arXiv:2305.01821v1 [q-bio.QM] 2 May 2023

pyPESTO：用于动态模型参数估计的模块化且可扩展工具

YannikSchälte[1,2,3,*]、Fabian Fröhlich[4,*]、Paul J. Jost[1,*]、Jakob Vanhoefer[1,*]、Dilan Pathirana[1],Paul Stapor[2,3],Polina Lakrisenko[2,5],DantongWang[2,3],ElbaRaimúndez[1,2,3],Simon Merkt[1],Leonard Schmiester[2,3],PhilippStädter[2,3,6],StephanGrein[1],ErikaDudkin[1],Domagoj Doresic[1], Daniel Weindl[2],和 Jan Hasenauer[1,2,3,†]

[1]生命与医学科学研究所（LIMES）、波恩大学，波恩 53113，德国[2]计算健康中心、亥姆霍兹慕尼黑中心—德国健康与环境研究中心（GmbH），诺伊赫贝格 85764，德国[3]数学系、慕尼黑工业大学，加兴 85748，德国[4]系统生物学系、哈佛医学院，波士顿，MA 02115，美国[5]生命科学学院、慕尼黑工业大学，弗赖辛 85354，德国[6]莱布尼茨天然产物研究与感染生物学研究所，耶拿 07745，德国，[*]这些作者贡献相同

[†]联系邮箱：*jan.hasenauer@uni-bonn.de*

摘要

机制模型是描述和理解生物过程的重要工具。然而，它们通常依赖未知参数，而对于大型复杂系统，估计这些参数可能颇具挑战。我们介绍 pyPESTO，这是一个用于系统化参数估计的模块化框架，提供了用于优化和不确定性量化的可扩展算法。虽然 pyPESTO 针对常微分方程问题进行了优化，但也广泛适用于黑箱参数估计问题。除自有实现外，它还提供了统一接口，可调用多种常用的模拟和推断方法。pyPESTO 使用 Python 实现，并以三条款BSD许可证开源。代码和文档可在 GitHub 上获取（https://github.com/icb-dcm/pypesto）。

1 引言

在包括计算生物学在内的许多研究领域，数学模型是研究复杂系统并理解其潜在机制的重要工具（Kitano，2002）。尽管描述生物系统有多种形式化方法，常微分方程（ODE）模型仍广受欢迎，因为它们能够以自然的方式描述和解释实验生物学中常见的扰动后动态变化（Fröhlich等，2019）。然而，模型通常包含未知参数，需要根据观测数据估计其数值和不确定性（Tarantola，2005）。我们介绍了 pyPESTO，这是一种基于Python的参数估计工具，通过简化流程以模块化方式提供多种推断方法（图 1；有关工具比较，请参见补充信息第 1 节）。pyPESTO 可应用于多种问题类型，包括大规模问题。



图 1：概念示意图。pyPESTO 涵盖完整的参数估计工作流程，从问题定义、参数优化、不确定性量化，到可视化与分析。

2 特性

2.1 问题定义

为了便于应用于生物系统，pyPESTO 支持用于规范参数估计问题的社区标准 PEtab（Schmiester 等，2021），尤其对常微分方程模拟和敏感性分析引擎 AMICI 提供接口（Fröhlich等，2021）。此外，pyPESTO 允许用户通过标量目标函数或向量值残差函数定义连续参数估计问题。这包括基于对数似然和对数后验的目标函数，在此情况下，估计值和不确定性可以进行统计学解释。由于许多推断方法受益于目标函数导数，pyPESTO 支持用户提供用于计算导数的函数，同时也提供自适应有限差分。

2.2 优化

将全局最优参数作为点估计求出，是参数估计中常见的起点。由于许多问题具有非凸性且存在多个局部最优解，因此需要采用全局化策略。为此，pyPESTO 提供了全局优化器接口，以及适用于局部和全局优化器的多起点全局化策略；该策略在生物学问题中表现良好（Raue 等，2013）。pyPESTO 为 Ipopt、Dlib、PySwarms、pycma、SciPy、NLopt 和 Fides 等局部及全局优化库提供了统一接口（见补充信息第 2.1 节）。扩展到其他工具也很容易。此外，pyPESTO 还提供了一种分层方法，可高效处理相对数据和噪声参数（Schmiester 等，2019a）以及序数数据（Schmiester等，2019b）。

