



2024.5.28

EVERY AUTHOR

从理解到精通：科研论文核心研究方法全解读

RESEARCH METHODOLOGY 101

FREE WEBINAR

免費公益課程開放報名

国内专场

北京時間 2024年5月28日下午7:00-8:00 (GMT+8)



微信搜一搜

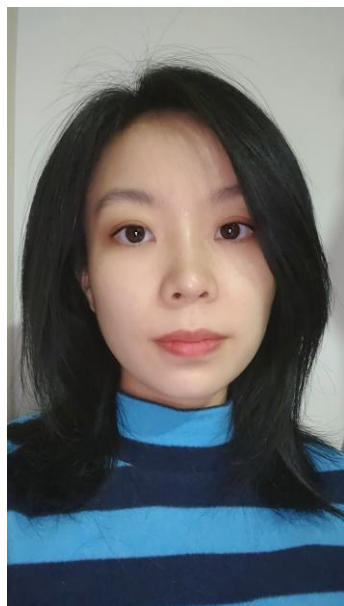
🔍 Every Author

www.everyauthorpub.com

此培训活动一般以普通話与英文进行。

Our training workshops are normally conducted in Mandarin Chinese and English.

www.everyauthorpub.com



主攻教学设计与评估, 研究方法, 英语语言学。

负责科技、工程、医学 方向, 包括该领域相关的学术论文写作的作者培训与相关课程的教学。

高校教师|资深编辑|国际期刊审稿人

Maisie Liao

Author Consultant

Every Author 作者服务





Every Author是一家成立于香港深受全球新人作者信赖和喜爱的学术出版服务与培训机构，为不同阶段的科研新人提供专业化的学术服务支持

高质量编辑团队

专业化国际师资

行业认证



Association of Learned
and Professional
Society Publishers

© Every Author





Editing Services 编辑服务

论文语言编辑润色

论文翻译

期刊选择

预审/同行评审服务等





国际期刊论文零基础出版全程指导

精品课程

EVERY AUTHOR
国际期刊论文
零基础出版指导

PUBLISHING GUIDE FOR THE BEGINNER (MANDARIN CHINESE)

限时优惠228元

国际期刊论文零基础出版指导
¥268.00-2500.00

长按识别小程序码进入

EVERY AUTHOR
Supporting brave
new authors

引导学术规范
和准则

提供反馈和修
订建议

提高发文成功率



From Understanding to Mastery

A Comprehensive Interpretation of
Core Research Methodologies in Academic Papers

Maisie Liao

2024. 05.28



Chapter A: Understanding Methods



The role of the methodology part

- **在创作上的意义**

如实展现你的科研流程

说明你的results是如何产生

- **对于论文发表的意义**

向编辑证明你的科研成果是合规，具有学术的严谨性

在peer review环节体现你的原创性，提供审核你成果的标准

- **对于未来的影响**

推动业内的学术交流

增加后期被专业人士引用的几率



Before you writing, please address research deficiencies

Defects and solutions

- Sampling Bias: 选择的样本不具代表性，无法反映目标人群的特征。

使用随机抽样技术确保样本的代表性，或者采用分层抽样方法来包括所有重要的亚群体。

Yin et al. (2024): “Pregnant women were eligible in the UNIHOPPE cohort if they attended regular antenatal care and delivered at the included hospitals” (研究父母年纪对分娩结果的影响)

- Measurement Errors: 使用不准确或不一致的工具和方法来收集数据。

确保所有测量工具都经过验证和校准，使用标准化的数据收集程序来减少误差

Chaudhuri (2024): In order to achieve reliable vehicle detection dataset classes are labelled. Based on motivation from Ref.5 an image tool has been used here towards labeling and annotation of this dataset. As shown in Fig. 3a,b for each object in image, a bounding box is assigned manually. (为了收集车辆在道路行驶上的数据)



