生物多样性数据GAP分析
Biodiversity Data Gap Analysis (DGA)

林聪田

2016-10-11

第十二届生物多样性科学与保护研讨会

DGA目的

- 现有分布数据虽然不少，但在类群上和地理上都有缺失（GAP）
- 需要方法和工具来衡量GAP，然后进一步判断数据的适用性
- 为新的生物多样调查，物种采样提供指导

DGA对象

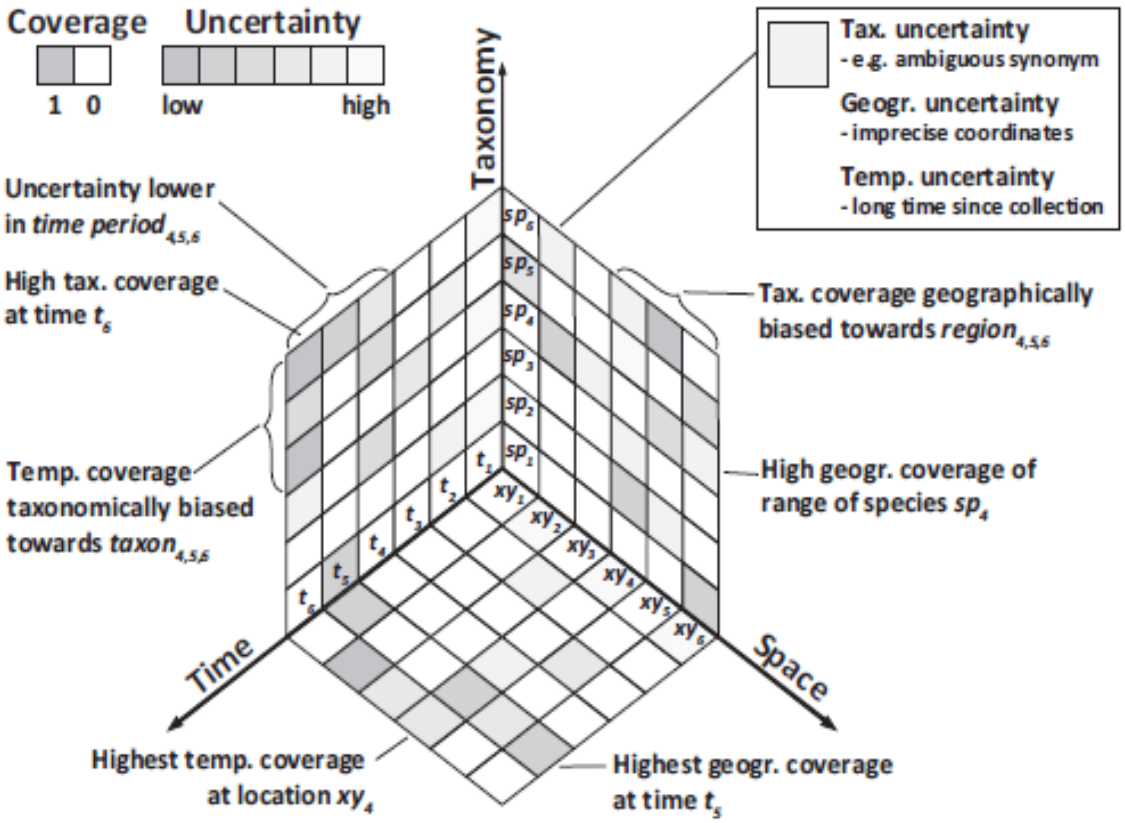
- 分布数据
- 类群 & 物种
- 维度：

  Taxonomic

  Geographic

  Temporal

  Environmental



Meyer et al., 2016

DGA技术与工具

- 数据库技术(Database engines)
- 地理信息系统工具（GIS and mapping tools ）
- 统计工具（Statistical packages or languages (e.g. *R*) ）
- 数据服务接口技术（Web services）
- 可视化工具（Visualization tools ）

