外交评论　2020 年 第 2 期
DOI:10.13569/j.cnki.far.2020.02.121

非洲恐怖袭击时空规律的大数据分析
——基于 GIS 技术和分离总体持续期模型

陈　冲　庞　珣

摘　要　大数据及其分析技术的发展极大地推动了国际关系研究的发展，使得传统国际关系研究无法完成的任务变为可能。恐怖主义是当今全球治理中重大而棘手的问题，也是国际关系领域长期关注的议题。本文运用海量事件数据和地理信息系统（GIS）技术，将恐怖袭击的地理单元细粒化到以 0.5×0.5 经纬度作为分辨率的时空网格，寻找微观分析单元上恐怖袭击的时空规律，并运用分离总体持续期模型对恐怖袭击风险的空间分布和发生时机进行解释和预测。通过对非洲 10674 个时空网格在 1994—2013 年间恐怖袭击的时空特征进行分析，本文对非洲恐怖袭击发生时机进行精准捕捉，对恐袭地点进行细粒化定位，并对恐袭地理分布进行“疫区”与“免疫区”区分，不但深化了对恐怖主义的认识，而且提高了恐袭预测的精度和实际应用价值。本文的研究尝试表明，在大数据时代进行国际关系研究，不仅要充分利用海量数据提供的前所未有的丰富信息，更要选择和运用恰当的统计技术对复杂时空相关性进行细粒度分析，从而形成有效的创新路径。

关键词　大数据　恐怖主义　非洲　地理信息系统　时空网格　国际关系研究　事件数据

---

*　陈冲，清华大学国际关系学系助理教授；庞珣，清华大学国际关系学系教授。（北京 100084）

**　本文是国家社科基金重大项目“地缘政治风险预测的理论与方法研究”（项目编号：17ZDA110）的阶段性成果。复制本文结果的代码和数据可以在 Harvard Dataverse（https://doi.org/10.7910/DVN/UN9JWF）下载。

一、导　　言

当今大数据国际关系研究方兴未艾。前所未有的数据容量、多变的数据形式以及近乎实时的数据产生速度,[①] 使得众多传统上近乎无法进行实证研究的任务成为可能。[②] 在冲突研究中,大数据技术不仅催生海量新数据和新测量方法,也直接推动了理论创新和方法革新,形成了实证化、微观化和细粒化的研究大趋势。研究者努力超越传统以国家为基本分析单元的路径,在寻求"次国家层次"(sub-national level)的解释中不断细粒化分析单元,追求精准定位地理空间。[③] 在此背景下,地理信息系统(Geographical Information System,简称GIS)技术成为与大数据国际关系分析伴生的重要分析手段,用以帮助研究者在高维空间[④]中重新审视战争与和平、稳定与动荡、发展与贫困等一系列重大问题。[⑤]

恐怖主义是国际关系研究长期关注的重大理论和政策问题,但这一领域的实证研究面临一些最具挑战性的困难。由于数据和方法上的限制,传统研究以国家作为单元来分析恐怖主义这种非国家暴力形式,[⑥] 势必造成分析上的重大缺陷,从而强化"恐怖主义袭击随机而不可预测"的悲观判断,[⑦] 更无助于反恐防恐的政策应用。然而,近年来大数据信息和方法的蓬勃发展为恐怖主义研究提供了新的思路。GIS技术超越国家层次的地理追踪,可以发现恐怖袭击危险在地理分布上的细微和动态特征,克服传统研究过于宏观的

① Burt L. Monroe, "The Five Vs of Big Data Political Science: Introduction to the Virtual Issue on Big Data in Political Science", *Political Analysis*, Vol. 21, No. 5, 2013, pp. 1-9.

② 相关讨论参见 Andrej Zwitter, "Big Data and International Relations", *Ethics & International Affairs*, Vol. 29, No. 4, 2015, pp. 377-389; 庞珣、刘子夜:《基于海量事件数据的中美关系分析——对等反应、政策惯性及第三方因素》,《世界经济与政治》,2019年第5期。

③ 漆海霞:《大数据与国际关系研究创新》,《中国社会科学》,2018年第6期。

④ 在国际关系中,空间不仅指地理意义上的空间,还包括经济、政治、战略、文化等非地理意义上的空间。

⑤ Lars-Erik Cederman and Manuel Vogt, "Dynamics and Logics of Civil War", *Journal of Conflict Resolution*, Vol. 61, No. 9, 2017, pp. 1992-2016; Jordan Branch, "Geographic Information Systems (GIS) in International Relations", *International Organization*, Vol. 70, No. 4, 2016, pp. 845-869.

⑥ Joseph K. Young and Michael G. Findley, "Promise and Pitfalls of Terrorism Research", *International Studies Review*, Vol. 13, No. 3, 2011, pp. 411-431.

⑦ Josiah Marineau, et al., "The Local Geography of Transnational Terrorism", *Conflict Management and Peace Science*, online first, 2018, pp. 1-32, https://doi.org/10.1177/0738894218789356.

缺陷。在更精细的层面上发现差异，对于探寻规律、定位风险、识别关键因素、制订有效反恐行动方案等均有重大价值。

本文是运用大数据对恐怖主义袭击进行分析预测的一次创新性尝试，通过应用 GIS 技术将分析层次细粒化到比国家更细微数十个层级的时空网格，[①] 对恐怖袭击的时空规律进行把握，从而尝试更为准确地寻找影响恐袭的因素，更为精确地进行预测。具体而言，本文聚焦非洲 1994—2013 年间的恐怖袭击风险，以 0.5×0.5 经纬度作为分辨率，运用 GIS 技术将非洲分为 10674 个时空网格，并把恐怖袭击的地理定位“叠加”（overlay）在时空网格上，以便精准观察恐怖袭击的时空规律。本文进而利用卫星遥感技术等大数据方法获取网格层面分析所需数据，通过机器学习和理论研究建立统计模型，对恐怖袭击在网格层面上的时空特征进行解释和预测，通过分离总体持续期模型（split-population duration model，简称 SPDM）聚焦地理单元面对恐怖袭击的“免疫力”和发生恐袭的“时机”（timing）。

选择非洲作为分析地域，一方面是由于恐怖主义近年来在非洲急速增长并广泛蔓延，非洲业已成为恐怖主义研究的热点地区。根据澳大利亚经济与和平研究所 2019 年 11 月发布的“2019 全球恐怖主义指数”，全世界受恐怖主义影响严重的国家有一半位于非洲大陆。布基纳法索、尼日尔、索马里、刚果（金）等非洲国家遭受恐怖袭击日益频繁和严重，造成大量人员伤亡和经济损失。[②] 另一方面，非洲的稳定与安全对于顺利推进我国“一带一路”倡议以及保护我国在非洲投资、贸易等合法权益至关重要，中非携手打造合作共赢的中非命运共同体的过程离不开非洲相对安全与稳定的政治环境。这要求我们对非洲的政治暴力尤其是恐怖主义活动进行系统性、规律性的认识和研究，从而准确、及时地预防和预警恐袭，规避恐袭风险。

本文发现，恐怖袭击追求“随机”的设计恰恰形成了恐袭的时空规律，识别这些规律对解释和预测恐怖主义风险的蔓延和恐袭的时机十分重要。从总体上看，随着恐怖主义“疫情”的加重和蔓延，原本具有较强恐袭免疫力的地区的感染风险会大幅度提高。同时，疫区可以被“治愈”而成为免疫区，但重复“感染”的风险却相对很高。自身或邻近地区曾经遭遇恐袭的地

① Andreas Forø Tollefsen，Håvard Strand and Halvard Buhaug，“PRIO-GRID：A Unified Spatial Data Structure”，*Journal of Peace Research*，Vol. 49，No. 2，2012，pp. 363-374.

② *Global Terrorism Index 2019*：*Measuring the Impact of Terrorism*，http://economicsandpeace.org/reports/；《非洲反恐形势频敲警钟》，《人民日报》，2019 年 12 月 30 日，第 16 版。

区具有更高和更紧迫的恐袭风险，需要特别加强风险防治。当邻国遭遇恐袭时，下一个恐袭的地理目标很有可能选在靠近邻国的边境地区，尤其是其中那些尚未遭受恐怖袭击的地点。经济发达的国家在整体上面临较低的恐袭风险，但其经济活动相对密集的地区对恐袭危险更为脆弱。这一发现得益于对卫星遥感获得的夜间灯光（nighttime lights emissions）数据对经济活动进行次国家层次的微观测量，由此精细区分恐怖主义袭击在宏观（国家）和微观地理（时空网格）层次的不同机制。这些被恐袭策划刻意隐藏的规律在以国家为单元和基于传统数据来源的分析中难以观察，但大数据信息和GIS技术的细粒度分析却可以对此进行较好的发现和识别。

本文对在地理网格层次恐怖袭击的时空规律的分析具有重要理论和现实意义。在理论上，作为政治暴力的一种，如何有效合理地结合海量数据与社会科学研究方法，准确检验和厘清恐怖主义爆发的相关机制，是解释政治暴力现象的重要研究课题。[①] 本文在地理网格层次上推进恐怖主义研究，所发现的恐怖主义活动时空规律，有助于加深对恐怖主义的动机和行为逻辑的理解。从大数据分析技术来看，本文运用GIS技术联接宏观与微观两个层次，根据预测的不同目标（时机和程度）建立预测模型，将社会科学理论与机器学习方法相结合对预测量进行高效选择，突出了大数据分析不能对单一技术进行机械运用，而应该根据研究任务在理论和现实基础上对多种技术进行甄选和综合的特点。从现实意义来看，恐怖主义对各国社会安全、政治进程和社会生态均产生了重要而持久的影响，[②] 反恐防恐任务紧迫而艰巨，且成本极高。进行细粒度的地理单元分析，掌握风险的动态变化和高精度的分布情况，有助于缩小反恐防恐目标范围、提高资源利用效率，把恐怖主义对国家社会生活的影响降到最低。同时，随着“一带一路”倡议的实施，准确评估沿线国家的政治风险是顺利推进“一带一路”建设、有效保护我国海外利益

① Michael G. Findley and Joseph K. Young, “Terrorism and Civil War: A Spatial and Temporal Approach to a Conceptual Problem”, *Perspectives on Politics*, Vol. 10, No. 2, 2012, pp. 285-305; Virginia Page Fortna, “Do Terrorists Win? Rebels' Use of Terrorism and Civil War Outcomes”, *International Organization*, Vol. 69, No. 3, 2015, pp. 519-556.

② Laia Balcells and Gerard Torrats-Espinosa, “Using a Natural Experiment to Estimate the Electoral Consequences of Terrorist Attacks”, *Proceedings of the National Academy of Sciences*, Vol. 115, No. 42, 2018, pp. 10624-10629; Tamar Mitts, “Terrorism and the Rise of Right-Wing Content in Israeli Books”, *International Organization*, Vol. 73, No. 1, 2019, pp. 203-224.

的必要智力支撑。[①]

本文提供的研究方法和路径适用于恐怖主义和其他风险种类的评估与预测，为大数据时代建立政治安全风险评估和预警体系，为国际关系研究在宏观与微观因素上的有效结合，提供了分析方法上的尝试和准备。需要特别指出的是，大数据分析所提供的恐怖主义袭击以及其他暴力事件的分析和预测主要服务于追踪性预防，并不致力于提供根除此类暴力行为的建议，更无法替代深入实地研究以厘清暴力根源的努力。但不同的研究路径并非对立，相反，它们在不同层次和时间窗口上，可以对控制和降低恐怖主义袭击的共同事业提供相互补充的理解和应对建议。

二、大数据、GIS 与恐怖主义研究

对暴力冲突事件发生的时空规律的把握和应用，是大数据冲突研究区别于传统研究的一个重要方面，GIS 技术作为与大数据伴生的新工具对此具有尤为突出的地位和作用。[②] 时空规律对于描述、解释和预测冲突的发生、蔓延和持续十分重要，大数据为把握时空规律提供了强大的信息和方法支撑。国际冲突、领土争端、边界谈判一直是国际安全研究中的焦点议题，而地缘政治因素是其中毋庸置疑的关键变量，但由于宏观地理空间的稳定性特征无法提供解释变化的信息，地缘因素仅能作为思考的背景与框架，在传统的实证研究中往往被忽略或隐去。[③] 大数据时代的地理空间数据和 GIS 技术的发展，为重新探究地缘政治这一经典概念提供了新的视角和分析路径，为地缘政治的定量研究带来了广阔前景。例如，尼尔斯·魏德曼等学者利用 GIS 技术研究提供了国家行政疆域历史变迁的空间数据，追踪国家在国际体系中的

① 赵敏燕等：《“一带一路”沿线国家安全形势评估及对策》，《中国科学院院刊》，2016 年第 6 期；周亦奇、封帅：《安全风险分析的方法创新与实践——以“一带一路”政治安全风险数据库建设为例》，《国际展望》，2017 年第 5 期。

② Nils B. Weidmann and Monica Duffy Toft, “Promises and Pitfalls in the Spatial Prediction of Ethnic Violence: A Comment”, *Conflict Management and Peace Science*, Vol. 27, No. 2, 2010, pp. 59-176.