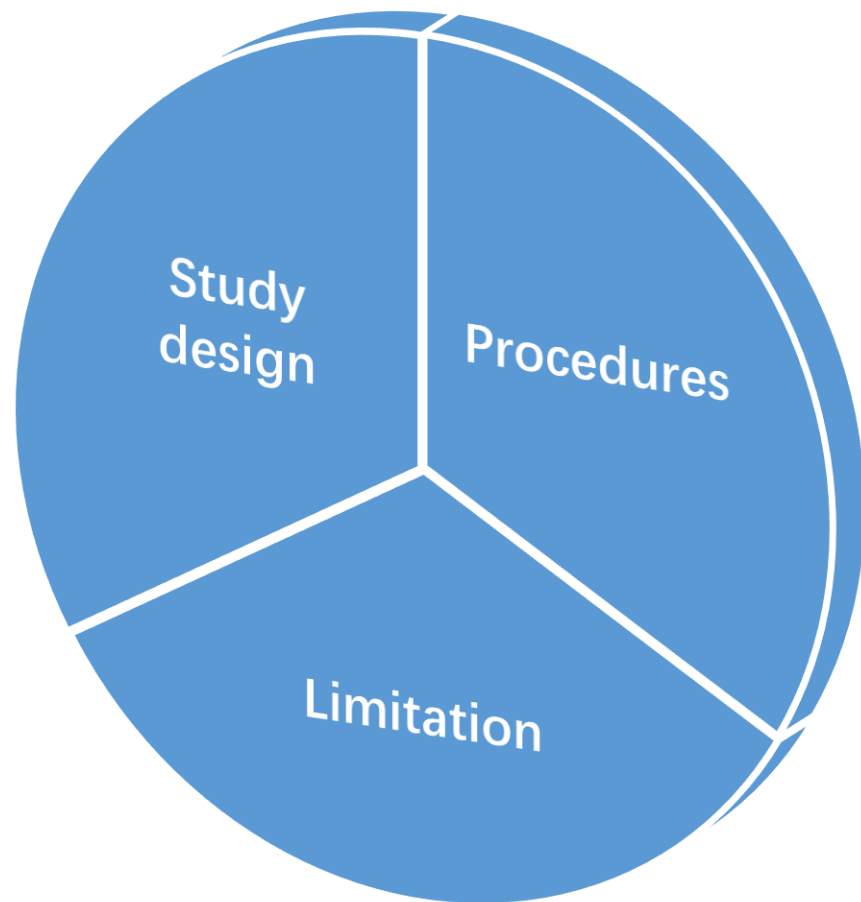
Main Components

内容:

- 哪些内容是硬性要求, 哪些内容我可以根据自己情况灵活创作的
- 哪些materials, participants和instruments的相关信息应该展示在论文中

结构:

- 关键信息 materials, participants和instruments我应该放在哪个部分写
- 如何有条理地展现我的工作流程, 避免像流水账





Chapter B: Structure and content arrangement



Detailed study design

- 通常适配于模型实验或者实验室测试
- 你的实验模型也是你的科研成果(应用型研究)
- 当你的实际操作流程是严格按照实验步骤进行的

Subheading: 明确内容 (design)

Content:

- 具体说明实验背景条件
- 实验的流程 (通常聚焦于数据如何产生的或者是实验产品的生产)
- 详细介绍materials, participants和instruments, 必要时引入图像