DGA scope

taxonomic / geographic / temporal

Data accessibility assessment

1. Data sources inventory
2. Data scope analysis
3. Data amount
4. How to access

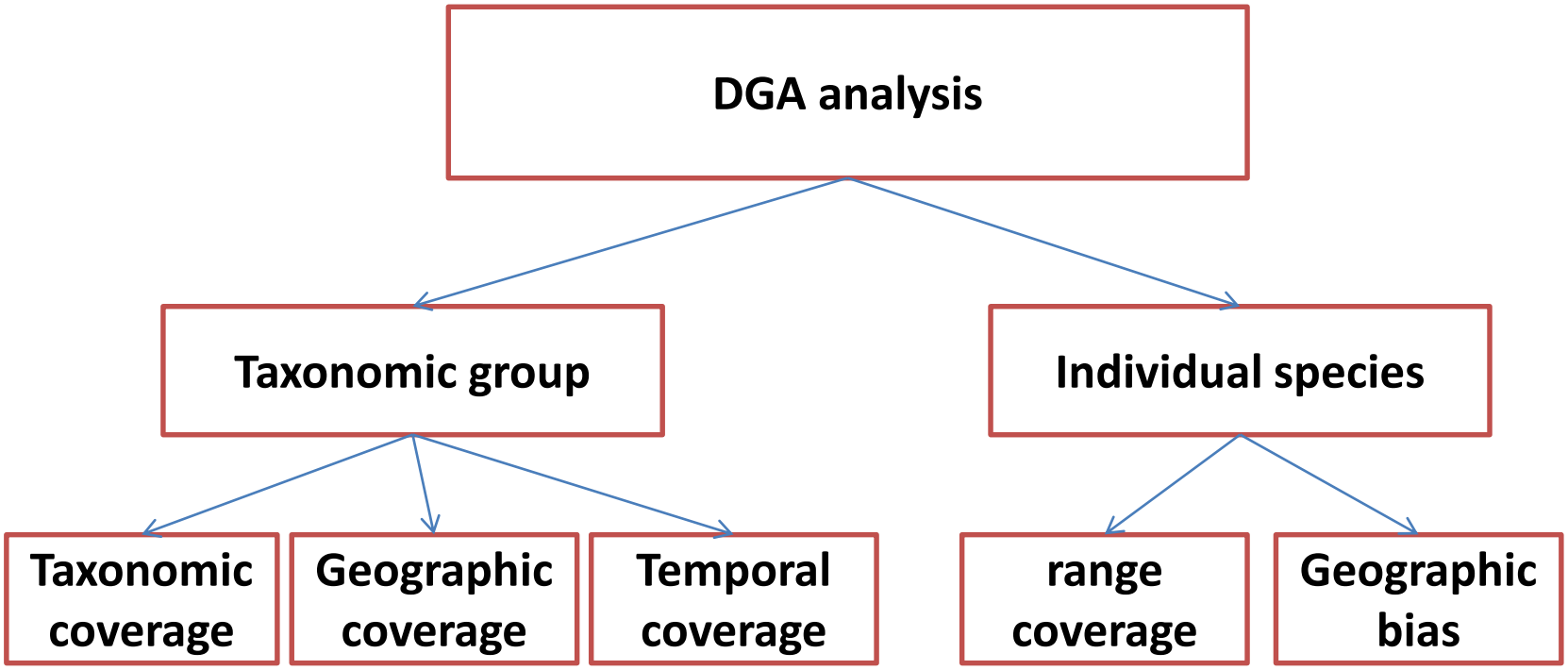
Data harvest

Data clean

1. Clean data without coordination
2. Clean data with wrong coordination
3. Clean data unmatched DGA scope

Data mapping

Mapping data to grid and shape map


DGA analysis
Taxonomic group
Individual species
Taxonomic coverage
Geographic coverage
Temporal coverage
range coverage
Geographic bias