③ Kristian Skrede Gleditsch and Nils B. Weidmann, “Richardson in the Information Age: Geographic Information Systems and Spatial Data in International Studies”, *Annual Review of Political Science*, Vol. 15, No. 1, 2012, p. 462.

演化过程。[①] 肯尼斯·舒尔茨运用GIS技术绘制了领土争端的地理位置，[②] 将地缘影响分析的精度提升到以1千米为单位，以准确检验领土争端中的物质因素和战略位置的作用。[③] GIS技术在国际关系研究中迅速上升的热度，表明地缘政治回归国际关系的科学探寻。在本质上，GIS技术重新定义了国际关系研究中的“空间”概念，空间精度的大幅提升，对于研究非传统行为体（如恐怖主义组织）实施的政治暴力事件极具价值。[④] 因此，GIS技术生成的微观层次数据在近年来高速积累，可以精细描述微观地理单元在政治、人口、社会、经济、环境等各个维度的特征。[⑤] 与此同时，GIS技术也改变了传统地缘政治的静态特征，时间被视为空间的一个维度，地缘因此不再是一个静态的存在，空间的动态化极大地提高了地缘因素的解释和预测能力。[⑥]

在国际关系研究所关注的政治暴力类型中，恐怖主义爆发的地缘因素尤为引人注意，因为恐怖袭击者对时间和地点的选择均经过精心设计。由于忽略对时空地缘规律的考察，传统的恐怖主义研究对恐袭“随机性”的特征常感到束手无策。但是，正因为这种精心设计的“随机性”具有非随机本质，某种时空规律必然存在于众多看似孤立事件的背后，只是传统的数据和方法无法让这些隐藏的时空规律浮出表面，[⑦] 而需要深入探索恐怖主义袭击的时间延续性和空间关联性。[⑧] 在已有研究中，理查德·梅迪纳等学者运用GIS技术考察了2004-2009年间发生于伊拉克的恐怖袭击的时空模式，发现恐怖

① Nils B. Weidmann, Doreen Kuse and Kristian Skrede Gleditsch, “The Geography of the International System: The CShapes Dataset”, *International Interactions*, Vol. 36, No. 11, 2010, pp. 86-106.

② Kenneth A. Schultz, “Mapping Interstate Territorial Conflict: A New Data Set and Applications”, *Journal of Conflict Resolution*, Vol. 61, No. 7, 2017, pp. 565-590.

③ Hein Goemans and Kenneth Schultz, “The Politics of Territorial Claims: A Geospatial Approach Applied to Africa”, *International Organization*, Vol. 71, No. 1, 2017, pp. 31-64.

④ Yuri M. Zhukov, “Roads and the Diffusion of Insurgent Violence: The Logistics of Conflict in Russia's North Caucasus”, *Political Geography*, Vol. 31, No. 3, 2012, pp. 144-156.

⑤ Jordan Branch, “Geographic Information Systems (GIS) in International Relations”, pp. 845-869; Kristian Skrede Gleditsch and Nils B. Weidmann, “Richardson in the Information Age: Geographic Information Systems and Spatial Data in International Studies”, pp. 461-481.

⑥ Halvard Buhaug and Kristian Skrede Gleditsch, “Contagion or Confusion? Why Conflicts Cluster in Space”, *International Studies Quarterly*, Vol. 52, No. 2, 2008, pp. 215-233.

⑦ André Python, et al., “A Bayesian Approach to Modelling Subnational Spatial Dynamics of Worldwide Non-State Terrorism, 2010-2016”, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, Vol. 182, No. 1, 2019, pp. 323-344.

⑧ Alex Braithwaite and Quan Li, “Transnational Terrorism Hot Spots: Identification and Impact Evaluation”, *Conflict Management and Peace Science*, Vol. 24, No. 4, 2007, pp. 281-296.

袭击的分布在时间和空间上并非随机，而是与人口密度相关。[①] 伊斯梅尔·奥纳特也运用 GIS 技术检验了 2008—2012 年间伊斯坦布尔的恐怖袭击的空间特征，同样发现恐怖袭击在空间上明显呈现出集聚特征。[②] 但是上述研究都忽略了恐怖袭击的时间维度，而且对一国或一城分析所得的空间规律是否具有普遍适用性的问题也值得探讨。

恐怖袭击研究中的地缘因素具有非静态特征，这就要求对恐袭风险在时间和空间维度上的蔓延过程进行动态追踪。学界近期的研究对此进行了有益的尝试，例如亚历克斯·布雷斯韦特和李泉考察了 1975—1997 年间 112 个国家的恐怖袭击事件，发现恐袭“热点”随时间变化而移动。但是这种以国家为分析单元的恐袭热点动态追踪过于宏观，忽视了国内不同地区的风险动态变化。另外一些研究尝试改进他们的方法，在时空网格单元上对恐袭热点进行识别并追踪其动态过程。[③] 利用识别的网格层次的恐袭热点，进而用当地地理和人文环境因素来解释恐袭热点的形成。然而，这些研究将明显不存在恐袭风险的地方（如海洋、荒漠以及其他无人居住地区）均纳入分析之中，并用同一个模型来分析具有高度异质性的不同地区，事实上影响了分析结果的可靠性。

运用大数据方法将恐怖主义袭击研究在空间上推向微观层次，同时考察恐怖主义事件发生的时空规律和追踪其时空动态特性，对创新恐怖主义研究路径和提升研究的应用价值极具潜力。然而，现有恐怖主义大数据研究在“普遍性—精确性”平衡方面尚待提高且缺乏预测功能。本文在现有研究的基础上，一方面力求探寻恐怖袭击的普遍性时空规律，另一方面承认观察单元的异质性，将普遍的规律和各异的特质结合起来，意在更为贴切地描述和解释恐袭风险，并且提高恐袭预测的针对性和准确率。

---

① Richard M. Medina, Laura K. Siebeneck, and George F. Hepner, “A Geographic Information Systems (GIS) Analysis of Spatiotemporal Patterns of Terrorist Incidents in Iraq 2004-2009”, *Studies in Conflict & Terrorism*, Vol. 34, No. 11, 2011, pp. 862-882.

② Ismail Onat, “An Analysis of Spatial Correlates of Terrorism Using Risk Terrain Modeling”, *Terrorism and Political Violence*, Vol. 31, No. 2, 2019, pp. 277-298.

③ 参见 Stephen C. Nemeth, Jacob A. Mauslein, and Craig Stapley, “The Primacy of the Local: Identifying Terrorist Hot Spots Using Geographic Information Systems”, *The Journal of Politics*, Vol. 76, No. 2, 2014, pp. 304-317; Josiah Marineau, et al., “The Local Geography of Transnational Terrorism”, pp. 1-32。

三、时空网格中的恐怖主义袭击

传统的政治空间多以行政区划定义，这源于对行政单位的统计数据的依赖。在大数据时代，数据来源多样化，官方统计数据不再是唯一甚至重要的数据来源，研究者可以根据研究需求灵活选择空间层次。在冲突研究中，GIS技术让冲突区（conflict zone）、族群、村落等成为可供定量分析的空间单元。[①] 但空间层次选择的灵活性也使得不同层次的数据难以整合。为解决这一问题，地理网格化方法应运而生，它不但进一步提高了空间单元选择的灵活性，也便于根据统一的标准和编码对地理编码数据进行整合。此外，地理网格独立于政治实体之外，外生于研究对象（如内战），可以很好地避免分析中的内生性问题。[②] 本文将在地理网格单元观察和分析恐怖主义袭击事件。

（一）时空网格：细粒度分析单元

地理网格化方法中最著名的是奥斯陆和平研究所（PRIO）发布的“时空网格”（PRIO-GRID）数据。[③] 时空网格以0.5×0.5经纬度作为分辨率，涵盖了1946—2014年全球所有国家和地区的陆地和海洋。[④] 由于地球的球面性质，0.5×0.5经纬度定义的网格在面积上不等，以赤道上网格的面积最大，约为55千米×55千米（约3200平方千米），相当于北京市面积的五分之一。随着纬度的增加，网格的面积会逐渐减少，在南北极点面积为0。全球网格数据一共包括360行×720列，共259200个网格。由于地球大部分为海洋，全球陆地共包含64818个网格。[⑤] 数据库使用最主流的“世界大地坐标系”

① Andreas Forø Tollefsen, Håvard Strand and Halvard Buhaug, “PRIO-GRID: A Unified Spatial Data Structure”, p. 365.

② Hanne Fjelde, Lisa Hultman, and Desirée Nilsson, “Protection Through Presence: UN Peacekeeping and the Costs of Targeting Civilians”, *International Organization*, Vol. 73, No. 1, 2019, pp. 103-131.

③ 由于PRIO-GRID数据既包括空间也包括时间维度的变量，我们因此将其简称为“时空网格”。

④ 关于网格的设计和介绍，参见Andreas Forø Tollefsen, Håvard Strand and Halvard Buhaug, “PRIO-GRID: A Unified Spatial Data Structure”, pp. 363-374；陈冲：《机会、贪婪、怨恨与国内冲突的再思考——基于时空模型对非洲政治暴力的分析》，《世界经济与政治》，2018年第8期。

⑤ 陆地网格指网格中陆地面积须大于0.01平方千米，略大于一个足球场，否则视为海洋网格。

（即 WGS84），因此非常便于与其他空间数据如 GeoEPR[①]、UCDP Georeferenced Event Dataset[②]等合并。时空网格数据的开发是在关系数据库管理系统（RDBMS）中完成的。具体来说，研究者使用 PostgreSQL 对象—关系型数据库，利用 Python 语言将 SQL 数据库存储的空间地理数据对接起来，并利用 Python 语言循环地将不同时间维度的数据统一自动加以整合。[③] 时空网格数据一共包括两部分：静态数据和动态的年度数据。静态的数据主要涉及恒定的变量，如地形和网格的面积等。而动态数据以年份作为单位，目前的版本包括 1946—2014 年间的一系列具有空间性质的数据，其中最具代表性的是以网格为单位计算的夜间灯光、降水、网格中心到首都和边界的最短距离等动态变量。[④]

本文的分析聚焦于非洲的恐怖袭击风险，将非洲进行网格化，得到共计 10674 个时空网格。[⑤] 相比于以国家为空间单元（非洲共 53 个国家），网格化在空间层次上的分辨率提高约 200 倍，研究的细粒度大为提高。网格划定后，本文根据恐怖袭击及其他变量的地理位置和时间信息将它们叠加到对应的时空网格中，重组分析单元各个维度的信息。

（二）恐怖袭击的大数据：时机和空间

在时空网格为单元的细粒度上，我们可以更为细致地观察非洲恐怖袭击的分布情况。本文使用的“全球恐怖主义数据库”（Global Terrorism Database，简称 GTD）为目前世界上关于恐怖袭击最全面的公开数据库。[⑥] GTD 将恐怖主义袭击定义为“非国家行为体威胁使用或实际使用非法的武力或者暴力，

① Julian Wucherpfennig, et al., “Politically Relevant Ethnic Groups Across Space and Time: Introducing the GeoEPR Dataset”, *Conflict Management and Peace Science*, Vol. 28, No. 5, 2011, pp. 423-437.

