个属性值的数据是否为另一个属性值的背景数据，由两个因素决定，一是能否发挥制约作用；二是能否容易获得。

### 3. 资料与方法

#### 3.1. 数据来源

国内外对地震损失评估的相关研究成果众多，文章数量庞大，本研究数据来源为 Web of Science

核心数据库。在数据的采集上，主题检索方式为高级检索：设定检索条件为“TS（主题） = ‘earthquake（地震）’、‘loss（损失）’ AND ‘assessment（评估）’”；检索时间段设定为“1996 - 2020”；数据采集按时间为 2020 年 9 月 20 日；检索文献类型包括了 Article 和 Review；最后共获得有效文献 2217 篇，每篇文献的记录内容包括题目、作者、摘要、机构、关键词、发表年份、发表期刊及参考文献等。

#### 3.2. 研究方法

自从文献计量学开创以来，基于已发表的文献进行定量分析的方法开始出现，并逐步走向成熟。由美国 Drexel 大学信息科学与技术学院陈超美教授研发的信息可视化软件 Citespace，在基于文献计量学进行学科领域研究热点演变分析中得到广泛应用。该软件是基于 JAVA 程序语言编写的一款着眼于分析科学研究中蕴含的潜在知识，是在科学计量学、数据可视化背景下逐渐发展起来的引文可视化分析软件，通过可视化手段呈现科学知识的结构、规律和分布情况，以此得到“科学知识图谱”的可视化图形。目前，该软件已经广泛应用于国内外各大学科领域，在中国，使用该软件的研究多集中在管理学、教育学、社会学、经济学等领域。

本研究主要采用 CiteSpace 软件平台，运用文献计量学、信息可视化分析与统计分析方法，对获取的文献数据进行定量研究与定性分析相结合，以实现地震灾害损失评估研究文献的特征和热点演变可视化的目标。

### 4. 结果与分析

#### 4.1. 年文献计量分布

为了更好地了解地震灾害损失评估研究内容的特征和演变趋势，首先对从 Web of Science 数据库中筛选的 2217 篇文章进行年文献数量变化的分析，结果如图 2 所示。由图可见，自 2008 年以来，文献数量整体表现为明显的上升趋势。

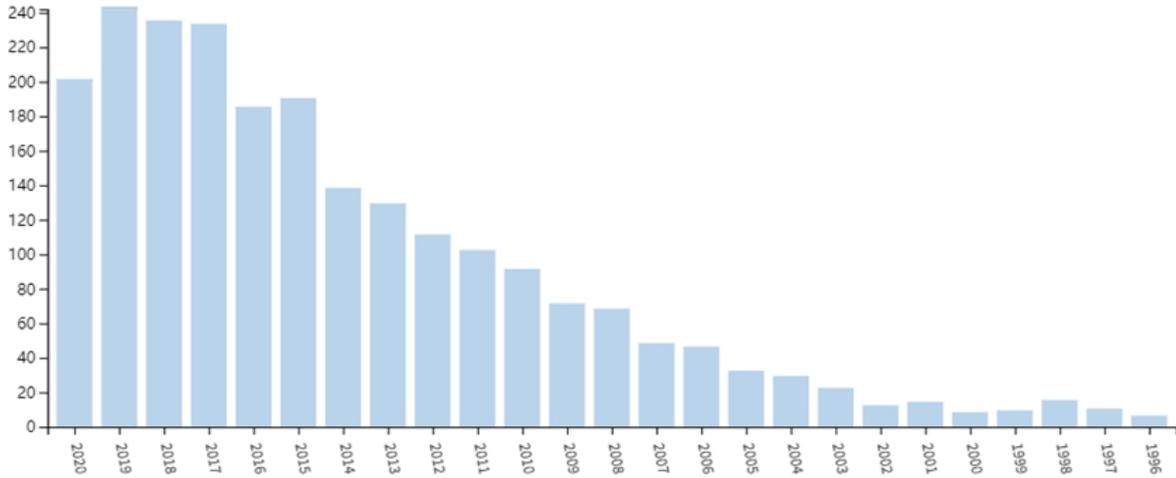


图 2 1996-2020 年地震损失评估领域年文献数量

#### 4.2. 高频关键词

词频是所分析的文档中词语出现的次数。词频分析方法就是在文献信息中提取能够表达文献核心内容的关键词或主题词频次的高低分布，来研究该领域发展动向和研究热点的方法。关键词是能反应文章主题或思想的关键性词语，它是对文章主题的高度概括，同时是贯穿文章的主要线索。关键词的共现分析是对所提供的关键词的分析。在 CiteSpace 对关键词进行分析时，分析步骤与界面如图 3 所示。