2.3 不确定性分析

在不确定性分析方面，pyPESTO 使用前述面向不同优化库的接口，实现了基于优化的（频率学派）剖面似然和（贝叶斯）剖面后验方法（Raue等，2009年）。此外，pyPESTO 还实现了自适应 Metropolis（Haario 等，2001年）和自适应并行回火（Miasojedow等，2013）贝叶斯采样算法，并为流行的采

样工具 Emcee、PyMC 和 Dynesty（参见补充信息第 2.2 节）提供统一接口，尤其支持基于梯度的采样。

2.4 进一步方面

pyPESTO 提供了多种例程，用于可视化和分析所得结果。结果可以采用基于 HDF5 的紧凑存储格式保存、恢复和共享。此外，pyPESTO 支持通过多线程和多进程实现共享内存并行化。因此，pyPESTO 可灵活部署在桌面计算机和高性能计算基础设施上。

3 实现与可用性

pyPESTO 使用 Python 实现，采用三条款BSD许可证开源。该代码旨在实现模块化和可扩展性，托管于 GitHub，也可从 PyPI 安装。ReadTheDocs 上提供了详尽的文档，其中包括许多教程笔记本，介绍 pyPESTO 的功能。我们通过单元测试和持续集成确保开发过程中的正确性。

4 讨论

pyPESTO 已被至少 12 篇出版物采用，目前由 3 家机构的至少 4 位核心贡献者持续开发。未来，我们计划实现更多优化和不确定性量化算法，使 pyPESTO 能够与其他流行工具对接，并扩展和进一步标准化其支持的参数估计工作流程。我们预计，pyPESTO 将继续广泛应用于计算生物学及其他领域。

资金支持

本研究获德国研究联合会（DFG）通过德国卓越战略（EXC 2047/1 390685813 和 EXC 2151 390873048）、SFB 1454（代谢炎症）、TRR333（BATenergy）、AMICI（443187771）、人类前沿科学计划（LT000259/2019-L1；FF）、德国联邦教育与研究部（BMBF）（01ZX1916A；DW、PL）、波恩大学施莱格尔教授职位（JH）以及约阿希姆·赫茨基金会（YS）资助。

参考文献

Fröhlich, F. 等（2019）。生化过程常微分方程模型的可扩展推断。载于 G. Sanguinetti 和 V. A. Huynh-Thu 编，《style id='3'>基因调控网络：方法与实验方案，第 1883 卷《style id='5'>分子生物学方法，第 16 章，第 385–422 页。Humana Press，第 1 版。

Fröhlich, F. 等（2021）。AMICI：面向大型常微分方程模型的高性能灵敏度分析。《生物信息学》，**btab227**。btab227。Haario, H. 等（2001年）。一种自适应 Metropolis 算法。《Bernoulli》，**7**（2），223–242。Kitano, H.（2002）。系统生物学：简要概述。《科学》，**295**（5560），1662–1664。Miasojedow, B. 等（2013）。一种自适应并行回火算法。《计算与图形统计学期刊》，**22**（3），649–664。Raue, A. 等（2009年）。通过利用剖面似然对部分观测动态模型进行结构和实际可辨识性分析。《生物信息学》，**25**（25），1923–1929。Raue, A. 等（2013）。从系统生物学中的定量动态建模中汲取的经验。PLoS ONE，**8**（9），e74335。Schmiester, L. 等（2019a）。基于相对测量的大规模动态模型高效参数化。《生物信息学》，**36**（2），594–602。Schmiester, L. 等（2019b）。使用高效最优尺度方法从定性数据进行机制模型的统计推断。BioRxiv，第 848648 页。Schmiester, L. 等（2021）。PEtab——系统生物学中参数估计问题的可互操作规范。《PLOS计算生物学》，**17**（1），1–10。Tarantola, A.（2005）。《模型参数估计的逆问题理论与方法》。SIAM。