② Ralph Sundberg and Erik Melander, “Introducing the UCDP Georeferenced Event Dataset”, *Journal of Peace Research*, Vol. 50, No. 4, 2013, pp. 523-532.

③ 参见 Andreas Forø Tollefsen, Karim Bahgat, Jonas Nordkvelle, and Halvard Buhaug, “PRIO-GRID v. 2. 0 Codebook”, June 2, 2016, http://www.nber.org/ens/feldstein/ENSA_Sources/UCDP_PRIO/PRIO%20GRID/PRIO-GRID-Codebook.pdf。

④ 时空网格的距离等变量来源于 CShapes 数据，因而考虑了首都、边界等的变迁问题。参见 Nils B. Weidmann, Doreen Kuse, and Kristian Skrede Gleditsch, “The Geography of the International System: The CShapes Dataset”, pp. 86-106。

⑤ 由于南苏丹 2011 年独立后并没有收录在时空网格数据中，因此 2011—2013 年的时空网格总数为 10469 个。

⑥ Gary LaFree and Laura Dugan, “Introducing the Global Terrorism Database”, *Terrorism and Political Violence*, Vol. 19, No. 2, 2007, pp. 81-204.

通过恐惧、强制或恐吓从而达到其政治、经济、宗教或者社会目标”。[①] GTD进而对恐怖主义袭击事件进行分类，包括暗杀、劫机、绑架、爆炸、武装袭击、破坏设施等不同类型，自1970年以来全球共记录超过18万个恐怖袭击事件。数据库中每个袭击事件包含45—120个维度上的信息，不但包含精确的时间和地理信息，还包含如使用武器类型、死亡人数、是否自杀性袭击等变量。根据GTD数据库的记录，非洲在1970—2016年间共发生21689次恐怖袭击，其中暗杀2222次、劫机145次、绑架1980次、武装袭击7079次。

图-1（a）报告了非洲1970—2016年间恐怖袭击类型的时间分布。[②] 从时间趋势看，非洲的恐怖主义风险整体呈增长趋势。自1970年以来，恐怖主义袭击数量在1988—2000年出现第一个高峰期，2000—2003年短暂回落后持续上升，在2010年之后出现暴增，各种类型恐怖袭击的数量均呈跳跃式增长，爆炸和武装袭击尤其高发，在2014—2015年间达到顶峰，而2016年虽有所下降但仍保持高位。

为了观察恐袭事件的空间分布，本文根据袭击事件的时间和地理编码信息，将这些恐袭事件投射到非洲国别行政版图上。如图-1（b）所示，1970—2016年，恐怖袭击在地理位置上呈现出高度的集聚特征，主要集中在北非地中海沿岸、萨赫勒地区（伊斯兰马格里布基地组织在2009年后活跃于该地区）、刚果民主共和国北部与南苏丹临近地区、东北部边境地区、索马里南部和南非东部地区。鉴于2004年和2014年分别为进入21世纪以来非洲恐怖袭击最低和最频繁的年份，图-1（c）和（d）分别展示了2004年和2014年恐怖袭击呈现出的“蔓延”趋势。这两幅分布图的一个明显变化是尼日利亚和非洲东北部索马里恐怖袭击的大量增加。值得一提的是，如图-1所示，恐袭频率在同一个国家内部不同地区之间的差异度并不亚于不同国家之间的差异。例如，在通常以国家为单元的认识中，北非的阿尔及利亚、突尼斯、利比亚、埃及均为恐袭的高发国家，但是，从图-1（b—d）可以看到，恐袭的风险主要集中在其国土的沿地中海地区，而其国内其他绝大部分地区的恐袭风险并不比相对安全的马达加斯加更高，尤其是南部的沙漠地区鲜有恐袭

① 参见“Codebook: Inclusion Criteria and Variables”, October 2019, https://www.start.umd.edu/gtd/downloads/Codebook.pdf。在实践中，恐怖主义袭击必须同时满足三个标准：该事件必须是有意图的，必须包含一定程度的暴力或者暴力威胁，施害者必须是非国家行为体。关于恐怖主义定义的更多讨论，参见Joseph K. Young and Michael G. Findley, “Promise and Pitfalls of Terrorism Research”, pp. 411-431。

② 本文图片全部依次列在文章后面。

发生。由此可见，以国家为单元的恐袭风险分析针对性相当不足，不但容易遗漏真正的重要因素，而且政策应用价值也很低。

相反，以时空网格为地理观察单元，我们则可以更细致地考察恐袭分布的时空规律。利用GIS技术，本文将恐袭事件叠加到对应的10674个时空网格中进行分析。[①] 图-2以尼日利亚2009—2012年间发生的892次恐怖袭击为例，观察风险在空间上“蔓延”的情况。尼日利亚在2009年发生了42次恐怖袭击，主要集中在其国土南部的三角洲州和河流州，而除去东北部的博尔诺州之外，大部分地区属于“安全”地带。进入2010年，由于“博科圣地”的崛起，[②] 恐怖袭击开始增多（63次），并向中部蔓延。这一趋势在2011年进一步恶化（174次），中部和东北部的恐怖袭击增多，而南部开始减少。到了2012年，尼日利亚的恐怖袭击达到了613次，集中分布在中部的卡杜纳州和东北的包奇州、约贝州以及博尔诺州。这些存在于国家内部的恐怖袭击的差异性和动态趋势是在以国家为单元的研究中无法观察到的，也进一步显示了GIS方法的优势。[③]

本文将对恐怖袭击的发生风险和时机进行研究，将因变量定义为某网格是否在某年遭遇恐怖袭击，数据形式为以时空网格为地理单元、以年度为时间单元、以恐怖袭击发生与否为因变量的面板数据（panel data）。由于受到解释变量网格数据可得性的限制，样本时间段选定在1994—2013年。[④]

四、模型和变量

本文以恐袭时机为研究问题，将恐袭事件的发生看成是安全状态的结束，试图回答安全状态在各个时间点上脆弱性（恐袭风险）的问题。本文的

① 本文使用R软件中的GIS技术进行叠加，相关代码参见本文在Harvard Dataverse公布的复制数据。

② Max Gallop，Shahryar Minhas，and Cassy Dorff，“Networks of Violence：Predicting Conflict in Nigeria”，*Journal of Politics*，forthcoming.

③ 值得特别说明的是，时空网格划分的GIS技术给予研究者自主选择分辨率的自由。本文选择0.5×0.5经纬度作为分辨率是考虑到解释变量的可得性。在完全进行描述和追踪的研究中，研究者可以以更高的精度进行研究，例如0.01×0.01经纬度，即相当于赤道上1平方千米的面积。

④ 主要受到夜间灯光数据的限制，该数据涵盖了1992—2013年。同时，由于GTD曾经丢失了1993年的数据，即使后来恢复了一些，依然是不全面的。鉴于此，我们将研究的时间段限制在1994—2013年。我们期待随着夜间灯光数据的更新，未来的研究可以进一步延伸时间跨度。

模型和变量即是基于这一研究任务而进行选择的。

（一）持续期模型：恐怖袭击的“时机”

基于本文的研究任务，持续期模型（或名生存模型）是一个显而易见的模型选择。持续期模型是对一种状态结束的时机的分析方法。例如，不受恐怖袭击的安全状态是一种状态，恐袭事件的发生即是安全状态的结束，而持续期模型所关注的研究问题就是安全状态结束的“时机”——在任何时间点上这种状态均有可能结束（即恐袭均有可能发生）。但在任何时间点上，恐袭风险会因之前安全状态的持续期长短而不同。例如，保持长时间安全状态的地区遭受袭击的风险会增大，因为恐怖分子预计长时间的安全会使该地区放松防范，恐袭更易成功，从而选择这种地区作为恐袭对象。或者安全状态的持续期越长，恐袭的危险也有可能越小，因为持续期拉长或许表明稳定社会的形成、防恐体系的构建、对恐怖主义根源的成功治理等。本文将使用持续期模型对这一时机问题进行实证考察。①

然而，对研究恐怖袭击而言，传统的持续期模型有一个重要缺陷，即它假定一种状态的持续期终将结束。虽然这一假定适用于某些研究，尤其是生物学中关于生存持续期的研究，因为生物体终将死亡，但是对于恐怖袭击而言，尤其是当目标单元细化到时空网格层面，我们很难合理地假定所有的网格恐袭最终都会发生。换言之，并非所有网格均会“必有一袭”。承认这一网格之间的异质性非常重要，不但有助于深化对恐袭风险的理解，而且在反恐防恐实践中排除安全地区以提高效率十分必要。例如，图-1（b）对1970—2016年非洲所有恐怖袭击的空间分布描述表明，大量地区在1970—2016年从未发生过恐怖袭击。从网格层次上看，遭遇恐怖袭击的单元毕竟是极少数（即稀缺事件），从而使得“持续期终将结束，只是时间问题”的假定尤其不合理，对异质性的考虑更为必要。鉴于此，本文使用了一种不同于

① 持续期模型在国际关系中的应用，参见 Janet M. Box-Steffensmeier and Christopher J. W. Zorn，“Duration Models and Proportional Hazards in Political Science”，*American Journal of Political Science*，Vol. 45，No. 4，2001，pp. 972-988；Bruce Bueno De Mesquita and Randolph M. Siverson，“War and the Survival of Political Leaders：A Comparative Study of Regime Types and Political Accountability”，*American Political Science* Review，Vol. 89，No. 4，1995，pp. 841-855；Caroline Hartzell and Matthew Hoddie，“Institutionalizing Peace：Power Sharing and Post-Civil War Conflict Management”，*American Journal of Political Science*，Vol. 47，No. 2，2003，pp. 318-332。

传统持续期模型的混合（mixture）持续期模型——分离总体持续期模型（SPDM）。[①]

（二）分离总体持续期模型：区分疫区与免疫区

作为持续期模型的一种，SPDM 较好地处理了时空网格是否终究会遭受恐怖袭击问题上的异质性。SPDM 承认一些网格由于各种原因（如地理条件、人口稀少、防卫森严等）对恐袭或许具有“免疫力”而处于恐袭“免疫区”，而另一些地区无论是否已经遭遇恐袭，都处于恐怖袭击的“疫区”，具有感染风险，是否实际遭到恐袭只是时间问题。

我们用 $Z\in\{0,1\}$ 来表示网格是否经历恐怖袭击，$Z=1$ 表示发生袭击，$Z=0$ 表示没有发生。由于我们无法观测到数据中右截尾（right-censored）——在观察期结束还没有发生袭击——的网格是否经历恐怖袭击，因此事件 Z 对于右截尾（$Z=0$）的网格而言是无法观测的潜在事件（latent event）。换言之，我们可以观察到 $Z=1$，即事件最终发生，但无法观察到没有发生恐袭。因为在任何时间点上观察到事件没有发生并不等于观察到“事件最终没有发生”，即使分析单元消亡，我们仍然不能认为观察到了 Z。用 π 来表示网格经历袭击的概率，即

$$\Pr(Z=1)=\pi$$
$$\Pr(Z=0)=1-\pi$$

然后，定义另一类事件——“网格 i 在安全状态持续了 t 时长后是否发生恐怖袭击”，记为 $Y_i \mid t=Y_{it}$，发生为 1，不发生为 0。事件 Y 是一个可以被观察到的事件，即使对未来的 Y 进行预测，在所预测的时间来到之时，Y 是可以观察到的。在区分样本异质性的基础上，SPDM 对恐袭风险随时间推移的动态变化进行分析。SPDM 可以修改传统生存模型的密度函数 $f(t)$ 为条件分布 $f(t \mid Z=1)$，只有最终会经历的事件才有必要在不同时间点上考察风险密度。考虑到时空网格在恐袭风险上的异质性，我们对 Y_{it} 的不同情况进行以下区分：首先，当 $Y_{it}=1$ 时，一定是 $Z_{it}=1$，因为既然已经经历了恐怖袭击，则“恐怖主义袭击最终是否会发生”这一事件也即发生。在这种情