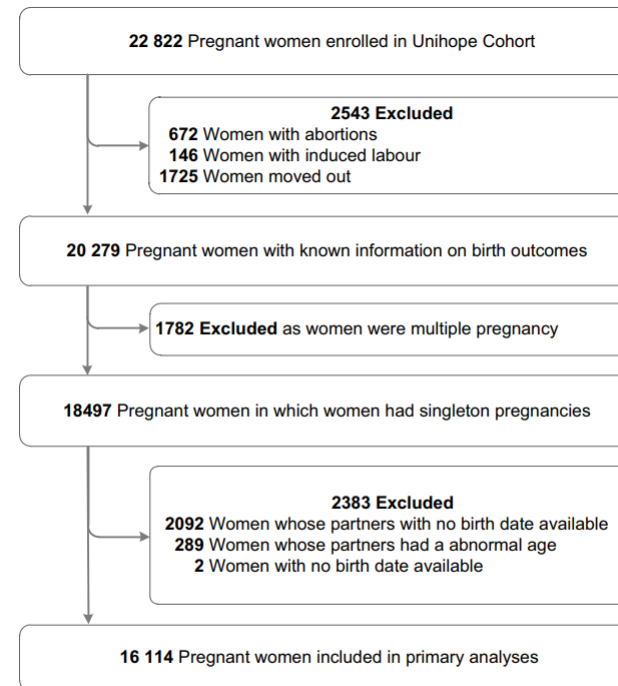


Fig. 1 Flow chart of participants selection

2.1 Study design and participants

The multicenter prospective UNIHOPe cohort was conducted in China from July 2016 to June 2021, aiming to explore potential predictors of adverse pregnancy outcomes among women of advanced age (ClinicalTrials.gov: NCT03220750) [13]. The cohort was conducted in eight public referral hospitals located in Beijing, Shanghai, Guangzhou, Shenyang, Wuhan, Chongqing, and Chengdu, scattered through the eastern, central, and western regions of China. Pregnant women were enrolled before 14 weeks of gestation, and followed up at 24–28 weeks, 32–34 weeks of gestation, delivery, and 6–12 weeks postpartum, by obstetrician or nurse in the hospitals. Information on sociodemographic characteristics and lifestyle (maternal and paternal), medical and reproductive history, prenatal care of the current pregnancy, and pregnancy outcomes was collected at enrollment and subsequent follow-ups using a structured questionnaire.

Pregnant women were eligible in the UNIHOPe cohort if they attended regular antenatal care and delivered at the included hospitals. Women would be excluded if they had mental disorders or had no ability to provide informed consent. Initially, a total of 22,822 pregnant women were enrolled in the UNIHOPe cohort. For this study, 6708 women who were ended with spontaneous or induced abortions ($n=818$), moved out ($n=1725$), had multiple pregnancies ($n=1782$), had no birthdate ($n=2$), or whose partners had no birthdate ($n=2092$) or abnormal age (<20 or >70 years, $n=289$) were excluded. Finally, 16,114 pregnant women remained in the analysis (Fig. 1).



Detailed study design

subheading: 明确这一步的具体工作

Note: 不要用“Procedures”之类的笼统表达作 subheading

content:

- 完善（具体化）整个实验流程，对于后续步骤进行详尽的描述
- 注意明确一些不涉及“participants”入选标准，但是已经对结果有影响的参与者特质的因素

Note: 前文出现的materials, participants和instruments不需要再次介绍; 数据观测, 收集分析等阶段, 如有前文未提及的materials, participants和instruments参与其中, 请及时说明

2.2 Exposures and covariates

The main exposure was paternal age, which was defined as the male partner's age at the time of conception, calculating as subtracting the partner's birthdate from the last menstrual period date. Paternal age was used continuously or categorically (<30, 30–34, 35–39, 40–44, and ≥ 45 years) in this analysis, as appropriate. Maternal age was calculated as subtracting the maternal birthdate from the date of delivery, and categorized into four groups: <35, 35–39, 40–44, and ≥ 45 years.

Other covariates included paternal occupation, alcohol consumption and smoking, as well as maternal ethnicity, education, occupation, alcohol consumption and smoking

2.3 Outcomes of interest

2.4 Statistical analysis

To assess the robustness of the results, sensitivity analyses were conducted by using multiply imputed missing values for covariates, or by excluding women who conceived with ART. Additionally, ordinal regression models were used to assess the associations of the paternal and maternal age with perinatal outcomes.

Statistical analyses were performed using SAS software, version 9.4 (SAS Institute, Inc). A 2-sided $P < 0.05$ was deemed statistically significant.



Brief study design

In this section, smart traffic vehicle management using Faster R-CNN based deep learning based ensemble method is highlighted. The research problem revolves around traffic management which involves vehicle segmentation at different background levels⁵. In order to achieve this, we study Faster R-CNN⁵ which analyses vehicles in smart traffic. The segmentation activity on various strategic areas of smart traffic analytics provide information related to decision-making activities. (Chaudhuri, 2024)

- 粗略说明实验的方向以及涉及的关键要素（通常是对应实验问题，呼应introduction）

The proposed system is based on the steps for classifying protein sequence families shown in Fig. 1.



(Abu-Qasmieh et al., 2023)

Figure 1: The proposed classification method design

- 简单的流程图预告实验的全过程也是常见的方式

Brief study design

subheading: 具体展示实验步骤

content:

- 根据实际情况，展示从数据产生到分析的全流程，按照流程顺序展示对应环节materials, participants和instruments的具体信息
- 所有的关键信息都应该按照最后的实际情况进行描述，可借助图表展示相关信息

Dataset	Low density traffic	High density traffic
Source images	5000	1400
Annotations	50,595	2000
Classes		
Two wheeler	9000	120
Three wheeler	9351	200
Four wheeler	10,879	880
Six wheeler	10,500	600
Eight wheeler	5990	110
Ten wheeler	4875	90

Table 1. Statistics of dataset images.