DGA 分析实例

GBIF的中国鸟类分布数据

DGA scope

taxonomic: 鸟类
geographic: 中国
temporal: 全部采样年份

Data accessibility assessment

1. Data sources: GBIF
2. Data amount: **389,456**
3. How to access: Web API

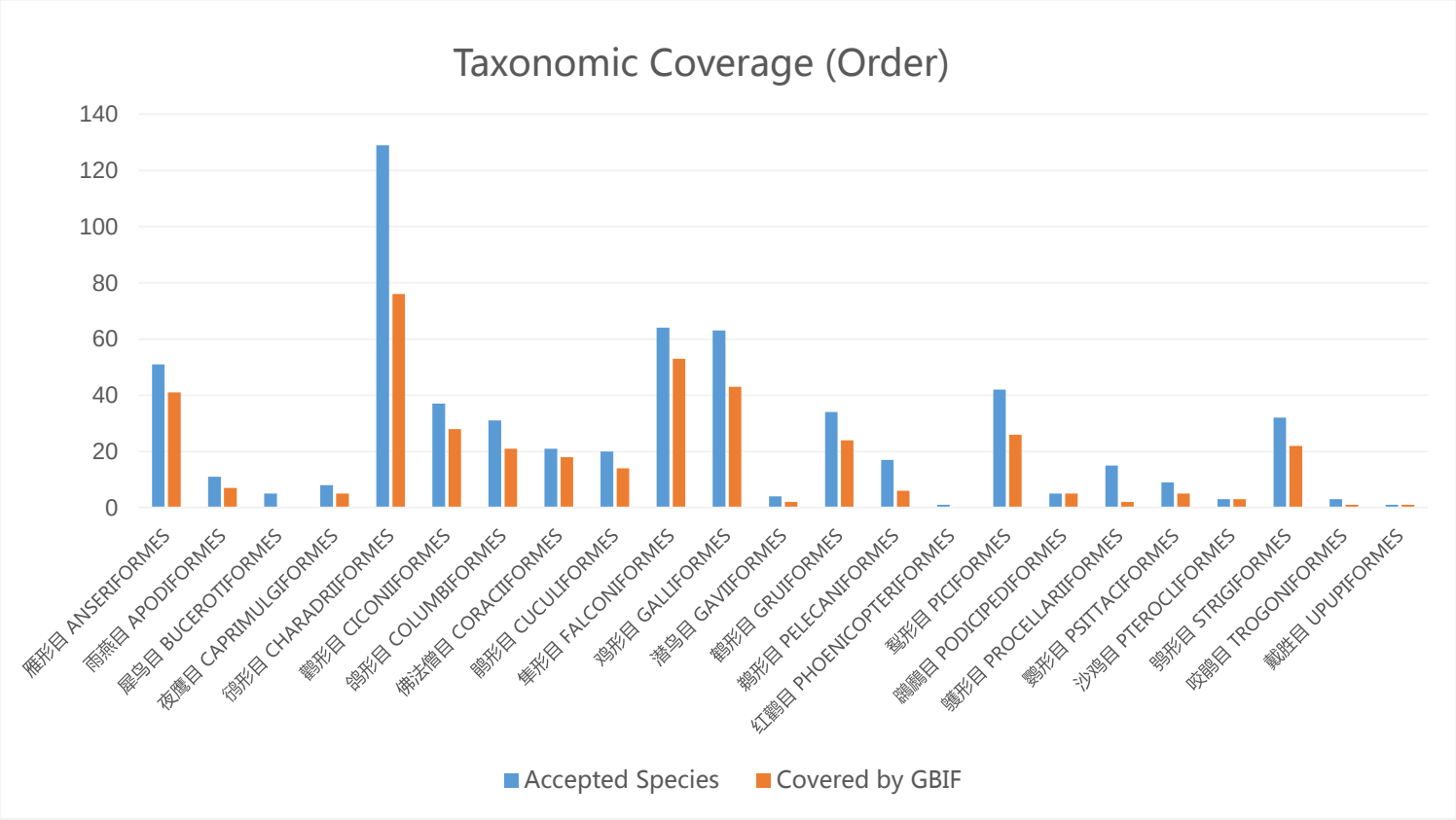
Data harvest

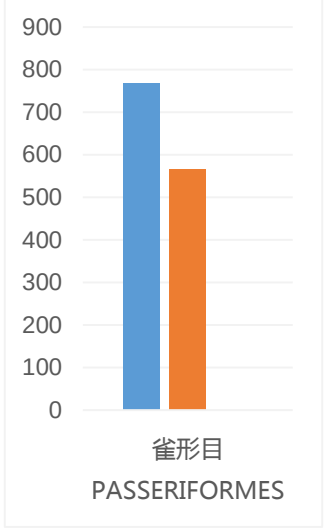
Data clean

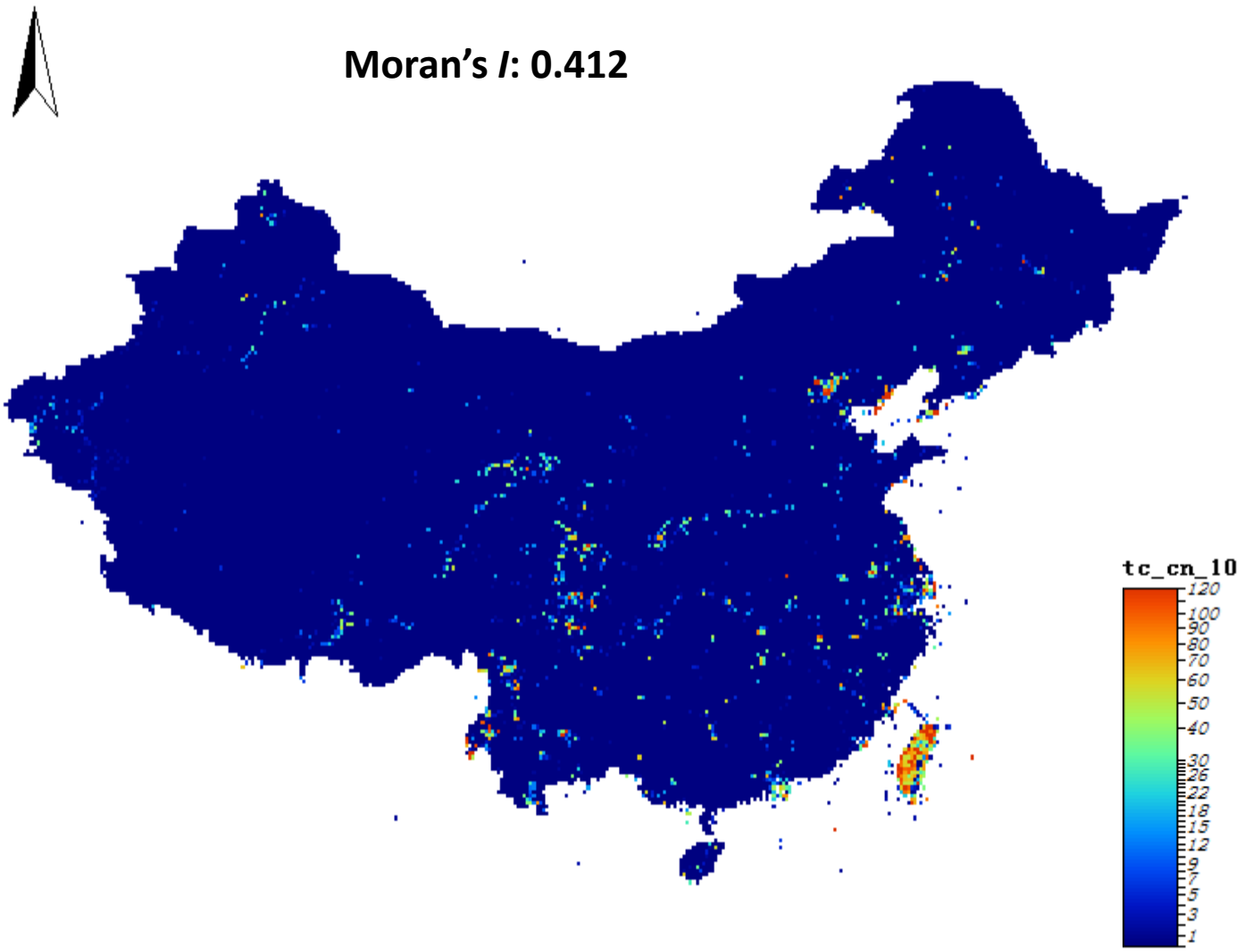
1. without coordination: -78,841
2. unmatched DGA scope: -503