① SPDM 在国际关系中的主要应用，参见 Milan Svolik，“Authoritarian Reversals and Democratic Consolidation”，*American Political Science Review*，Vol. 102，No. 2，2008，pp. 153-168；Andreas Beger，Cassy L. Dorff and Michael D. Ward，“Irregular Leadership Changes in 2014：Forecasts Using Ensemble，Split-Population Duration Models”，*International Journal of Forecasting*，Vol. 32，No. 1，2016，pp. 98-111；Michael D. Ward and Andreas Beger，“Lessons from Near Real-Time Forecasting of Irregular Leadership Changes”，*Journal of Peace Research*，Vol. 54，No. 2，2017，pp. 141-156。

况下，时空网格在持续期 t 时的恐袭概率可以表达为：

$$\Pr(Y=1 \mid t)=\Pr(Z=1) f(t \mid Z=1)=\pi f(t \mid Z=1)$$

第二种情况是 $Y=0$，即没有经历恐怖袭击。$Y=0$ 则有两种可能，一是网格对恐怖袭击具有“免疫能力”，即 $Z=0$，恐袭始终不会发生，$\Pr(Z=0)$；另一种可能是，最终也会遭到恐怖袭击，但只是时机未到，即 $(Z=1) * s(t \mid Z=1)$。于是 $Y=0$ 的概率为两种可能性的概率相加，可以表达为：

$$\begin{aligned}\Pr(Y=0 \mid t) &= \Pr(Z=0)+\Pr(Z=1) * s(t \mid Z=1) \\ &= (1-\pi)+\pi s(t \mid Z=1)\end{aligned}$$

综合 Y 的两种可能情况，模型可更普遍地表达为：

$$\Pr(Y \mid t)=[\pi f(t \mid Z=1)]^{\delta} \times[(1-\pi)+\pi s(t \mid Z=1)]^{(1-\delta)}$$

其中，δ 是指示函数（indicator function）$\in\{0, 1\}$，取值为1时在 $<t$ 这一观察期（spell）结束遭遇到恐怖袭击，而0则为在该观察期未发生恐怖袭击。在此基础上，对于 n 个地理网格观察单元，SPDM模型的似然方程（likelihood function）可以表达如下：

$$L\{\theta \mid t_1, \cdots, t_n\}=\prod_{i=1}^{N}\{\pi_i f(t_i)\}^{\delta_i} \times\{(1-\pi_i)+\pi_i s(t_i)\}^{1-\delta_i}$$

值得注意的是，在SPDM中，π 是通过 $f(t \mid Z=1)$［或者 $s(t \mid Z=1)$］联合估计（estimated jointly）的。

最后需要决定似然方程中风险函数 $f(t)$ 以及生存函数 $s(f)$ 的分布，以此决定危险率（hazard rate），$h(t)=\frac{f(t)}{s(t)}$。本文采取比例参数（scale parameter）为 λ，形状参数（shape parameter）α 的log-logistic分布。[①] Log-logistic分布允许估计的危险率在特定的生存时间上有一个峰值。在log-logistic中，如果 $\alpha>1$，那么危险率严格递减；如果 $0<\alpha<1$，那么危险率先递增，然后递减。[②] 这一统计分布假定也符合图-1中我们所观察到的恐袭分布特征。我们使用 R 统计软件中的spduration包来估计模型。[③]

（三）变量选择

为了考察恐怖袭击的空间关联和空间蔓延规律，并对这些关联和规律进

① Andreas Beger, et al., “Splitting It up: The Spduration Split-Population Duration Regression Package for Time-Varying Covariates”, *R Journal*, Vol. 9, No. 2, 2017, p. 476.

② Milan Svolik, “Authoritarian Reversals and Democratic Consolidation”, p. 158.

③ Andreas Beger, et al., “Spduration: Split-Population Duration (Cure) Regression”, R Package Version 0.17.0, 2017, https://CRAN.R-project.org/package=spduration. 本文的分析代码和数据可通过Harvard Dataverse下载。

行解释和预测，本文在时空网格层次上构建和整合了一系列变量，包括时空变量和其他在恐怖主义研究中常见的解释和预测变量。由于这些变量数量众多，本文采用理论和机器算法相结合的降维方法，最终选取了 20 个变量建立解释和预测模型。[①] 下文主要对时空变量的构建和变量选择的方法和结果进行报告。

1. 关于微观风险的时空变量

本文构建用于捕捉恐袭地缘特征的时空变量主要如下。首先，为了分析恐袭在空间上分布的集聚特征，本文根据沃尔多·托布勒地理学第一定律——“任何事物都是与其他事物相关的，只不过邻近的事物关联更紧密”，[②] 考察相邻地理网格恐袭风险的“传染性”。由于分析单元为 10674 个地理网格，本文创建了一个 10674×10674 维度的“邻接矩阵”（adjacency matrix）W_{ij}。[③] 矩阵中每一个元素 w_{ji} 的值由网格 i 与网格 j 是否在地理上相邻决定，相邻为 1，否则为 0（1＝是，0＝否），其中 w_{ii} 设为 0。由于“相邻”是个相对概念，我们确切定义相邻的标准为“k-最近相邻”（k-nearest）方法，[④] 设定 k＝10（即搜索的邻近地理网格数量最多为 10），以半径为 5 个单位，搜索以网格中心为圆心的方圆 390 千米（即 5 个单位的半径×78 千米）范围的最多 10 个相邻地理网格为相邻网格，即它们与 i 网格的 w_{ij} 取值为 1，而邻接矩阵第 i 行的其他元素为 0。[⑤] 用这样的方法可以得到邻接矩阵中每一个元素的取值。然后，我们可以生成空间滞后变量 WY_{t-p}，其中 Y_{t-p} 为所有网格滞后 p 年的恐袭发生向量。本文取 p＝1，2，即考察滞后一期和两期的恐袭危险时空蔓延。

恐袭地点的选择往往十分微妙，不但在于袭击实际所能造成的伤害，而

---

① 本文对缺失值进行了“多重替代”（multiple imputation）。利用贝叶斯高斯关联结果替代（Bayesian Gaussian copula imputation）方法，通过 10000 次马尔科夫—蒙特卡洛（Markov Chain Monte Carlo，MCMC）将缺失值进行了替代。参见 Peter D. Hoff，“Extending the Rank Likelihood for Semiparametric Copula Estimation”，*The Annals of Applied Statistics*，Vol. 1，No. 1，2007，pp. 265-283。

② Waldo R. Tobler，“A Computer Movie Simulating Urban Growth in The Detroit Region”，*Economic Geography*，Vol. 46，No. S1，1970，pp. 234-240.

③ 由于南苏丹独立，2011—2013 年的邻接矩阵为 10469 ×10469。

④ Sunil Arya，et al.，“An Optimal Algorithm for Approximate Nearest Neighbor Searching Fixed Dimensions”，*Journal of the ACM*（*JACM*），Vol. 45，No. 6，1998，pp. 891-923.

⑤ 关于该方法的介绍，参见陈冲：《机会、贪婪、怨恨与国内冲突的再思考——基于时空模型对非洲政治暴力的分析》，第 107 页。我们使用 *R* 中的 RANN 软件包进行“k-最近相邻”分析，参见 Sunil Arya，et al.，“RANN：Fast Nearest Neighbour Search（Wraps ANN Library）Using L2 Metric”，R Package Version 2. 5. 1.，2017，https：//CRAN. R-project. org/package＝RANN。

且受害地区还需具有象征意义并受到更多媒体关注。一国的首都往往是最有象征意义的城市，更能吸引公众和媒体注意力，能够达到恐袭需要公开宣传的效果。[①] 距离首都近的地理单元可能被“殃及池鱼”而更容易遭到恐袭，因此空间上与首都的距离很有可能是影响恐袭风险的一个重要地缘因素，这一变量的数据来源于PRIO-GRID项目。

不但到“腹地”的距离会影响恐袭风险，到边界的距离也同样可能与风险相关。毗邻边界为进行袭击的恐怖主义分子脱身提供了便利，成为比较方便的恐袭目标地点。另外，距离边界近的地方，更容易受邻国因素的影响而增加易感性。例如，来自邻国的难民、移民等更容易聚集在边界地区生活，这就增加了网格“感染”恐怖袭击的风险，甚至增加了这些难民和移民被招募进入恐怖组织的可能。距离边界近的地方也增加了恐怖组织成功逃避政府追捕和反恐打击的可能性，甚至以邻国作为避难所。[②] 距离边界距离的数据来自PRIO-GRID数据项目。

2. 关于宏观风险的时空变量

恐袭的风险不但受到最直接邻近区域的风险影响，而且宏观的风险形势对于微观单元的风险状况也具有广泛的地缘影响。本文着眼于国家层次和国际层次的风险如何侵蚀局部地区的安全状况。[③] 从国家层次来看，当一国过去本身遭受较多恐怖袭击时，其未来被袭击的整体概率也可能增加，因此我们纳入本国恐怖袭击总数。[④] 此外，邻近国家处于高烈度政治动荡，可能滋生恐怖主义，而恐怖主义集团的活动和藏身会直接殃及邻近国家的特定地区。同时，内战等高烈度政治动荡还会产生大量的难民跨境流动，[⑤] 也会增加恐怖袭击的可能性。[⑥] 本文用两个变量来考察邻国政治动荡对恐袭风险的

① Konstantinos Drakos and Andreas Gofas, “The Devil You Know but Are Afraid to Face: Underreporting Bias and Its Distorting Effects on the Study of Terrorism”, *Journal of Conflict Resolution*, Vol. 50, No. 5, 2006, pp. 714-735.

② Idean Salehyan, “Transnational Rebels: Neighboring States as Sanctuary for Rebel Groups”, *World Politics*, Vol. 59, No. 2, 2007, pp. 217-242.

③ Melissa M. Lee, “The International Politics of Incomplete Sovereignty: How Hostile Neighbors Weaken the State”, *International Organization*, Vol. 72, No. 2, 2018, pp. 283-315.

④ Sara Jackson Wade and Dan Reiter, “Does Democracy Matter? Regime Type and Suicide Terrorism”, *Journal of Conflict Resolution*, Vol. 51, No. 2, 2007, pp. 329-348.

⑤ Idean Salehyan and Kristian Skrede Gleditsch, “Refugees and the Spread of Civil War”, *International Organization*, Vol. 60, No. 2, 2006, pp. 335-366.

⑥ Vincenzo Bove and Tobias Böhmelt, “Does Immigration Induce Terrorism?” *The Journal of Politics*, Vol. 78, No. 2, 2016, pp. 572-588.

空间影响[①]——“邻国是否处于内战”和“邻国是否处于国际冲突”。邻国的定义采取战争相关数据库（COW）数据中“直接毗邻”（Direct Contiguity，v3.2）的标准。[②] 内战的数据则来自乌普萨拉冲突数据项目。[③] 这两个变量来源于凯特林·韦伯斯特等的复制数据。[④] 最后，从国际的宏观风险形势来看，当整个非洲的恐袭风险增大时，也会危及微观网格单元的安全，尤其是整个非洲大陆处于恐怖袭击的高峰时期，所谓“覆巢之下安有完卵”。为此，我们引入了变量“非洲当年恐袭总数”来反映恐袭风险的宏观程度。

3. 机器学习降维与变量选择

除了时空关联关系能够帮助我们描述、解释和预测恐怖主义袭击之外，恐怖主义作为复杂的国际政治现象，有众多层面和维度上的因素均可影响恐袭发生的时机和风险。GTD 和 PRIO-GRID 两个数据库中有超过 200 个相关备选解释变量，恐怖主义研究的文献中还有众多传统上经常使用的国家或者国际层次数据，再加上本文构建的用以探索时空依赖特征的变量，用以建立 SPDM 模型的备选变量超过 300 个。这就形成了高维度的数据，需要采用大数据降维方法进行建模。[⑤] 本文采取根据理论与贝叶斯模型筛选相结合的方法，[⑥] 首先根据理论和文献筛选出对恐怖主义最为相关的变量，[⑦] 继而运用“贝叶斯自适应抽样算法”（Bayesian Adaptive Sampling algorithm，BAS），[⑧] 采

① Alex Braithwaite and Tiffany S. Chu，“Civil Conflicts Abroad，Foreign Fighters，and Terrorism at Home”，*Journal of Conflict Resolution*，Vol. 62，No. 8，2018，pp. 1636-1660.