Datasets used

The experimental datasets are adopted from Ref.⁵⁸. The datasets are prepared keeping in view different traffic conditions. At the very first instance multi-class vehicle objects are considered. Several challenge factors such as traffic jams and overlapping vehicles are incorporated in dataset. Broadly speaking datasets are developed with respect to two different situations viz high density and low density traffic scenes. The former comes with several objects in single image while later has single class per image having no overlapping. In order to achieve better training, images from both above mentioned situations are placed in different datasets. The high density traffic scenes are considered from several places having traffic which are less crowded in nature. A total of 1000 images from six classes of vehicles viz two wheelers, three wheelers, four wheelers, six wheelers, eight wheelers and ten wheelers are developed. Figure 2a,b highlight some sample images. High density traffic scenes are considered from daily traffic which are congested in nature. 5000 images from above stated classes are created. Certain critical factors such as illumination, occlusions etc. are incorporated irrespective of appearance, shape, scale and size in this dataset. Table 1 highlights certain high and low density dataset statistics along with annotations of class. These datasets are preprocessed and augmented as discussed in “Data annotation and augmentation” section. Each of these datasets are benchmarked with respect to standard datasets⁵⁸.

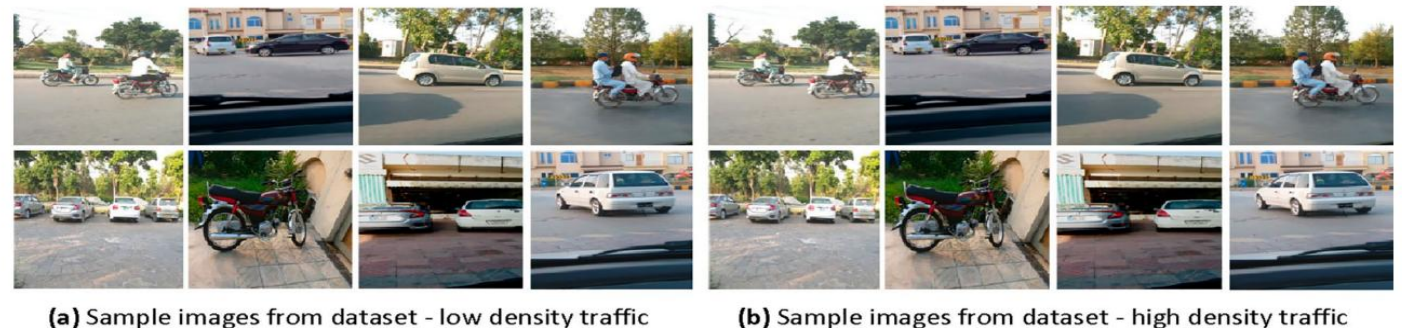


Figure 2. (a) Sample images from dataset—low density traffic (b) sample images from dataset—high density traffic.



Brief study design

由于study design没有一个完整的流程展现后文需要一个对于整体流程的展示，可以借助视图

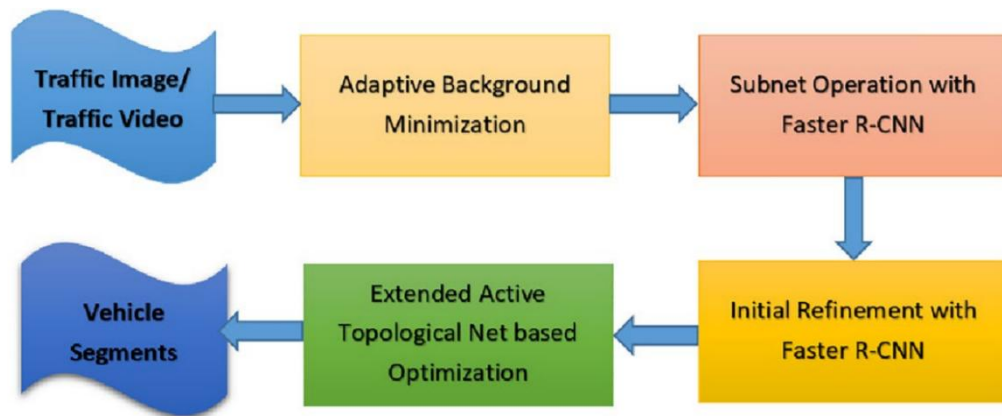


Figure 4. Architecture of vehicle segmentation method.