设置时间切片为 1 年，Node Type 选择 Keyword，连线强度选择 Cosine，网络裁剪使用 MST+Pruning the merged network + Pruning the sliced networks。得到地震损失评估高频共现关键词知识图谱如图所示。从图 4 可以看出，研究文献中出现频次最大的关键词是“earthquake（地震）”，同时也是文献获取中数据来源的基础条件。除“地震”外，“建筑结构易损性”、“脆弱性分析”、“地震工

程”、“倒塌”、“断裂带”、“灾害管理”、“工程保险”等出现频次也较高。可见，在地震灾害损失评估的研究中关于建筑物结构易损性的研究是热点。



图 3 Citespace 共词分析主要步骤

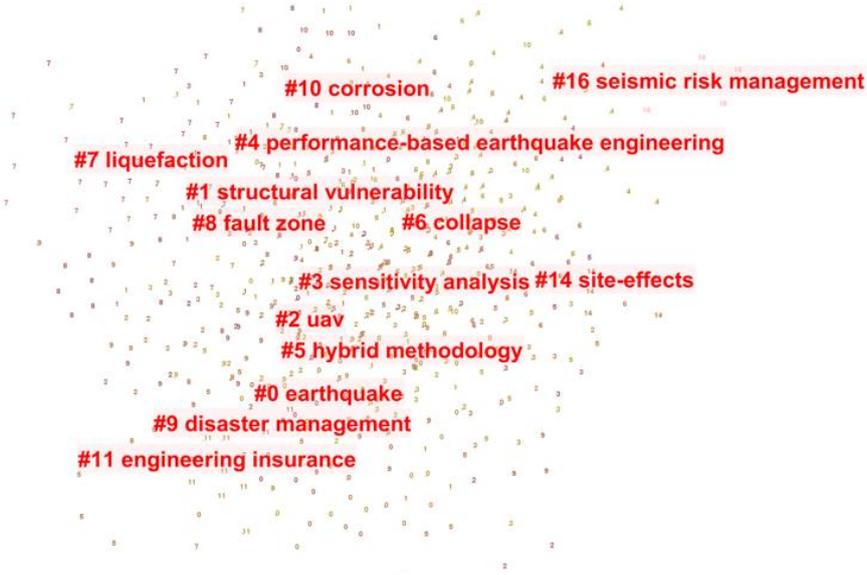


图 4 地震灾害损失评估高频共现关键词知识图谱

### 4.3. 基于文献共被引的研究热点进展分析

在文献计量学中，有学者认为引文分析法就是利用数学及统计学方法进行比较、归纳、抽象、概括等，对科学期刊、论文、著者等分析对象的引用和被引现象进行分析，以揭示其数量特征和内在规律的一种信息计量研究方法。

选择 2010 - 2020 的文献数据，设置 Node Types 为 Cited Reference，连线强度选择 Cosine，网络裁剪使用 Pruning the sliced networks，可视化方式选择 Cluster View - Static 和 Show Merged Network。当

网络计算完成后，点击“Visualize”对分析的数据网络进行可视化。

如图 5 所示，限定参与高频共被引分析的文献发表时间在 2010 年以后，2010 - 2020 年地震灾害损失评估研究文献显示 11 个高频聚类词，包括建筑破坏 (#0)、地震风险 (#1)、地震风险评估 (#2)、破坏 (#3)、衰败 (#4)、地震 (#5)、成本效益分析 (#6)、地震灾害评估 (#7)、地震易损性 (#8)、工程保险 (#9)、地震风险管理 (#10)。可以看出，高频聚类词涉及到建筑结构的易损性和地震的危险性。