② Douglas M. Stinnett，Jaroslav Tir，Paul F. Diehl，Philip Schafer，and Charles Gochman，“The Correlates of War（COW）Project Direct Contiguity Data，Version 3.0.”，*Conflict Management and Peace Science*，Vol. 19，No. 2，2002，pp. 59-67.

③ Nils Petter Gleditsch，et al.，“Armed Conflict 1946-2001：A New Dataset”，*Journal of Peace Research*，Vol. 39，No. 5，2002，pp. 615-637.

④ Kaitlyn Webster，Chong Chen，and Kyle Beardsley，“Conflict，Peace，and the Evolution of Women's Empowerment”，*International Organization*，Vol. 73，No. 2，2019，pp. 255-289.

⑤ Burt L. Monroe，“The Five Vs of Big Data Political Science：Introduction to the Virtual Issue on Big Data in Political Science Political Analysis”，pp. 1-9；David Lazer and Jason Radford，“Data ex Machina：Introduction to Big Data”，*Annual Review of Sociology*，Vol. 43，No. 1，2019，pp. 19-39；庞珣：《定量预测的风险来源与处理方法——以“高烈度政治动荡”预测研究项目的再分析为例》，《国际政治科学》，2017 年第 3 期。

⑥ 庞珣：《定量预测的风险来源与处理方法——以“高烈度政治动荡”预测研究项目的再分析为例》。

⑦ Walt L. Perry，et al.，*Predicting Suicide Attacks：Integrating Spatial，Temporal，and Social Features of Terrorist Attack Targets*，Rand Corporation，2013.

⑧ Merlise A. Clyde，Joyee Ghosh，and Michael L. Littman，“Bayesian Adaptive Sampling for Variable Selection and Model Averaging”，*Journal of Computational and Graphical Statistics*，Vol. 20，No. 1，2011，pp. 80-101.

用MCMC抽样的方法根据模型的后验概率进行变量筛选。[①] 根据机器降维建议的变量共20个，其中地理网格层次变量10个、国家层次变量7个、国际层次变量3个，变量描述性统计和数据来源报告在表-1中。

表-1 变量的描述性统计和来源

变量	样本量	均值	标准差	最小值	最大值	来源
国际层次						
邻国处于国际冲突	212，865	0.021	0.143	0	1	Webster et al. 2019
邻国处于内战	212，865	0.981	1.038	0	4	Webster et al. 2019
非洲恐袭总数	212，865	379.623	268.639	0	1，282	GTD
国家层次						
本国恐袭总数	212，865	16.762	46.814	0	606	GTD
本国政体分值	212，865	0.298	5.051	−9	10	Polity IV
本国人均GDP（对数）	212，865	7.150	1.070	4.752	9.920	WDI
本国总人口（对数）	212，865	16.671	1.105	11.215	18.962	WDI
本国军费开支	212，865	13.671	6.838	0.675	88.206	WDI
本国非正常领导更迭	212，865	0.035	0.180	0	1	Archigos
本国内战数量	212，725	0.425	0.598	0.000	2.000	UCDP
网格层次						
网格夜间灯光（最大值）	212，865	7.339	15.526	0	63	PRIO-GRID
网格的山地面积百分比	212，865	0.142	0.260	0.000	1.000	PRIO-GRID
网格距首都距离	212，865	649.146	416.586	3.703	2，482.531	PRIO-GRID
网格到最近邻国距离	212，865	183.776	167.103	0.003	1，971.659	PRIO-GRID
网格中被排斥族群	212，865	0.337	0.594	0	5	PRIO-GRID
网格中政府暴力数	212，865	0.061	1.264	0	240	UCDP-GED
网格蕴藏石油	212，865	0.011	0.106	0	1	PRIO-GRID
网格中恐袭总数，t-1	212，865	0.036	0.930	0	147	GTD
邻近网格遭袭的数量，t-1	212，865	0.322	3.446	0	186	GTD
邻近网格遭袭的数量，t-2	212，865	0.302	3.197	0	186	GTD

① BAS方法已经在R软件包BAS中实现，参见Merlise Clyde，“BAS：Bayesian Adaptive Sampling for Bayesian Model Averaging”，R Package Version 1.4.7。关于BAS的使用介绍，参见Merlise A. Clyde，“Using the Bayesian Adaptive Sampling（BAS）Package for Bayesian Model Averaging and Variable Selection”，https://merliseclyde.github.io/BAS/articles/BAS-vignette.html。机器降维的代码参考本文的复制数据。

五、实证结果及解读

图-6 报告了模型的回归系数及其标准差。如前文所述，SPDM 的分析对象包含两种风险的动态变化：其一是对具有异质性的地理单元的“免疫能力”进行评估（即参数 π_{it}，本文称之为风险函数），其二是遭遇恐怖袭击的时机（即危险率 $h(t)$，本文称之为持续期函数）。下文将对两类风险的实证结果进行报告和解读。

（一）恐怖袭击的风险：疫区与免疫区

由于风险是不可直接观测的存在，我们通过数据和模型对风险进行测量和估计得到 $\hat{\pi}_{it}$。使用 GIS 和 SPDM 的一大优势是可以直观地获得对 t 年 i 网格处于免疫区可能性的估计。

图-3 从时间维度报告了模型估计的免疫区概率分布情况，并与非洲 1994—2013 年间恐袭实际发生情况相比照。结果显示，近年来非洲的恐怖袭击风险在增大，呈现出蔓延趋势。以 $\hat{\pi}<0.05$ 作为临界值来区分“免疫区”与“非免疫区”，图-3（b）展示了 $\hat{\pi}<0.05$ 的“免疫区”的密度分布。$\hat{\pi}$ 越趋近于 0，被袭击的风险也越趋近于 0。在实际中，$\hat{\pi}<0.05$ 意味着网格几乎不存在被袭击的风险（因此是免疫区）。图-3（b）表明，1994—2004 年，$\hat{\pi}$ 趋近于 0 的网格越来越多，表明这段时间整体上非洲的恐怖袭击风险是下降的。2004 年之后，$\hat{\pi}$ 的密度分布开始向右边移动，表明 $\hat{\pi}$ 趋近于 0 的网格在减少，处于免疫区的网格数量在萎缩，进入 2012 年后恐袭风险感染情况严峻，大片原来的安全区域沦陷。将统计估计的疫区蔓延情况与现实相对照，图-3（a）报告了每年尚未遭到恐袭和已遇恐袭的网格数量对比。图中蓝色虚线表示每年尚未观察到袭击的网格数量（左侧 y 轴），黑色实线表示每年观察到经历袭击的网格数量（右侧 y 轴）。2002—2004 年非洲遭遇袭击的时空网格数量明显下降（从 115 个网格减少到 50 个网格），但是 2004 年之后，每年被袭击的网格数量开始明显增多。2013 年非洲遭遇恐袭的网格达到 330 个（约占所有网格的 3%），显示出非洲恐怖主义危险的严峻形势。这也与图-3（b）估计的风险情况变化在趋势上一致。

本文运用 GIS 技术将 $\hat{\pi}_{it}$ 投射到第 t 年的非洲地图对应网格 i 中，得到恐

怖袭击疫区和免疫区的空间分布情况。[①] 为了更清晰地展示疫区和非疫区的分布，图-4 截取尼日利亚恐怖主义疫区在 1994—2013 年的变化作为示例，将实际发生恐怖袭击的网格中心点投射到地图中（图中黑点）。图-4 显示，从 1994 年起的较长一段时间内，尼日利亚境内只有极少部分地区为恐怖主义高感染风险区。但是从 2004 年起，国土南端有一些安全地区陷入疫区，并不断加重和扩大，恐袭感染“病毒”向东北部和中部转移，至 2011 年，东北部和中部大部分地区沦为恐袭易感染区。2012—2013 年，风险向全国扩展，全境变为恐袭易感染区域，尽管在易感程度上仍有地区差异。2012—2013 年的地图清晰显示，尽管尼日利亚西部地区和东南边境地区尚未发生任何恐怖袭击事件，随着疫区在其他区域的扩大和疫情的加重，这些原本具有较强恐袭免疫力的地区的感染风险大幅度提高，能够处于安全区的可能性减小。图-4 的另一个有趣发现是，疫区可以被治愈而成为免疫区，但重复感染的风险依然存在。在 1997 年和 1998 年，尼日利亚境内最为脆弱的地区在西南角，但从 1999 年开始，西南角地区的恐袭易感性逐年下降，到 2005 年呈现出较为安全的态势。但是，随着南端地区的感染情况日渐严重，西南角地区于 2007 年开始被再次感染。图-4 也表明，恐袭疫区的疫情实际爆发情况也更为糟糕，但是易感性和实际爆发恐袭事件并非百分百对应，也即易感性是一种潜在的状态，即使没有实际发生恐袭，这种感染风险依然存在。

（二）疫区内恐怖袭击的时机：危险率

SPDM 通过各种变量对恐袭时机的影响，统计估计出网格在各时间点上遭受恐怖袭击的危险率（$h(t)$），从而绘制出每个时空网格的危险率曲线（hazard curve）。危险率曲线反映了随着“存活”（安全）持续时长而变化的恐袭危险率。本文从 10674 个样本时空网格中随机挑选 4 个危险率估计结果为例进行实证解读。研究者或政策制定者可以挑选样本中任何感兴趣的时空网格或者城市，进行如下的危险率曲线分析。我们随机挑选的 4 个时空网格分别位于刚果民主共和国和阿尔及利亚境内。图-5（a）显示了它们在地图上的位置（图中方格即网格编号）以及它们距离首都金沙萨、阿尔及尔（图中三角形）的相对位置。在 GTD 数据中，网格－172432 在 1994—2013 年间未经历恐怖袭击，而网格－182529 在此期间每年面临 1—22 次不等的袭击。网格－127852 距离首都金沙萨和最近的邻国边界距离分别为 1210 千米和 370

① 完整的彩色版非洲疫区图可在 Harvard Dataverse 下载。

千米，但是1994—2013年间未经历恐怖袭击。网格－135058离首都金沙萨1740千米，距离最近的边界70千米，在2008—2011年间经历了1—8次不等的恐怖袭击。图-5（c）展示了这4个不同时空网格估计出的危险几率在不同年份的变化，其中网格－172432和网格－182529的危险率几乎为0。

在图-5（d—g）中，我们以2010年这4个时空网格的取值来估计其危险率与存活持续期t（survival up to time t）的关系。总体上，网格－127852和网格－172432的危险率非常低，最大值分别低于0.01和0.0025，危险率曲线表明，这两个网格的安全状态很可能持续26年以上。相反，网格－135058和网格－182529的危险率相对较高，2010年的危险率分别为0.1和0.9。网格－135058在之后的4年内危险率依然上升，然后开始下降。特别是网格－182529的危险率在所有年份均偏高，即便在2010年之后的20年内依然存在较大的危险（$\hat{\pi}>0.1$）。值得注意的是，由于网格－182529持续偏高的危险率，图-5（g）所估计的90%置信区间也非常窄，表明估计的危险率具有较高的确定性。

（三）恐怖袭击风险和时机的影响因素

在对恐怖袭击的感染风险和发生时机进行规律总结的基础上，我们进一步考察模型分析结果中各解释变量对恐袭风险和时机的影响程度和方向，加深对恐怖主义袭击时空规律的理解。

1. 影响恐袭风险免疫力的因素

图-6下图是对风险方程估计的结果，其系数解读与标准的Logistic回归系数一致，正系数表示与袭击风险增加相关，负系数则相反。从分析结果可见，时空变量具有显著的影响，对恐怖袭击的“抵抗力”在时间和空间上都具有“毗邻性”。此种时空毗邻性可以通过理论和逻辑得到解释。恐怖袭击是高风险、高成本的暴力活动，重复选择之前袭击地点或者与之前袭击相邻的地点可以利用以往收集的情报信息，因而产生的信息成本更低。[①] 另外，被袭击过的地方往往已经被媒体报道过，提高了“知名度”，而对同一个地方进行重复袭击或对其周边实施恐怖袭击，能够更有效地增加恐慌和摧毁安全信心。此外，时空网格的相对地理位置——距最近邻国边界的距离和距首都的距离——对网格的恐袭易感性均具有显著性的负向影响。从宏观地缘风险来看，邻国发生内战或卷入国际冲突以及非洲的恐怖袭击危险形势（“非

① Richard M. Medina, Laura K. Siebeneck, and George F. Hepner, “A Geographic Information Systems (GIS) Analysis of Spatiotemporal Patterns of Terrorist Incidents in Iraq 2004-2009”, pp. 862-882.