(Chaudhuri, 2024)

对于前文有流程图进行粗略展示的完整流程的，后文的subheading通常是对应前文的流程图的

The proposed system is based on the steps for classifying protein sequence families shown in Fig. 1.



Figure 1: The proposed classification method design

2.1 The Selected Protein Sequence Superfamily

2.2 Encoding Systems

2.3 Bispectrum

2.4 Deep Learning (Abu-Qasmieh et al., 2023)

Abu-Qasmieh, I., Fahoum, A. A., Alquran, H., & Zyout, A. (2023). An innovative bispectral deep learning method for protein family classification. *Computers, Materials & Continua (Print)*, 75(2), pp. 3971–3991.

Chaudhuri, A. (2024). Smart traffic management of vehicles using faster R-CNN based deep learning method. *Scientific Reports*, 14(1). <https://doi.org/10.1038/s41598-024-60596-4>



Chapter C: Specific writing guide



Tense, voice and writing style

"We measured canopy leaf temperature at a 30-m canopy walk-up tower between July to December of 2004 and July to December of 2005 at the same site. We initially placed 50 thermocouples on canopy-exposed leaves of *Sextonia rubra*, *Micropholis sp.*, *Lecythis lurida* (originally published in Doughty and Goulden⁸). Fine-wire thermocouples (copper constantan 0.005 Omega) were attached to the underside of leaves by threading the wire through the leaf and inserting the end of the thermocouple into the abaxial surface. The thermocouples were wired into a multiplexer attached to a data logger (models AM25T and 23X, Campbell Scientific) and the data were recorded at 1 Hz."

1-时态上：描述实际的科研工作，通常使用一般过去时

2-语态上：主动式(We measured / We initially placed)与被动式(were attached / were wired into / were recorded at)相结合

主动语态能够凸现科研人员的行动，使整体表达效果显得更加积极，生动

被动语态方便作者组织句子，突出一些重要的【客体】

3-写作风格上：简练明确

基本不使用长传名词做修饰

使用专业名词，而非带形容词的描述性表达

副词的使用格外小心，有重要的实际含义，以及表达逻辑的连接词



How to show the details of materials, participants and instruments?

"The ECOSTRESS mission is a thermal infrared multispectral scanner with five spectral bands at 8.28, 8.63, 9.07, 10.6 and 12.05 μm . The sensor has a native spatial resolution of $38 \times 68 \text{ m}$, resampled to $70 \times 70 \text{ m}$, and a swath width of 402 km (53°).....ECOSTRESS is an improvement over other thermal sensors because..... To ensure the highest quality data, we used ECOSTRESS quality flag 3520, which identifies the best-quality pixels (no cloud detected) We accessed ECOSTRESS LST data through AppEEARS (<https://lpdaac.usgs.gov/tools/appeears/>) for the following products.....We calculated 'peak median', which is defined as the average of the highest three medians of each granule"

1-时态上：介绍materials, participants和instruments的基本情况时使用一般现在时 (is) , 当你切换回实际的科研工作时, 切换回一般过去时 (we assessed)

(同理, 原理, 标准和公式等, 也用一般现在时: which is defined as the average of)

2-信息点

影响实验结果的相关背景或者规格

为什么选择ta? (适配性与先进性: ECOSTRESS is an improvement over other thermal sensors)

具体说明如何使用它 (以什么状态: we used ECOSTRESS quality flag 3520 /we accessed)



How to show the details of materials, participants and instruments?

更加简单的展示

适用于参与实验，但是其规格型号对于结果影响较少的

We estimated canopy temperature at the KM 83 eddy covariance tower in the Tapajos region of Brazil^{41–43} using a pyrgeometer (Kipp and Zonen) mounted at 64 m to measure.....

We created a model of individual leaves on a tree (100 × 100 grid where each pixel is a leaf) using MATLAB (Mathworks version 2022a) to estimate the upper limit

The thermocouples were wired into a multiplexer attached to a data logger (models AM25T and 23X, Campbell Scientific) and the data were recorded at 1 Hz.