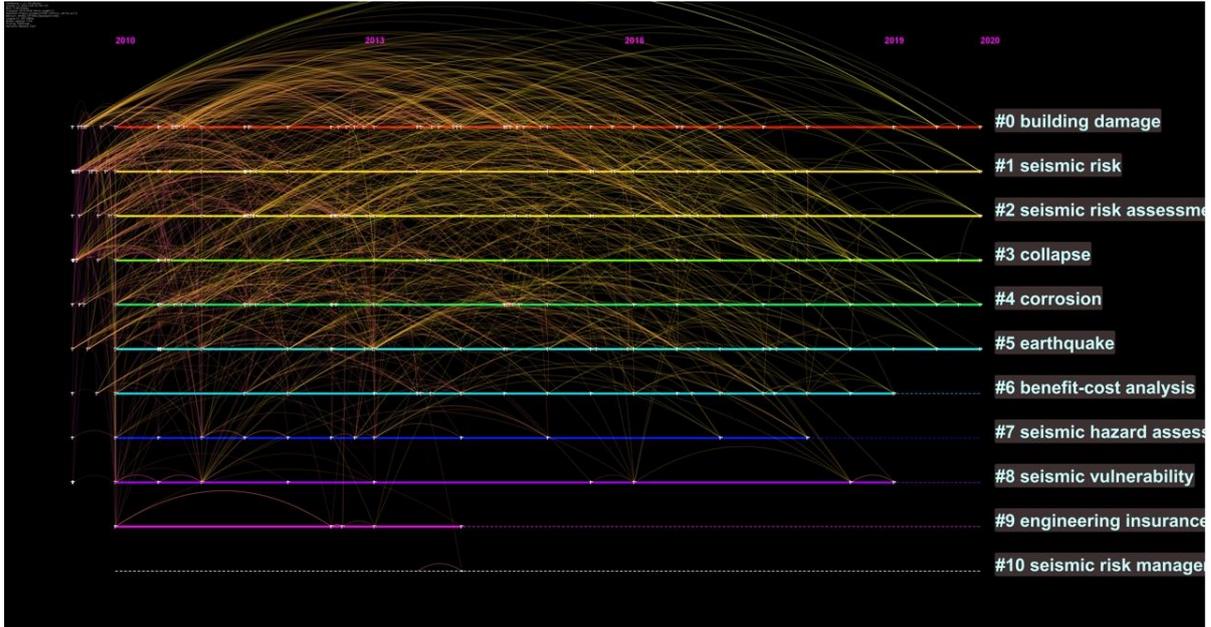


图 5 地震灾害损失评估高频共被引文献研究热点演变 (2010-2020 年)

### 4.4. 地理空间信息扩散技术参数选择

根据 1996 - 2020 地震灾害损失评估高频共现关键词与 2010 - 2020 地震灾害损失评估高频共被引文献研究热点演变，应基于地震危险性分析与结构易损性分析，结合损失评估理论，对不同强度水平的地震动作用下的建筑物或人员遭受的损失进行评估。评估流程如图 6 所示。

而在背景数据上，可以选择人口、人口密度、GDP、人均 GDP、地面峰值加速度-结构超越概率图、破坏状态-破坏概率图以及灾后可迅速获得的震级、震中距等，当以经济损失为损失评估指标时，可以引入建筑物的重置成本，当以人员伤亡为损失评估指标时，可以引入人员在室率作为背景数据。

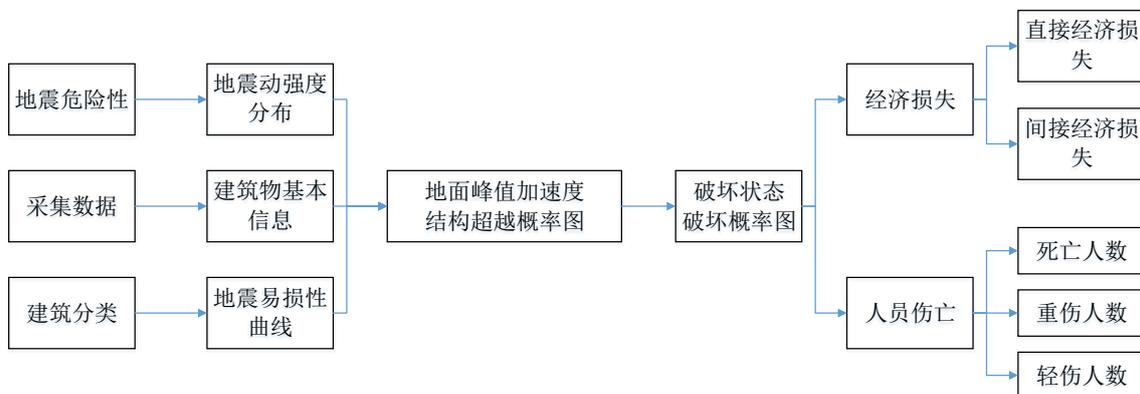


图 6 地震灾害损失评估流程

## 5. 结论与讨论

### 5.1. 结论

(1) 1996 年至今, 地震损失评估领域的相关文献数量不断增多, 自 2008 年以来, 呈明显的上升趋势。

(2) “建筑结构易损性”、“脆弱性分析”、“地震”、“地震工程”、“倒塌”、“断裂带”、“灾害管理”、“工程保险”等是地震损失评估领域的研究热点词汇。2010 - 2020 年该领域研究热点也越来越关注建筑破坏。