洲恐袭总数”）显著削弱时空网格的恐袭“免疫能力”。宏观地缘风险对微观地理单元恐袭易感性具有显著影响。

除此之外，地缘经济、政治和社会因素也影响到网格单元对恐袭风险的免疫能力（见图-6下图）。夜间灯光这一变量影响显著且回归系数为正数。越明亮的网格是恐袭风险越易感的地区，即经济活动比较集中的地区更难避免恐袭。这一发现和国家层次上的GDP变量的影响呈现有趣的对照。人均GDP对“免疫能力”的影响为正，相对富裕的国家对恐袭的易感性低，但夜间灯光变量的影响则表明相对富裕的地区恐袭风险易感性高。这两个变量实际反映的是恐袭风险在宏观和微观层次上机制的差异。人均GDP较高的国家，政府能动用更多的资源增加抵御恐袭的能力，而且社会中的“怨恨”群体也较少，使用极端暴力的恐怖袭击的风险就会降低。[①] 在宏观的国家层次，恐袭风险感染主要遵从“防御方”的免疫力机制，而在微观的网格单元层次，恐袭风险变动的机制主要是“进攻方”即该网格对恐袭的吸引力。

此外，模型还发现，国家层次上的政治动荡（如内战、非正常的领导人更替以及本国曾经历的恐怖袭击总数）会破坏网格单元的恐袭“免疫力”。国家的政体也与网格处于“疫区”的概率相关——非洲的所谓民主国家对恐袭风险的抵抗力相对更弱。但是，国家层次上的人口规模和军费开支并未发现具有任何显著的影响。网格层次的解释变量如山区地形、族群因素[②]以及政府暴力[③]都显著地增大了“感染”恐怖袭击的风险。如图-6下图所示，除了石油变量[④]之外，网格层次上的变量几乎在95% 置信区间都是显著的。在宏观的国家和国际层次上并非所有入选变量均具有可靠影响，显示各地理点上的恐袭“免疫能力”虽然受到宏观和微观因素的共同作用，但微观因素能够提供更直接的解释。

---

① Jennifer Kavanagh，“Selection，Availability，and Opportunity：The Conditional Effect of Poverty on Terrorist Group Participation”，*Journal of Conflict Resolution*，Vol. 55，No. 1，2011，pp. 106-132.

② Julian Wucherpfennig，et al.，“Politically Relevant Ethnic Groups Across Space and Time：Introducing the GeoEPR Dataset”，pp. 423-437.

③ Ralph Sundberg and Erik Melander，“Introducing the UCDP Georeferenced Event Dataset”，pp. 523-532.

④ Päivi Lujala，Jan Ketil Rod，and Nadja Thieme，“Fighting over Oil：Introducing a New Dataset”，*Conflict Management and Peace Science*，Vol. 24，No. 3，2007，pp. 239-256.

2. 影响恐袭时机的因素

图-6 上图展示了对持续期方程估计的回归系数。持续期方程使用的是加速失效时间（accelerated failure time），回归系数是距离最终遭遇恐怖袭击时间长度的对数（log of the expected time to failure）。因此，回归系数为正数表示在保持其他因素不变的情况下，该自变量取值的增大可以推迟网格遭遇恐袭，而为负数的回归系数表明恐袭的危险迫近，也即危险率在各个时点上升。例如，人均 GDP 对数的系数是 0.1534，并且在 95％的置信水平上是显著的。该估计结果的解读为，其他条件不变的情况下人均 GDP 的对数每增加一个单位，网格距遭遇袭击的时间长度会增加 16.6％。从图-6 上图进一步发现，一些影响网格遭遇袭击风险的因素并不一定影响网格遭遇袭击的时机。例如，本国领导人的非正常更替或者网格中山地面积的比例会显著增加袭击的风险，但没有明显的证据表明会影响袭击的时机。

鉴于篇幅限制，我们重点选择四个时空变量报告它们对恐袭时机的影响。图-7 报告了在其他条件不变的情况下，四个网格层次上的时空变量和经济地缘变量对危险率曲线变化的影响。每个图包含两条危险率曲线的信息，反映所考察变量的变化带来的危险率曲线的变化。每条曲线包含危险率估计量的期望值和 90％ 置信区间。①

图-7（a）显示了网格与最近邻国边界的距离为 10 千米（绿色）和 500 千米（红色）对危险率曲线的影响。第一个持续期，距离最近边界距离为 10 千米网格的危险率是距离为 500 千米网格的 4.3 倍，在第二个持续期为 3.7 倍左右，在第六个持续期大约为 2 倍，大约到第 9 个持续期后两者的危险率无显著差别。这表明，当安全持续较长状态后可以达到一个稳定的有效防止恐怖主义的状态，距离最近边界的距离不再重要。但是在稳定状态到来之前，这一空间因素具有显著影响，尤其在第二个持续期其影响达到高峰，当安全状态逐渐稳定后，距离的影响也逐渐减弱。

图-7（b）所显示的网格距离首都 100 千米和 1000 千米对袭击时机的影响。距离首都 100 公里的网格在第一个持续期内的危险率是距离 1000 公里的网格的 4 倍，在前面的四个持续期内的袭击危险率持续升高，在第四个持

① 本文用计算模拟的方法得到危险率曲线的置信区间。在每一次模拟中，我们将其他变量控制在其平均值，只允许感兴趣的变量从一个值变化到另一个值，进而比较它们取不同值时的危险率变化。进行 1000 次模拟后，我们获得模拟结果并计算出均值以及定位出 5％ 和 95％ 百分位的值形成 90％ 置信区间。

续期影响达到最大，而后逐年下降。图-7（c）比较了反映经济地缘情况的夜间灯光变量对危险率的影响，比较了取值为5（低夜光）和取值为30（高夜光）两种情况下危险率曲线的变化。非常有趣的发现是，对于低夜光的网格，几乎在所有的持续期内其遭受恐袭的危险大致小于0.01，即经济活动很少的地区的安全状态会持续很长，也即恐怖袭击对于这些地区而言是极为遥远的事件。但是，在第一持续期内，夜间灯光高的网格使得袭击的危险明显增加了5.6倍，并在第二持续期内达到最高。图-7（d）比较了邻近网格在前两年遭遇100次袭击和0次袭击对于网格危险率曲线的差异。[①] 当周边地区在过去两年中均处于安全状态时（即0次），图中的红色曲线非常接近安全状态，表明网格的安全持续期结束的概率非常小。而如果周边为恐袭多发地带的话（即100次），危险率则在第一个持续期飙升到0.15，第二个持续期继续走高到0.2，之后虽然有所下降，但直至网格本身的安全持续期长达26年，只要其周边为恐袭高发地带，其安全状况就无法进入真正的稳定，其持续26年的成功防恐反恐也没有取得决定性胜利，仍然具有大于0.06的危险率。

（四）恐怖袭击的样本外预测（2012—2013）

在描述和解释了恐怖袭击的风险和时机之后，本文进一步利用SPDM模型进行样本外（out-of-sample）预测，并利用GIS技术，将风险概率$\hat{\pi}_{it}$以及在未来遭遇袭击时机的预测结果映射到每一个时空网格中，以便直观观察恐袭风险和危险率的分布。

遵循大数据统计预测的标准程序，本文将样本分为两组：第一组为训练集数据（training data），用作建立预测模型和估计模型参数；第二组为测试集数据（testing data），即除因变量以外的其他自变量数据。测试集数据并不会参与模型的构建，因而是样本外预测。具体而言，训练集数据包含1994—2011年间时空网格的数据，用以估计模型的参数，然后将关于这些时空网格的预测量带入2012—2013年间的训练集数据中，计算出各个时空网格在2012—2013年的恐袭风险和时机。[②] 可视化预测结果可以如上文将疫区和免

① 在我们的数据中约有30个网格—年的取值超过100次。例如，在1999年的阿尔及利亚，与网格编号182525、181807、181808相邻的网格在前两年的袭击超过了170次。

② 一般使用机器学习方法，通过训练集数据对模型反复测试，找到最合适的阈值之后再用测试集数据进行预测，这一过程一般耗时较长。参见Robert A. Blair, Christopher Blattman, and Alexandra Hartman, "Predicting Local Violence: Evidence from a Panel Survey in Liberia", *Journal of Peace Research*, Vol. 54, No. 2, 2017, pp. 298-312; Michael Colaresi and Zuhaib Mahmood, "Do the Robot: Lessons from Machine Learning to Improve Conflict Forecasting", *Journal of Peace Research*, Vol. 54, No. 2, 2017, pp. 193-214。

疫区投射到地图上实际对应时空网格中，绘制成如图-5 一样的空间分布地图，从而提炼出具有政策价值的建议。鉴于具体思路与上文一样，在此不再赘述和解读样本外预测结果。

对于样本外预测任务而言，通常还包含对模型预测表现的评估。对于以两分变量为因变量的模型，冲突研究中一般利用受试者工作特征曲线（ROC 曲线）和 PR 曲线（Precision-Recall Curve）来比较模型在样本内和样本外预测的表现。[①] 在曲线下的面积（AUC）越大，表示预测的准确度越高。图-8 展示了 ROC 和 PR 曲线在训练集和测试集中的预测表现，可以看到利用 1994—2011 年的数据训练的模型可以在测试集中达到很高的准确度。图-8（a)表明，在 ROC 曲线下两者的面积非常接近。同样，在图-8（b）中，在 PR 曲线下的面积也非常相似。这表明，模型很好地实现了样本内和样本外预测的平衡，没有以过分拟合来提高样本内预测的精度，从而呈现出虚假的准确预测。

对于反恐政策制定者而言，在低层次上的细粒度单元和在地理定位上的精确度，将极大地提高恐怖袭击风险评估和预估的针对性，缩小重点关注的范围，可以在排除大量具有高免疫力的地区之后，对反恐防恐的重点地区集中资源和制订更具针对性的措施。否则，正如伊桑·德·梅斯基塔所指出的，政府大规模和极端的反恐政策有可能产生负外部性，从而更有利于恐怖组织动员潜在的参与者加入恐怖组织。[②] 本文使用 SPDM 模型引入对疫区和免疫区的区分，有利于政府有的放矢地将资源从“免疫区”转移到“热点区”，同时避免产生大规模全国性的负外部性。在此意义上，用时空网格细粒分析单元上的恐怖袭击预测比国家层次上的恐怖预测具有更强的政策指导意义。

---

① Brian Greenhill, Michael D. Ward, and Audrey Sacks, “The Separation Plot: A New Visual Method for Evaluating the Fit of Binary Models”, *American Journal of Political Science*, Vol. 55, No. 4, 2011, pp. 991-1002.

② Ethan Bueno De Mesquita, “The Quality of Terror”, *American Journal of Political Science*, Vol. 49, No. 3, 2005, pp. 515-530.