直接在相关的步骤描写中提及，以括号备注的方式提供必要信息（制造商，版本号等等），以及使用的方式或者目的

排除无效信息，有效避免列举式的流水账

保证描写的流畅度与完整性



Describe the complex process

- 首先明确这一实验（步骤）的目的以及必要的实验信息（地点，实验方式，分组情况）
- 在分组实验中，如果相互之间的实验本身是独立的，那么在写作的时候，本身实验的先后顺序不具备参考价值，要按照一定的“学术逻辑”去展示。
- 在每一组中，共有的基础信息是必须包括的（具体地点，物种类型，加热温度）。
- 可以根据实际情况说明每一组的特殊情况
- 前文可能已经在展示分组实验的时候提到了“实验数据”，但是不要忘记说明实验数据如何收集的

Tips:

The first / second /final experiment 这种写法是非常推荐的，用顺序词让架构更加清晰

一部分信息不需要单独描述，直接以括号的形式备注即可，避免信息过于冗杂且散碎

For model validation, we used the results of three upper canopy leaf and branch warming experiments of 2 °C (Brazil)¹⁷, 3 °C (Puerto Rico)¹⁸ and 4 °C (Australia)²⁰. The first experiment (Brazil) were four individual leaf-resistant heaters on each of six different upper canopy species at the Floresta Nacional (FLONA) do Tapajos as part of the Large-Scale Biosphere–Atmosphere Ecology Program in Santarem, Brazil¹⁷. On the same six species, black plastic passively heated branches by an average of approximately 2 °C. Initially, heat balance sap flow sensors and the passive heaters were added to 40 branches, but we had confidence in the data from 9 heated and 4 control in the final analysis. The second experiment (Puerto Rico) had two species (*Ocotea sintenisii* (Mez) Alain and *Guarea guidonia* (L.) Sleumer where leaves were heated by 3 °C at the Tropical Responses to Altered Climate Experiment canopy tower site at the Sabana Field Research Station, Luquillo, Puerto Rico¹⁸. The final experiment (Australia), which increased leaf temperatures by 4 °C, was conducted at Daintree Rainforest Observatory in Cape Tribulation, Far North Queensland, Australia²⁰. Leaf heaters were installed using a pair of 30-G copper-constantan thermocouples, one reference leaf and one heated with a target temperature differential of 4 °C. There were two pairs in the upper canopy of each tree crown installed in 2–3 individuals across four species with the thermocouples installed on the underside of the leaves. Two absolute 36-G copper-constantan thermocouples were installed in each species to measure the leaf temperatures of the reference leaves. Thermocouple wires connected into an AM25T multiplexer from Campbell Scientific connected to a CR1000 Campbell datalogger. More details about the experiment and sensors can be found in ref. 20.



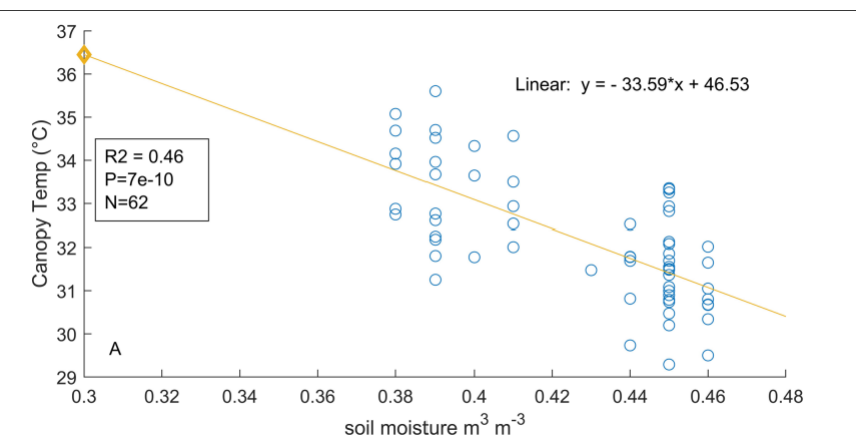
Functions and requirements of charts

- **Relevant:** 图表里展示的信息一定是和研究主题密切相关的
- **Clear and Concise:** 单一的图表包括的信息一定是围绕着同一个主题的，以一种清晰明了，让人能够理解方式呈现