中国鸟类分布记录 (GBIF)

Taxonomic coverage

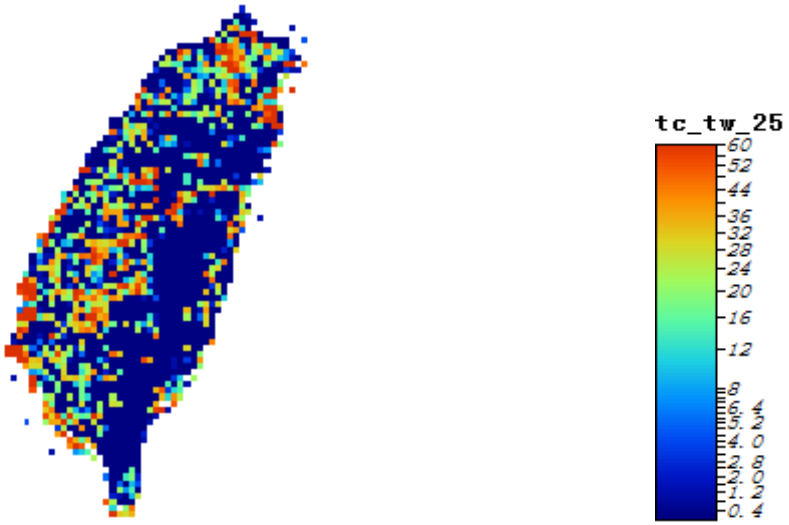






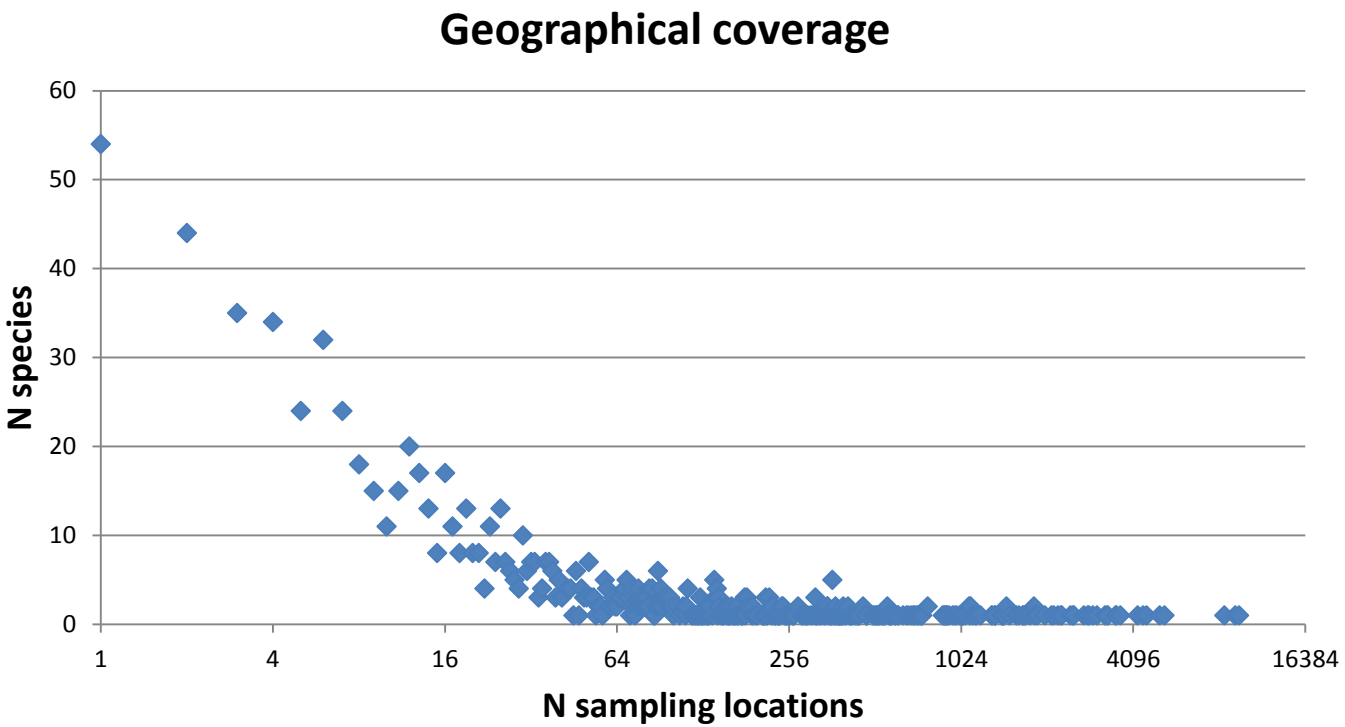
每个栅格被采样的物种数 20km X 20km

Moran's *I*: 0.298

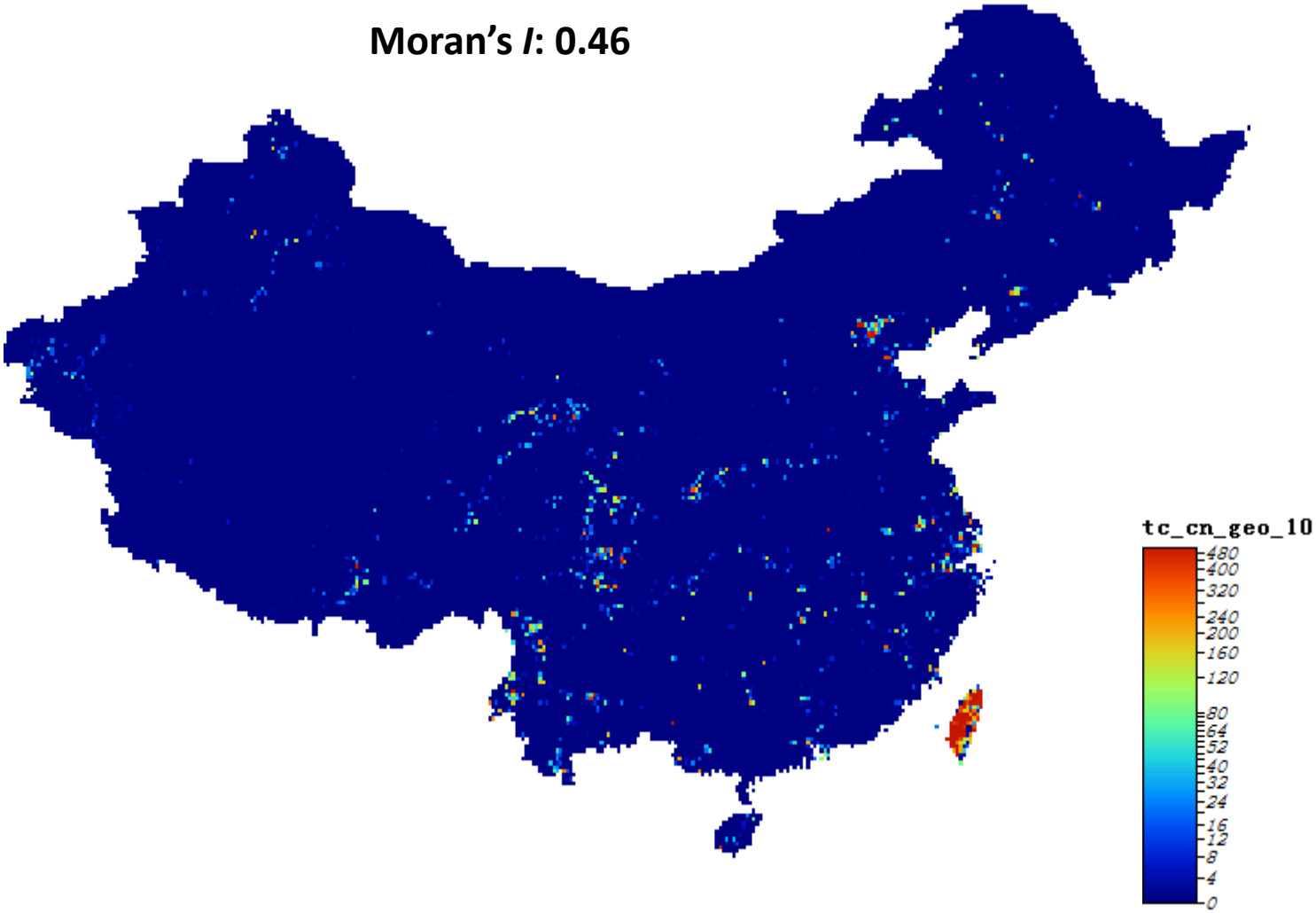


Species coverage in *5km X 5km* grids of Taiwan province

Geographical coverage

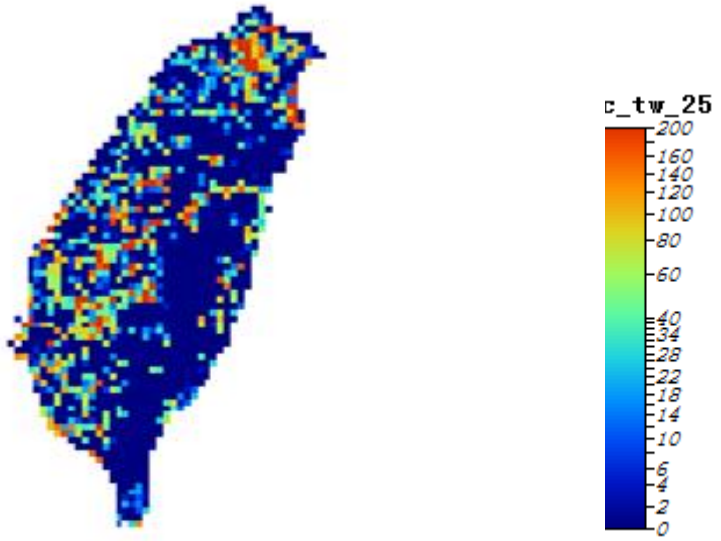


Moran's *I*: 0.46



每个栅格被采样的次数 **20km X 20km**

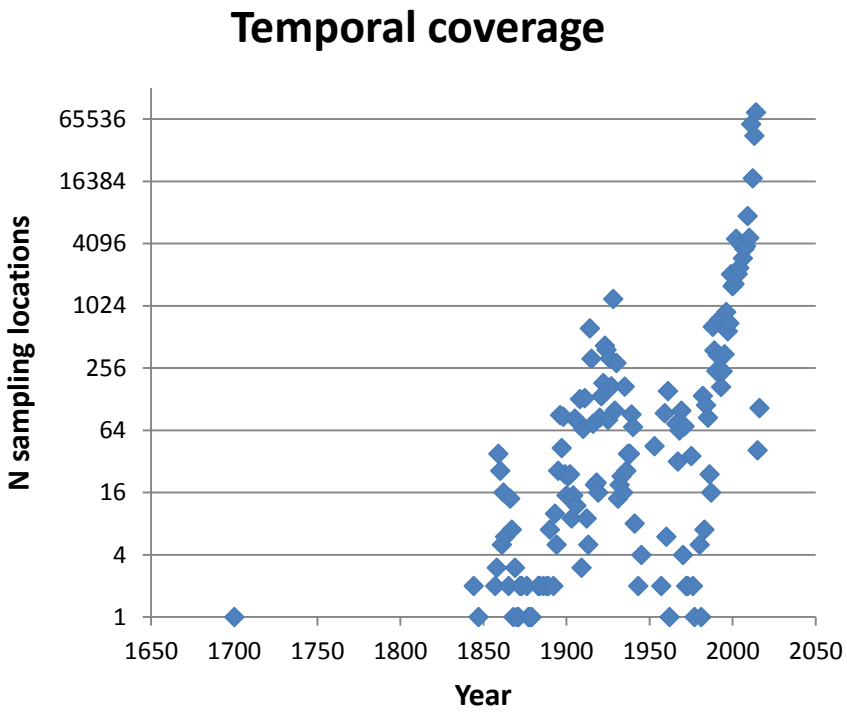
Moran's *I*: 0.234

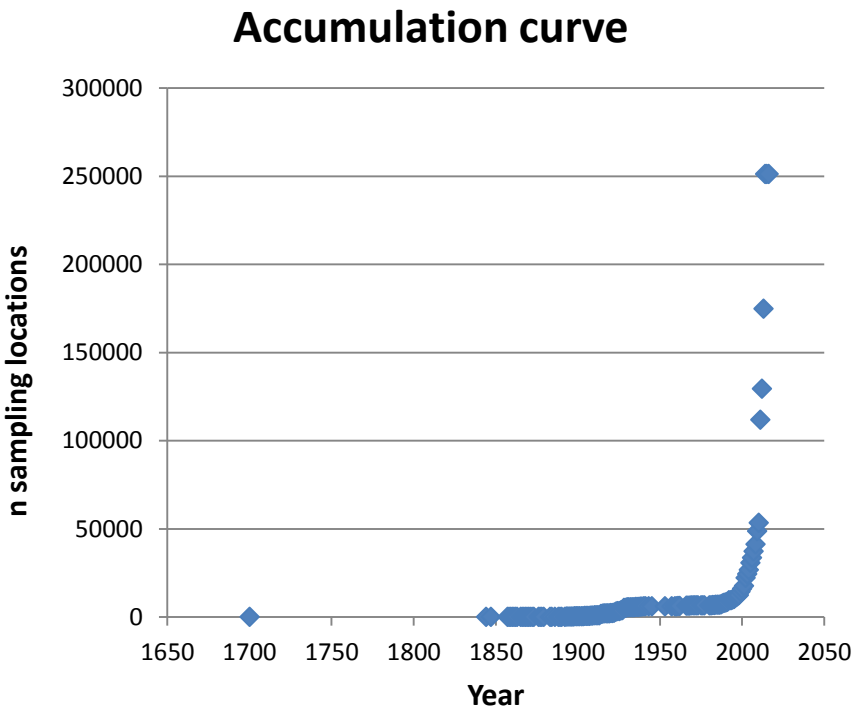


Geo-coverage of Birds sampling location in Taiwan province