arXiv:2305.01821v1 [q-bio.QM] 2 May 2023

补充信息：
pyPESTO：用于动态模型参数估计的模块化、可扩展工具

Yannik Sch¨alte[1,2,3,*], Fabian Fr¨ohlich[4,*], Paul J. Jost[1,*], Jakob Vanhoefer[1,*], Dilan Pathirana[1], Paul Stapor[2,3], Polina Lakrisenko[2,5], Dantong Wang[2,3], Elba Raim´undez[1,2,3], Simon Merkt[1], Leonard Schmiester[2,3], Philipp St¨adter[2,3,6], Stephan Grein[1], Erika Dudkin[1], Domagoj Doresic[1], Daniel Weindl[2], 和 Jan Hasenauer[1,2,3,†]

[1]生命与医学科学研究所（LIMES），波恩大学，德国波恩 53113[2]计算健康中心，亥姆霍兹慕尼黑研究中心（德国环境与健康研究中心），德国诺伊赫贝格 85764¨unchen ¨ur Gesundheit und Umwelt (GmbH), Neuherberg 85764, Germany[3]数学系，慕尼黑工业大学，德国加兴 85748¨at M¨unchen, Garching 85748, Germany[4]系统生物学系，哈佛医学院，美国马萨诸塞州波士顿 02115[5]生命科学学院，慕尼黑工业大学，德国弗赖辛 85354[6]莱布尼茨天然产物研究与感染生物学研究所，德国耶拿 07745，[*]这些作者贡献相同

[†]联系：*jan.hasenauer@uni-bonn.de*

目录

1 仿真与推断工具功能比较 **2**

2 已接入的接口工具列表 **2**

2.1 Optim化 2

2.2 Sampling 2

1　仿真工具与推断工具功能对比

表 1：各种系统生物学仿真与推断工具的功能比较（为便于阅读而换行）。

Tool	支持的模型	建模/数据标准	优化				不确定性分析				并行化
			局部	全局	多起点	分层	简介 全局	局部近似	采样 MCMC	PT	
AMICI	ODE	系统生物学标记语言、PySB、BNGL、PEtab	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	OpenMP
AMIGO2	Any	No	Yes	Yes	Yes	No	No	Yes	No	No	No
COPASI	常微分方程，马尔可夫跳跃过程	系统生物学标记语言、PEtab	Yes	Yes	No	No	Yes	Yes	No	No	No
data2dynamics	ODE	PEtab	Yes	Yes	Yes	No	Yes	Yes	Yes	No	pthreads、并行 for 循环
dmod	ODE	PEtab	Yes	Yes	Yes	No	Yes	No	No	No	mclapply、dopar
MEIGO64	Any	PEtab	Yes	Yes	Yes	No	No	No	Yes	No	jPar
parPE	ODE	PEtab	Yes	No	Yes	Yes	No	No	No	No	MPI
PEPSSBI	ODE	SBML	Yes	Yes	Yes	No	No	No	No	No	各种、
PESTO	任意；例如常微分方程	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	并行 for 循环
pyBioNetFit	ODE	BNGL、系统生物学标记语言	Yes	Yes	No	No	No	No	Yes	Yes	多种
Tellurium+SBstoat	常微分方程，马尔可夫跳跃过程	SBML	Yes	Yes	Yes	No	No	No	No	No	No
pyPESTO	任意；尤其是常微分方程	PEtab	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	多种