Note: 图表的阅读难度应该匹配本文的阅读难度。如果一位读者能够理解introduction和本图表出现之前的methods，那么视图传达的信息能够被其有效接收，则本图表可以被看作是“Clear”

如果图表很难被理解，可以尝试拆解或者换一种形式的图表

- **Self-contained:** 由标题，图例，数据，单位，注解以及图像构成的完整的单一图表，应该不需要正文中额外的文字解说，就能够传递有意义的信息



Extended Data Fig. 2 | Impacts on canopy temperature. (A) Linear regression of canopy temperature versus soil moisture (40 cm depth) at the km 83 eddy covariance tower ($r^2 = 0.46$, $P = 7e-10$, $N = 62$)



Chapter D: Tips for Writing



Some common questions

q: 有传言methods是写给专业人士看的，我还需要解释专业内容么？

a: 当你创作methods时，要理解发表的学术论文是具有一定科普性质的（避免对于读者过度专业化的推测），所以如果使用了特定的术语或非标准方法，请提供足够的解释或背景信息

Yin et al. (2024): abnormal age (<20 or > 70 years, n=289)

q: 为什么很多论文我都没看见study design?

a: 因为根据结构需要与连贯性问题

一些study design会在introduction尾部出现，

另外，比如有单独的Scope of Research或者

Research question的论文，也可能出现在

这两部分

Here we used data from the new ECOSTRESS sensor to estimate peak pantropical forest canopy temperatures. We began by ground truthing the satellite data with tower-based pyrgeometer data. We then used these data to determine what causes variation in peak temperatures at the canopy scale and show similar trends driving peak temperatures across all of the Tropics. Critically, we show that for a given canopy temperature, individual leaf temperatures display a 'long tail' of values in the distribution, in which the temperatures of a few individual leaves far exceed that of the overall canopy, and that this skewed distribution persists under leaf warming experiments of 2, 3 and 4 °C. Finally, we developed a simple empirical model to explore the implications of observed leaf temperatures on the fate of tropical forests under future climate change.

Doughty, C. E., Keany, J., Wiebe, B. C., Rey-Sánchez, C., Carter, K., Middleby, K. B., Cheesman, A. W., Goulden, M. L., Da Rocha, H. R., Miller, S. D., Malhi, Y., Fauset, S., Gloor, E., Slot, M., Menor, I. O., Crous, K. Y., Goldsmith, G. R., & Fisher, J. B. (2023). Tropical forests are approaching critical temperature thresholds. *Nature*, 621(7977), 105–111. <https://doi.org/10.1038/s41586-023-06391-z>

Yin, S., Zhou, Y., Zhao, C., Yang, J., Yuan, P., Zhao, Y., Qi, H., & Wei, Y. (2024). Association of Paternal Age Alone and Combined with Maternal Age with Perinatal Outcomes: A Prospective Multicenter Cohort Study in China. *Journal of Epidemiology and Global Health*. <https://doi.org/10.1007/s44197-023-00175-4>



Some common questions

q: 关于materials, participants和instruments, 我到底要展示哪些相关信息, 细化到什么程度? 看了很多论文, 有粗略的, 有细化的, 我应该如何把握?

a: 除之前提到对实验有影响的信息, 其余的信息还需要哪些, 你要参考你的target journal最近发表的论文, 相似的内容, 它细化到什么程度, 有什么必要信息

pyrgeometer (Kipp and Zonen) ; MATLAB (Mathworks version 2022a)→ nature至少要有制造商, 软件甚至需要版本 (Doughty, C. E., Keany, J., Wiebe, B. C., Rey-Sánchez, C., Carter, K., Middleby, K. B., Cheesman, A. W., Goulden, M. L., Da Rocha, H. R., Miller, S. D., Malhi, Y., Fauset, S., Gloor, E., Slot, M., Menor, I. O., Crous, K. Y., Goldsmith, G. R., & Fisher, J. B. (2023). Tropical forests are approaching critical temperature thresholds. *Nature*, 621(7977), 105–111. <https://doi.org/10.1038/s41586-023-06391-z>)

q: 数据都是大量的文本, 图片或者类似的数据, 要在methods展示么?

a: 传统的方法是将原始数据放在Appendix里, 比如社科研究的Questionnaire and responses, 目前也可以上传到云端, 然后提供一个可以查看的链接

Data availability

The data supporting the findings of this work are available in the paper, the Extended Data and the ETH Zürich Research Collection repository (open access <https://doi.org/10.3929/ethz-b-000627348>).