(3) 根据地震灾害损失评估领域高频共现关键词与 2010 - 2020 研究热点演变, 提出在地理空间信息扩散技术背景数据上, 选择人口、人口密度、GDP、人均 GDP、地面峰值加速度-结构超越概率图、破坏状态-破坏概率图、震级和震中距等, 并且根据损失评估指标不同时, 再加入不同的背景数据, 以经济损失为损失评估指标时, 引入建筑物的重置成本, 当以人员伤亡为损失评估指标时, 引入人员在室率作为背景数据。

### 5.2. 讨论

本研究基于 CiteSpace 软件进行了地震灾害损失评估的文献计量分析, 通过网络可视化分析的结果了解 1996 - 2020 年该领域的高频共现关键词、高频共被引文献研究热点演变等。相关结论将为地理空间信息扩散技术的参数选择提供思路, 也为该领域进一步明确研究方向、创新研究方法、提高未来研究结果精确度等方面提供借鉴意义。分析中所使用的数据来自 Web of Science 数据库, 重点考虑了世界范围内该领域的基本研究情况。CNKI、CSSCI 等基础数据库更有利于全面反应中国在该领域的研究变化情况, 如 2008 年后地震损失评估在中国的发展演变。可进一步结合 CNKI 等中国基础文献数据库, 采用 CiteSpace 软件进行分析。

## 致谢

本项目受国家重点研发计划课题“重大自然灾害多层次精准救助关键技术研究”(编号: 2017YFC1502902)资助。

## 参考文献

[1] 黄崇福, 地理空间上的信息扩散及其在风险分析中的应用, 一带一路背景下的风险分析和危机反应, 黄崇福王成军编(中国灾害防御协会风险分析专业委员会第八届年会论文集, 西安, 2018 年 10 月 20-21 日), Atlantis Press, Paris, 2018, pp.1-7. DOI: <https://doi.org/10.2991/rac-18.2018.1>

- [2] 黄崇福, 自然灾害风险分析与管理, 科学出版社, 北京, 2012 年.
- [3] 黄崇福, 借助背景数据的地理空间信息扩散技术, *Journal of Risk Analysis and Crisis Response*, 2019,9(1):2-10.DOI: [https://doi.org/10.1007/978-3-642-21455-4\\_3](https://doi.org/10.1007/978-3-642-21455-4_3)
- [4] M. Erdik, K. Şeşetyan, M.B. Demircioğlu, et al. Rapid earthquake loss assessment after damaging earthquakes. 2010, 31(2):247-266.
- [5] 杨天青,姜立新,杨桂岭.地震人员伤亡快速评估[J].地震地磁观测与研究,2006(04):39-43.
- [6] Bai Tao Sun, Hong Fu Chen. An Earthquake Loss Assessment Method to Buildings in Urban Zones and Its Application in Wenchuan Ms8.0 Great Earthquake. 2010, 850:817-820.
- [7] 刘如山, 余世舟, 颜冬启等, 2014. 地震破坏与经济损失快速评估精细化方法研究. 应用基础与工程科学学报, 22(5):928—940.
- [8] 黄崇福.自然灾害风险相对暴露度的研究, 第六届中国西部风险分析与风险管理学术研讨会论文集, 黄崇福、骆华松、杨永生编,巴黎,Atlantis 出版社,1-5
- [9] 陈悦,陈超美,刘则渊,胡志刚,王贤文. CiteSpace 知识图谱的方法论功能[J]. 科学学研究, 2015, 33(02):242-253.
- [10] 张露,张俊飏,童庆蒙,郭晴.农业碳排放研究进展: 基于 CiteSpace 的文献计量分析[J].科技管理研究,2015,35(21):219-223.
- [11] 袁一凡.四川汶川 8.0 级地震损失评估[J].地震工程与工程振动,2008(05):10-19.
- [12] Hsi-Hsien Wei, Igal M. Shohet, Mirosław J. Skibniewski, et al. Assessment of Casualty and Economic Losses from Earthquakes Using Semi-empirical Model. 2015, 123:599-605.