5km X 5km

Temporal coverage





台湾画眉（*Garrulax taewanus* (Swinhoe, R.,1859)）分布数据GAP评估

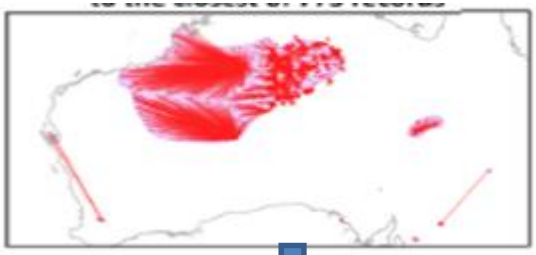
分布范围内随机产生1000个点

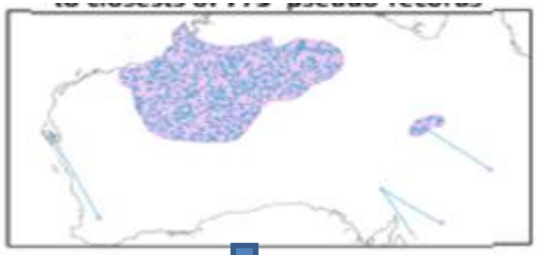
Null model（1000次重复）：
随机产生与实际点数相同的伪样本

实际分布点

计算随机点到实际点的最短距离

计算随机的1000个点到伪样本点的最短距离





Range coverage（RC）：随机点到实际点最小距离的平均值

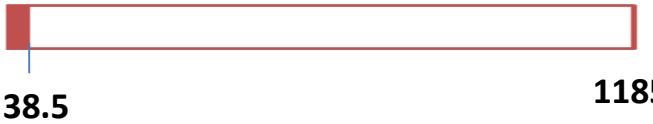
Geographical bias：实际RC 值与原假设模型RC值的差/原假设模型RC的标准差

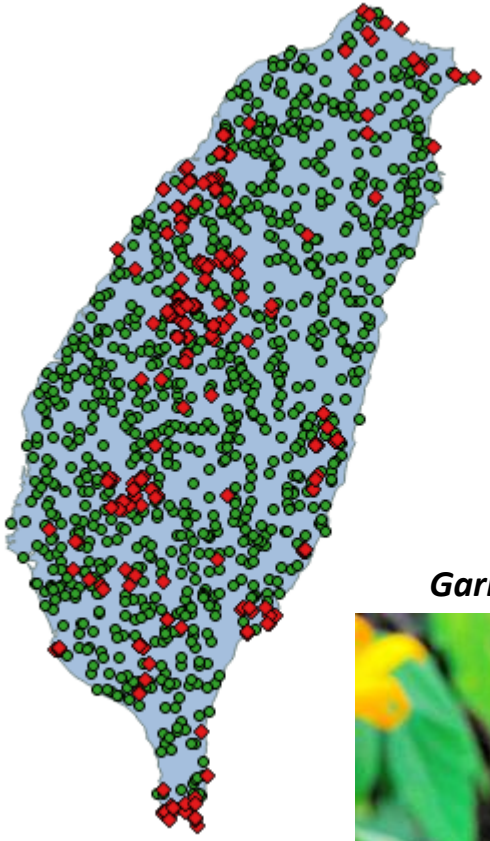
$$range\ coverage = -\mathrm{MMD} = -\frac{1}{1000}\sum_{i=1}^{1000}\mathrm{MinDistRP}_i$$

$$geographical\ bias = \frac{\mathrm{MMD}_{observed} - \mathrm{mean}(\mathrm{MMD}_{null\ model})}{\mathrm{SD}(\mathrm{MMD}_{null\ model})}.$$

Record count: 530
Duplicated: 269
Range area: 36192.8155 km^2
Range coverage： -13.8 km

Geographical bias: 38.5





台湾画眉分布点

1000个随机点

台湾画眉

***Garrulax taewanus* (Swinhoe, R.,1859)**

总结

- GBIF的中国鸟类数据在类群覆盖度方面比较好（>70%）
- 根据Moran's *I* 指数，采样的物种数在地理上相对集中，被采样的地点相对集中
- 多数样本集中在少数几个种
- 重复样本比较多
- 样本记录的时间多在2000年后，不利于时间序列分析

- 台湾省的鸟类数据分布相对均匀，地理覆盖度比较广
- 从覆盖度和地理偏差指标看，台湾画眉的地理分布数据覆盖范围大，偏差较小，补充少量样本就可以用于其分布地预测。
- 地理覆盖度和偏差指数大小对分布地模型预测的影响有待进一步研究。