Checklist of a good methods section

- **Verifiability and Replicability:**
 - 明确描述研究方法、实验的真实流程，以便其他研究人员可以验证你的结果。
 - 提供materials, participants和instruments的相关细节，例如人员背景，实验设备、软件版本、数据集来源等。
 - 引用相关文献以支持所采用的方法和技术。
 - 如果可能，提供实验代码、数据集和预处理步骤，以便其他研究人员可以轻松地重现结果。
- **Transparency and Comparability:**
 - 描述数据产生，收集和分析的过程，说明研究存在的限制和潜在偏见。
 - 在适当的情况下，提供实验的失败尝试，以展示研究的全貌。
- **Readability:**
 - 使用清晰、简洁的语言描述方法和结果。
 - 使用段落、标题和子标题来组织内容，使读者更容易理解。
 - 使用图表以及示例来辅助解释复杂的方法和技术。
- **Providing Evaluation Criteria**



Practical advices

- 为了避免Confounding Variables (未控制的混杂变量：未能识别或控制可能影响因变量的其他变量) 在实验设计初期进行变量的识别和分类，并采用随机分配或匹配等策略来控制这些变量
- 如果你“感觉”你的实验模型过于复杂 (Overfitting)，捕捉到数据中的随机噪声而非实际信号，不能展示你的真实情况，那么使用交叉验证技术来检验模型的泛化能力，避免使用过多的预测变量。
- 样本的大小需要在研究前进行功效分析来确定，以保证研究有足够的统计功效，能够尽可能地代表实际情况，避免样本过少 (Small Sample Size)。
- 在撰写methods之前，需要考虑实验的目的
 - 1-对照性分析或者是改进过程，就需要在methods包含“失败”的尝试
 - 2-如果是单纯追求某个结果，那么就不要再提及。注意results与discussion的部分保持和methods的逻辑一直



Practical advices

- 社科类研究，如果只有问卷统计的结果，那么除了问卷的内容与受访者以外，必须在写作时明确你的问卷是线上，还是线下，口头还是书面，即时反馈或者其他形式
- 写作中，注意区分：
 - 1-解释为什么要使用这些策略和方法，以及如果需要说明这些策略和方法的背景信息，需要放在 introduction (LR) 里面，把这些内容作为前人的研究成果展示。
 - 2-在methods的部分，通常只涉及介绍这个策略/方法具体是什么，然后如何利用或者遵照其展开研究
- 除了作者团队和参与者以外的人/团队，参与到数据观测，收集，处理等阶段，需要在methods里体现，说明他们是符合专业要求的（比如学术背景，职业背景，取得的执业资格等）
- 完成methods初稿之后，检查methods与introduction (research question) 与results的逻辑对照关系。

更多在线讲座以及课程

Join us for future Academic Publishing Upcoming on demand courses

精品课程



EVERY AUTHOR 学术写作常见错误 全解析

PUBLISHING GUIDE FOR THE BEGINNER (MANDARIN CHINESE)

FREE WEBINAR



EVERY AUTHOR 学术英语第一课 An Introduction to Academic English

October 16, 2023 02:00 PM Beijing CST
October 16, 2023 02:00 PM London BST
<https://www.everyauthorpub.com>

精品课程



EVERY AUTHOR 国际期刊论文 零基础出版指导

PUBLISHING GUIDE FOR THE BEGINNER (MANDARIN CHINESE)

FREE WEBINAR



EVERY AUTHOR 如何选择学术期刊

What to consider when
choosing an academic journal

To be confirmed
www.everyauthorpub.com

更多课程即将上架，请扫描二维码小程序关注！





EVERY AUTHOR

www.everyauthorpub.com



关注我们 Every Author 作者成长学院

更多课程即将上架，请扫描二维码小程序关注！



EVERY AUTHOR
作者成长学院
视频上线！



扫一扫
即刻关注



Every Author微信视频号

关注我们

Instagram: EveryAuthor_Asia

Twitter: EveryAuthor

Website: www.everyauthorpub.com

客服微信（普通话）：everyauthor