六、结　　论

本文是关于恐怖主义袭击的大数据研究的创新性尝试，将恐怖袭击事件通过GIS技术和空间层次（地理网格）数据进行合并，以非洲为案例，描述、解释和预测恐怖袭击的时空关联和动态变化，展示了大数据细粒度恐怖主义研究的优势和前景。本文的研究目的不仅在于推动大数据时代恐怖主义研究的科学化，也在更普遍的意义上丰富了大数据国际安全研究的思路、方法和路径。

首先，地理网格化是大数据时代国际安全研究细粒化的重要趋势，本文所使用的时空网格划分方法是一种灵活的方法，研究者可以根据其研究主题、目的和偏好，选择更大或更小的网格划分。GIS技术的发展使得研究者在未来可以充分开发不同空间分辨率的地理信息数据，进而在不同层次对感兴趣的政治现象进行研究。第二，研究者可以根据时空网格这一特殊的数据，进一步发展新的方法去测量不同层次的国际安全结果变量。SDPM也为研究者分析研究对象并非同质性分布的课题提供了统计分析方法。值得说明的是，本文基于数据的可获得性选择了以年作为时间维度，未来的研究可以充分利用大数据事件数据，更加灵活地将时间维度细粒化到以季度、月度甚至天为单位。时空网格完全适合更加精细的时间维度分析。第三，本文的分析和预测虽然是一个局部的尝试，但建立恐怖主义预警系统以及更为综合的国际安全预警系统，是科学研究的创新性任务，也是大数据安全研究在应用领域的重要使命。[①] 大型预警系统的成功建立往往以前期局部的学术研究为重要起点和基础。本文希望未来能够出现更多此类尝试，为我国建立和完善国际安全预警系统奠定更为坚实的学术基础。

① Håvard Hegre, et al., "ViEWS: A Political Violence Early-Warning System", *Journal of Peace Research*, Vol. 56, No. 2, 2019, pp. 155-174.

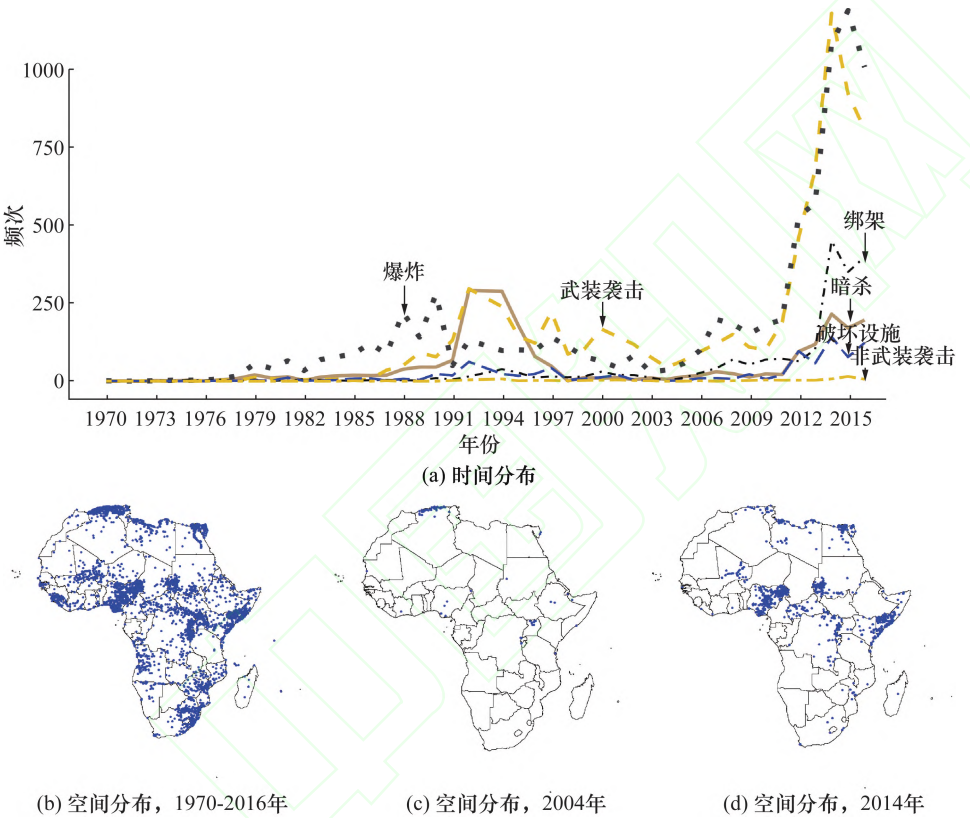


图-1　非洲恐怖主义袭击的时空分布（1970—2016）

注：图-1（a）展示了非洲1970—2016年间主要恐怖主义类型的时间分布。图-1（b）展示了1970—2016年非洲恐怖主义的空间分布。图-1（c-d）分别展示了2004年和2014年的恐怖袭击分布，圆点表示恐怖袭击的具体位置。

图片来源：笔者根据全球恐怖主义数据库（GTD）利用R软件绘制。非洲的“矢量文件”（shapefile）来源于R中的CShapes软件包，彩色版地图在作者的Harvard Dataverse可以获得，以下同。

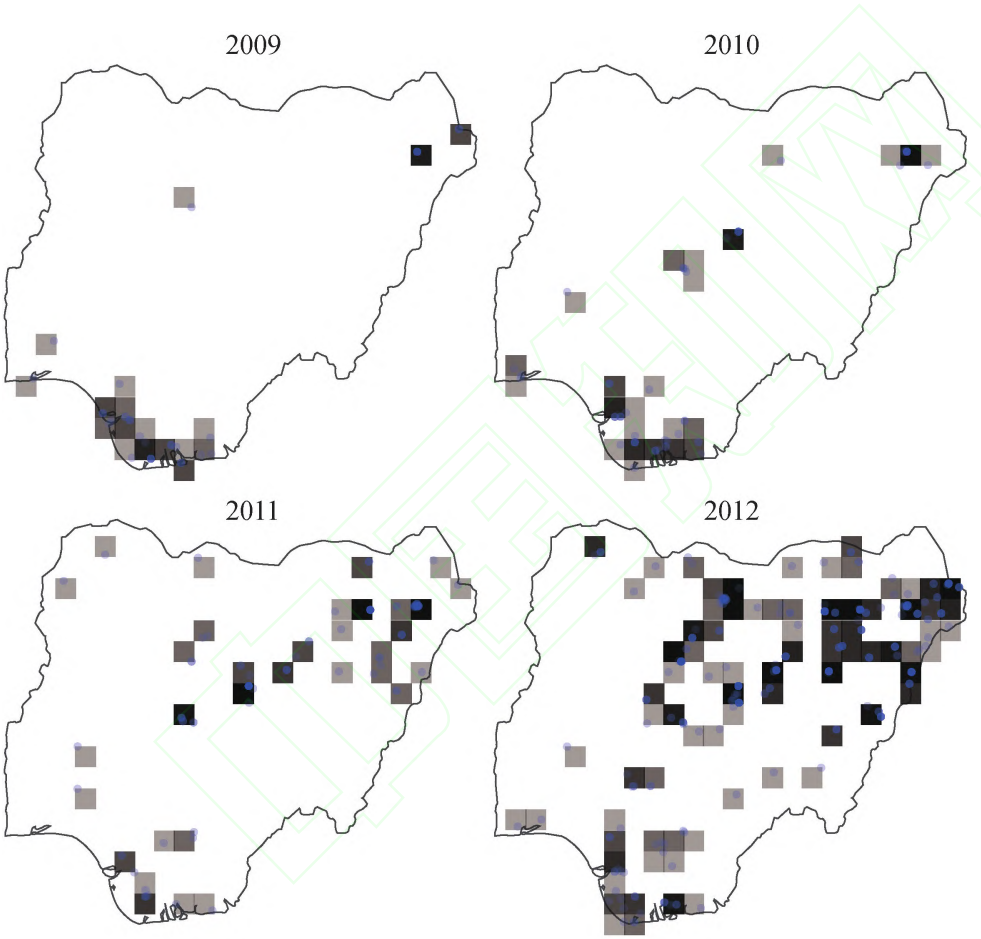


图-2　尼日利亚恐怖主义袭击的网格（2009—2012）

注：图-2 显示了尼日利亚 2009—2012 年间发生恐怖袭击的网格。图中每一个方块代表一个时空网格。为画面尽量简洁，没有发生恐怖袭击网格为无色且隐去轮廓。方格颜色越深表示当年该时刻网格遭遇袭击的次数越多。图中蓝色圆点代表恐怖袭击在网格内部的具体位置。

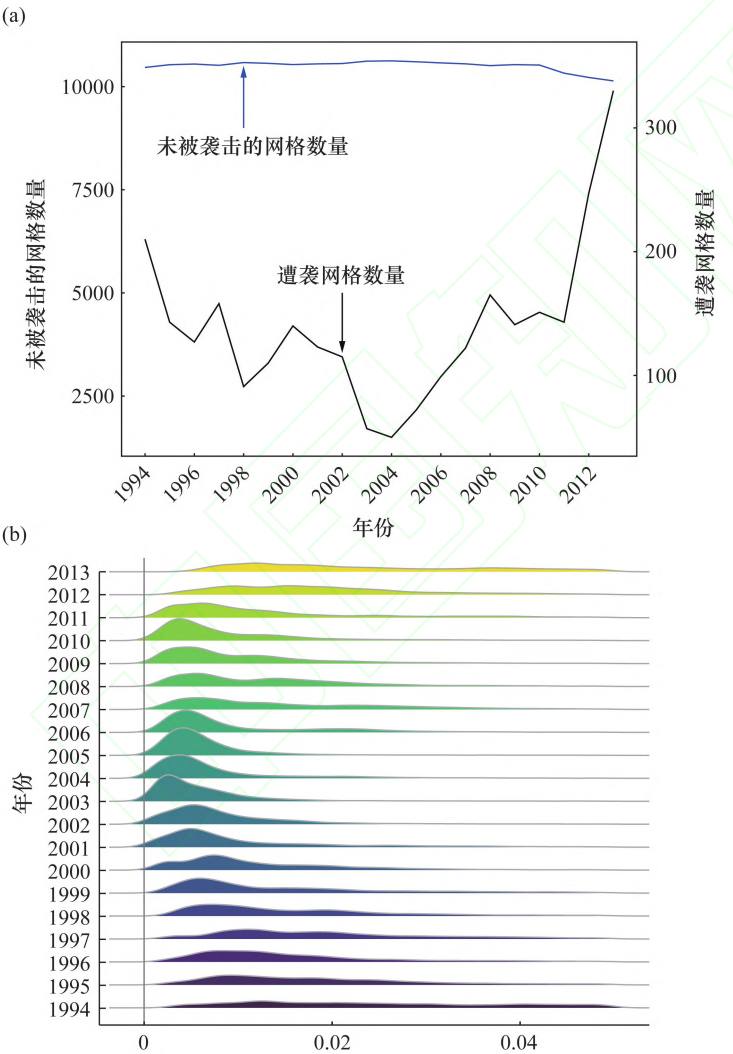


图-3　遭遇恐袭的实际网格数量与估计的免疫区密度分布（时间维度）

注：图-3（a）展示 1994—2013 年非洲经历恐袭的网格数量（右侧 Y 轴）和“尚未”经历恐袭的网格数量（左侧 Y 轴）对比。图-3（b）展示了根据图-6 模型所估计的免疫区概率分布。“免疫区”指模型所估计的袭击概率$\hat{\pi}$低于 0.05 的网格。