下一步工作

- 开发新的GAP分析方法
- 评估不同指标对生物多样性评价及物种分布预测的影响
- 完善DGA流程，并上线服务

- Amano T, Lamming JDL & Sutherland WJ (2016) Spatial Gaps in Global Biodiversity Information and the Role of Citizen Science. BioScience 66(5): 393-400. doi:10.1093/biosci/biw022
- JC Ganglo & SB Kakpo (2016) Completeness of Digital Accessible Knowledge of Plants of Benin and Priorities for Future Inventory and Data Discovery. Biodiversity Informatics 11: 23-39. doi:10.17161/bi.v11i0.5053
- Mair L & Ruete A (2016) Explaining Spatial Variation in the Recording Effort of Citizen Science Data across Multiple Taxa. PLoS ONE 11(1): e0147796. doi:10.1371/journal.pone.0147796
- Meyer C, Weigelt P & Kreft H (2016) Multidimensional biases, gaps and uncertainties in global plant occurrence information. Ecology Letters. doi:10.1111/ele.12624
- Meyer C, Jetz W, Guralnick RP, Fritz SA, Kreft H (2016), Range geometry and socio-economics dominate species-level biases in occurrence information. Global Ecology and Biogeography. doi:10.1111/geb.12483
- Orlikowska EH, Roberge JM, Blicharska M & Mikusiński G (2016) Gaps in ecological research on the world's largest internationally coordinated network of protected areas: A review of Natura 2000. Biological Conservation 200: 216-227. doi:10.1016/j.biocon.2016.06.015
- Stropp J, Ladle RJ, Malhado ACM, Hortal J, Gaffuri J, Temperley WH, Olav Skøien J & Mayaux P (2016), Mapping ignorance: 300 years of collecting flowering plants in Africa. Global Ecology and Biogeography. doi:10.1111/geb.12468
- Troia MJ & McManamay RA (2016) Filling in the GAPS: evaluating completeness and coverage of open-access biodiversity databases in the United States. Ecology and Evolution. doi:10.1002/ece3.2225
- Boakes EH, Fuller RA, McGowan PJK & Mace GM (2015) Uncertainty in identifying local extinctions: the distribution of missing data and its effects on biodiversity measures. Biology Letters 12(3): 20150824. doi:10.1098/rsbl.2015.0824
- Meyer, C, Kreft H, Guralnick R & Jetz W (2015) Global priorities for an effective information basis of biodiversity distributions. Nature Communications 6: 8221. doi:10.1038/ncomms9221
- Rubio Teso ML, Ronquillo Ferrero C, Nebreda Trejo A, Parra Quijano M, Torres Lamas E & Iriondo Alegría JM (2015) In situ conservation of CWR in Spain: present and future. Crop Wild Relative 10: 24-26. http://www.pgrsecure.bham.ac.uk/sites/default/files/documents/newsletters/CWR_Issue_10.pdf
- Pino-Del-Carpio A, Ariño, AH, Villarroya A, Puig J & Miranda R (2014) The biodiversity data knowledge gap: Assessing information loss in the management of Biosphere Reserves. Biological Conservation 173: 74-79. doi:10.1016/j.biocon.2013.11.020

- Wetzel F, Hoffmann A, Kroupa A et al. (2014) EU BON Deliverable 1.1: Gap analysis and priorities for filling identified gaps in data coverage and quality. Berlin: Museum für Naturkunde and Leibniz Institute for Evolution and Biodiversity Science. http://www.gbif.fr/sites/default/files/documents/eu_bon_deliverable_1_1_final_v30.pdf
- Gaiji S, Chavan VS, Ariño AS, Otegui J, Hobern D, Sood R & Robles E (2013) Content assessment of the primary biodiversity data published through GBIF network: Status, challenges and potential. Biodiversity Informatics 8: 94-172. doi:10.17161/bi.v8i2.4124
- Otegui J, Ariño AH, Encinas MA & Pando F (2013) Assessing the primary data hosted by the Spanish node of the Global Biodiversity Information Facility (GBIF). PLoS ONE 8(1): e55144. doi:10.1371/journal.pone.0055144
- Sousa-Baena MS, Couto Garcia L & Peterson AT (2013) Completeness of digital accessible knowledge of the plants of Brazil and priorities for survey and inventory. Diversity and Distributions 20(4): 369-381. doi:10.1111/ddi.12136
- Chavan VS, Gaiji S, Hahn A, Sood RK, Raymond M & King N (2010) State-of-the-Network 2010: Discovery and Publishing of Primary Biodiversity Data through the GBIF Network. Copenhagen: Global Biodiversity Information Facility http://www.gbif.org/resource/80666
- Hill AW, Otegui J, Ariño AH & Guralnick RP (2010) GBIF position paper on future directions and recommendations for enhancing fitness-for-use across the GBIF Network, Copenhagen: Global Biodiversity Information Facility. http://www.gbif.org/resource/80623
- Collen B, Ram M, Zamin T & McRae L (2008) The tropical biodiversity data gap: addressing disparity in global monitoring. Tropical Conservation Science 1(2): 75-88. http://tropicalconservationscience.mongabay.com/content/v1/08-06-09-Ben_Collen_et_al.html
- Peterson AT, Navarro-Sigüenza AG & Benítez-Díaz H (2008) The need for continued scientific collecting: A geographic analysis of Mexican bird specimens. Ibis 140(2): 288-294. doi:10.1111/j.1474-919X.1998.tb04391.x
- Guralnick RP, Hill A & Lane M (2007) Towards a collaborative, global infrastructure for biodiversity assessment. Ecology Letters 10(8): 663-72. doi:10.1111/j.1461-0248.2007.01063.x
- Condé S, Roekaerts M, Vignault MP & Richard D (1996) Databases on species, habitats and sites: Survey and analysis, 1995-96. Copenhagen: European Environment Agency. http://www.eea.europa.eu/publications/92-9167-034-0/at_download/file
- Soberón J, Llorente J & Benítez H (1996) An international view of national biological surveys. Annals of the Missouri Botanical Garden 83(4): 562–573. doi:10.2307/2399997

谢谢！
欢迎指正