Tool	梯度	模型模拟	语言	最后更新	Link
AMICI	FW 和 AJ	Yes	Python、C++	2023	https://amici.readthedocs.io/en/latest/
AMIGO2	FD	Yes	MATLAB	2021	https://sites.google.com/site/amigo2toolbox/
COPASI	FD	Yes	C++	2023	https://copasi.org/
data2dynamics	有限差分和 FW	Yes	MATLAB	2023	https://github.com/Data2Dynamics/d2d
dmod	FW	Yes	R	2022	https://github.com/dkaschek/dMod
MEIGO64	可传递	N/A	MATLAB	2023	https://github.com/gingproc-IIM-CSIC/MEIGO64
parPE	FW 和 AJ	不适用（使用 AMICI）	C++	2023	https://github.com/ICB-DCM/parPE
PEPSSBI	有限差分和 FW	Yes	Java	2020	https://bitbucket.org/andreadega/systems-biology-compiler
PESTO	有限差分，例如可传入 FW ＋ AJ	不适用（使用 AMICI）	MATLAB	2020	https://github.com/ICB-DCM/PESTO
pyBioNetFit	No	不适用（BioNetGen ＋ libRoadRunner 接口）	Python	2023	https://bionetfit.nau.edu/
Tellurium+SBstoat	FW	Yes	Python	2023	https://github.com/sys-bio/SBstoat
pyPESTO	有限差分，例如可传入 FW+AJ	不适用，模拟器的通用接口，尤其是 AMICI	Python	2023	https://pypesto.readthedocs.io/en/latest/

颜色编码缩写 很好 不适用 不适用（例如，因为该工具仅用于模拟或仅用于推断，而非两者兼有） 尚可 马尔可夫链蒙特卡罗（单链）
罗较差 并行回火 常微分方程 离散状态空间马尔可夫跳跃过程 有限差分 前向灵敏度计算 伴随灵敏度计算

据我们所知，本综述于2023年5月编写。pyPESTO 最初是对 MATLAB 工具 PESTO 的重新实现（Stapor等，2018年），如今则提供了简化流程，涵盖范围更广的现代方法和接口工具。

2 接口工具列表

2.1 优化

目前，pyPESTO 接口支持以下优化工具：Ipopt（W¨achter 和 Biegler，2006年）、Dlib（King，2009年）、PySwarms（Miranda，2018年）、pycma（Hansen，2022年）、SciPy（Virtanen 等，2020年）、NLopt（Johnson 和 Schueller，2021年）以及 Fides（Fr¨ohlich 和 Sorger，2022年）。

2.2 采样

目前，pyPESTO 实现了贝叶斯采样算法自适应 Metropolis（Haario等，2001年）和自适应并行回火（Vousden 等，2016年），并为流行的采样工具 Emcee（Foreman-Mackey 等，2013年）、PyMC（Salvatier等，2016年）和 Dynesty（Koposov，2022年）提供统一接口。

参考文献

Foreman -Mackey, D. 等(2013). emcee：MCMC 锤子。*arXiv*，**125**(925), 306 .

Fr¨ohlich, F. 和 Sorger, P. K. (2022). Fides：用于常微分方程模型参数估计的可靠信赖域优化。《*PLOS*计算生物学》，**18**(7), e1010322. Haario, H. 等(2001). 一种自适应 Metropolis 算法。《*Bernoulli*》，**7**(2), 223–242. Hansen, N. (2022). pycma：r3.2.2. Johnson, S. G. 和 Schueller, J. (2021). NLopt：非线性优化库。天体物理学源代码库，页码 ascl–2111. King, D. E. (2009). Dlib-ml：机器学习工具包。《机器学习研究期刊》，**10**, 1755–1758. Koposov, S. (2022). dynesty：v2.0.2. Miranda, L. J. V. (2018). PySwarms：用于 Python 粒子群优化的研究工具包。《开放源代码软件期刊》，**3**. Salvatier, J. 等(2016). 使用 pymc3 在 Python 中进行概率编程。《*PeerJ*计算机科学》，**2**, e55. Stapor, P. 等(2018). PESTO：参数估计工具箱。《生物信息学》，**34**(4), 705–707. Virtanen, P. 等(2020). SciPy 1.0：Python 科学计算的基础算法。《自然·方法》，**17**, 261–272. Vousden, W. 等(2016). 用于马尔可夫链蒙特卡罗模拟中并行回火的动态温度选择。《皇家天文学会月刊》，**455**(2), 1919–1937. W¨achter, A. 和 Biegler，L. T. (2006). 关于用于大规模非线性规划的内点过滤线搜索算法的实现。《数学规划》，**106**(1), 25–57.