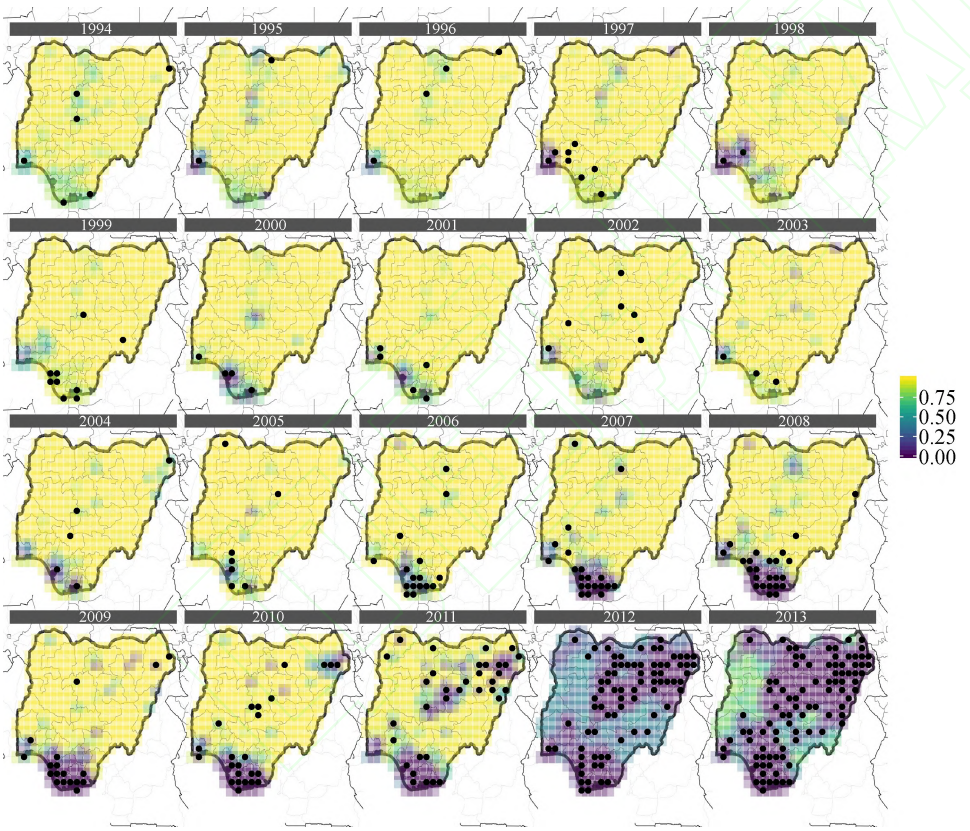
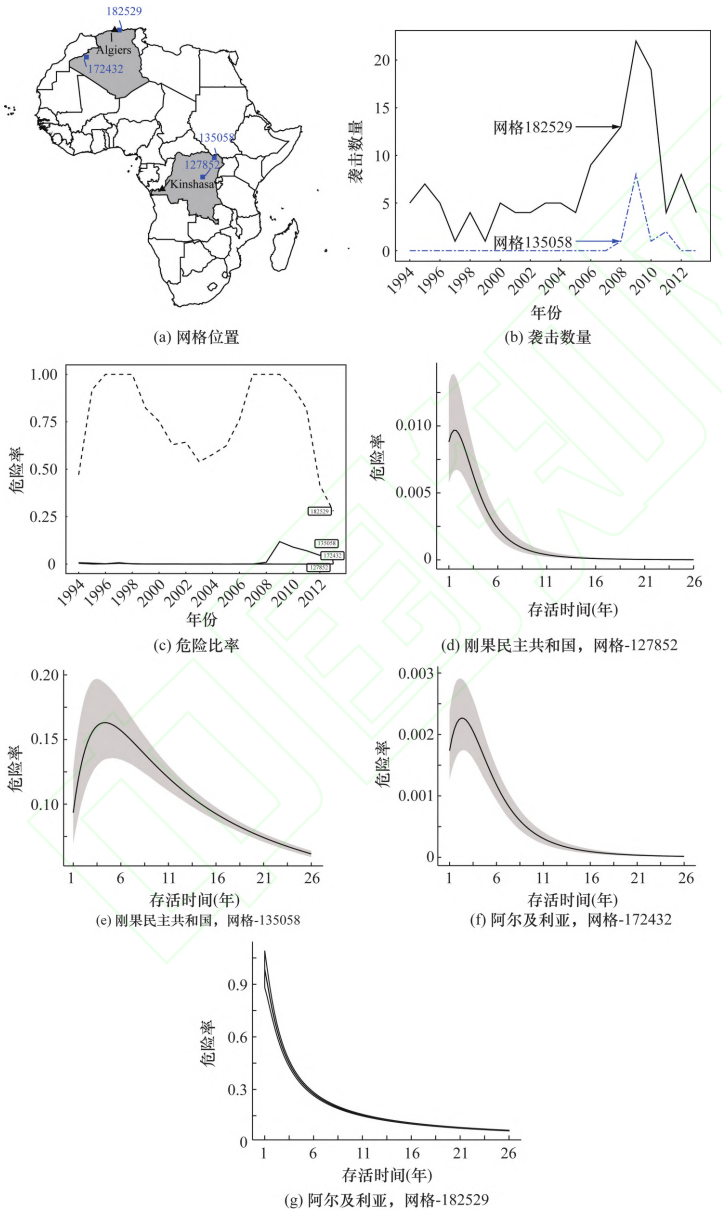


图-4　尼日利亚免疫区分布（1994—2013）

注：图-4 展示了根据图-6 估计的尼日利亚境内 1994—2013 年网格的免疫区（$1-\hat{\pi}_{it}$）在时空上的变化。黑色圆点表示实际遭遇恐怖袭击的网格。黄绿色（$1-\hat{\pi}_{it}$ 值越大）表示处于免疫区的概率越大，深紫色（$1-\hat{\pi}_{it}$ 值越小）表示越有可能划分为疫区。



(a) 网格位置　(b) 袭击数量

(c) 危险比率　(d) 刚果民主共和国，网格-127852

(e) 刚果民主共和国，网格-135058　(f) 阿尔及利亚，网格-172432

(g) 阿尔及利亚，网格-182529

图-5　时空网格层次上的恐袭时机

注：图-5（a）展示从刚果民主共和国和阿尔及利亚分别随机选取的 2 个不同网格的位置（方格）以及两国首都（三角形）。图-5（b-c）分别展示了网格-135058 和网格-182529 当年遭遇恐怖袭击的总数以及估计的危险比率。图-5（d-g）分别展示了四个网格的危险率曲线，阴影部分表示 90% 置信区间。

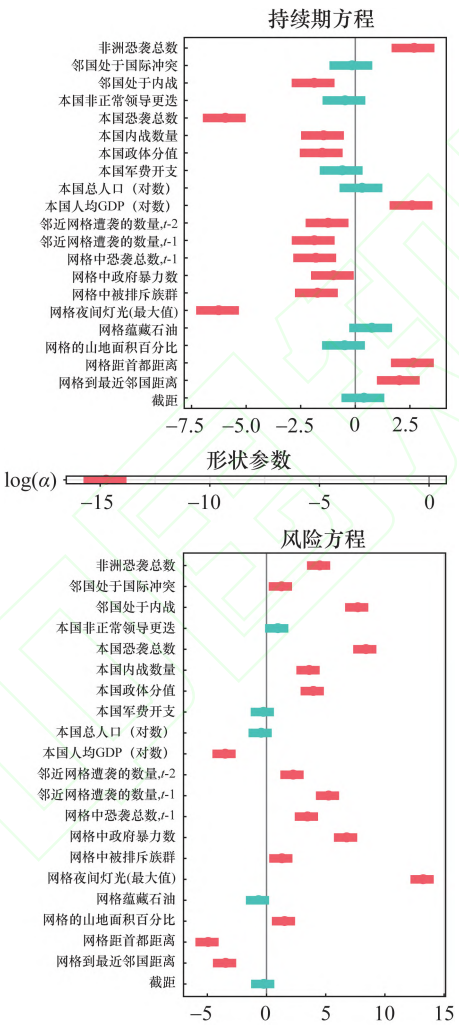


图-6　SPDM 回归系数：解释恐怖袭击的时机和地点

注：图-6 展示估计的回归系数（圆点）及其 95% 置信区间，红色点表示在 95% 置信区间显著性，而绿色点表示在 95% 置信区间不显著。模型采用的是 Log-logistic 分布，回归系数进行了标准化处理（rescale）。两个方程的分析层次都是网格—年。

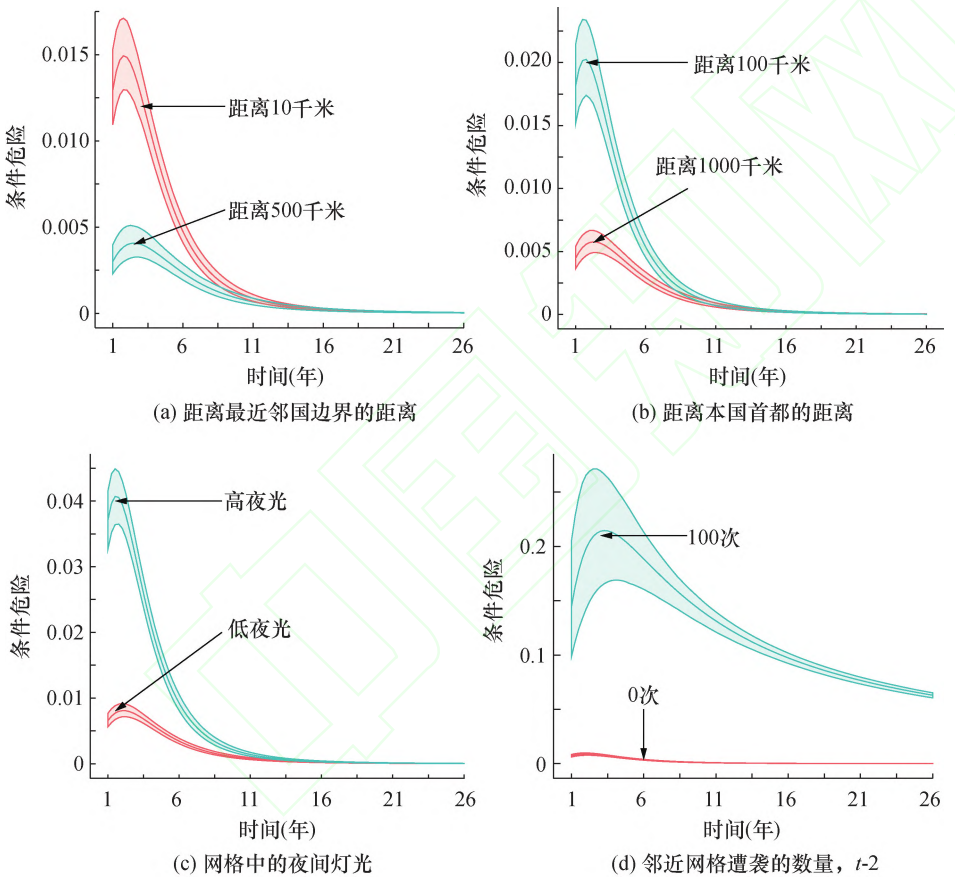


图-7　变量对危险率的影响：恐袭时机

注：图-7 展示了根据图-6 模型所估计的结果，通过 1000 次模拟计算出危险率的变化。阴影部分表示 90% 置信区间。

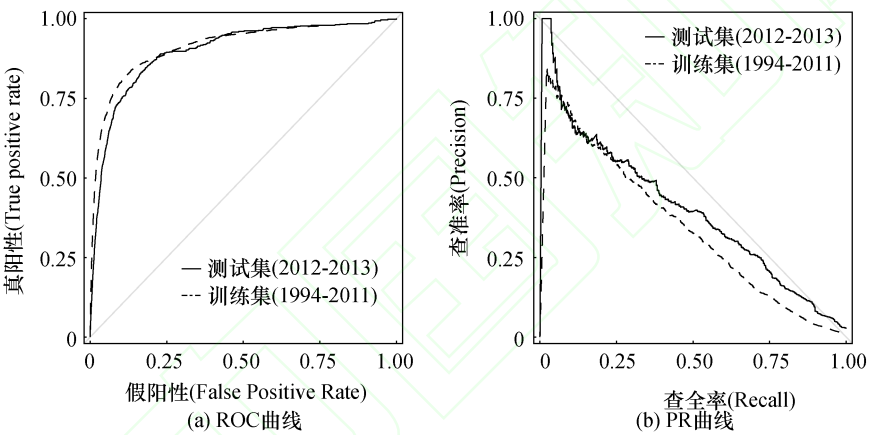


图-8　基于训练集数据和测试集数据的 ROC 与 PR 曲线

注：图-8 展示基于训练集数据（1994—2011）的样本内表现和基于测试集数据（2012—2013）的样本外预测表现。

（责任编辑：李　丹）