FERMILAB-CONF-26-0695-CSAID

从历史观测计划到在轨控制：4米布兰科望远镜的自主调度

Rachel Hur
芝加哥大学物理系，南埃利斯大道5720号，芝加哥，伊利诺伊州60637；NSF-西蒙斯天空人工智能研究所（SkAI），东栗树街172号，芝加哥，伊利诺伊州60611；rhur@uchicago.edu

Paul Chichura
NSF-西蒙斯天空人工智能研究所（SkAI）172 E. Chestnut St.,Chicago, IL 60611 芝加哥大学天文学与天体物理学系 5640 S. Ellis Ave., Chicago, IL 60637 pchichura@uchicago.edu

Alex Drlica-Wagner
费米国家加速器实验室 P.O. Box 500, 巴达维亚, IL 60510 芝加哥大学天文学与天体物理学系 5640 S. Ellis Ave., Chicago, IL 60637 NSF-西蒙斯天空人工智能研究所（SkAI） 172 E. Chestnut St., Chicago, IL60611 kadrlica@fnal.gov

Eric H.NeilsenJr.费米国家加速器实验室 P.O. Box 500, 巴达维亚, IL 60510 neilsen@fnal.gov

Guillermo Damke托洛洛山美洲际天文台/NSF NOIRLab, Casilla 603, La Serena, 智利 guillermo.damke@noirlab.edu

Aleksandra Ciprijanovic费米国家加速器实验室，邮政信箱 500，巴达维亚，伊利诺伊州 60510，美国 天文学与天体物理学系，芝加哥大学，芝加哥南埃利斯大道 5640 号，伊利诺伊州 60637 美国国家科学基金会-西蒙斯天空人工智能研究所（SkAI），芝加哥东栗树街 172 号，伊利诺伊州 60611，aleksand@fnal.gov

arXiv:2611.32963v1 [astro-ph.IM] 24 Nov 2026

摘要

地基天文台的科学效率取决于能否在动态、多时间尺度且科学奖励复杂的环境中有效决策。传统观测调度器由领域专家设计和调优，借助基于环境和观测系统模型的启发式规则来处理这些因素。在此，我们利用历史数据，通过行为克隆（BC）训练自动望远镜调度器。该行为克隆调度器完全使用历史数据集进行训练；训练时既未提供专家调度启发式规则，也未提供最优科学奖励。随后，我们按照真实程度由低到高的顺序对其进行评估：首先在专家状态上评估，然后进行闭环离线展开，最后在美国国家科学基金会的维克托·M·布兰科4米望远镜上，借助实时反馈进行自主运行。

第九届物理科学机器学习研讨会（ML4PS），2026。

在维克托·M·布兰科4米望远镜上。对于留出夜晚的专家状态，行为克隆策略在气团质量和月球角距方面与专家分布高度吻合。在同一批夜晚进行闭环滚动评估时，我们发现，每次观测的数据质量指标 t_{eff} 相较于专家方案下降了 $6.9 \pm 3.4\%$。在天文台现场部署期间，该策略在 5.2 小时内安排了 149 次曝光，平均$t_{\mathrm{eff}} = 0.75 \pm 0.26$，与闭环展开评估结果一致。过渡到真实观测条件后，未发现额外性能下降。这是领导级国家天文台首次在真实世界中实时部署学习型自主调度器，也为评估基于历史调度方案并辅以反事实轨迹训练的基于强化学习的调度器开辟了道路。

1 引言

天文巡天调度，即决定望远镜观测什么以及何时进行观测，是部分随机环境中的高维、多目标序贯决策问题。该问题类似于著名的旅行商问题——一种已知为NP困难的组合优化问题 [15]。天文巡天调度还受到不断变化的环境（天气、天空亮度、目标可用性）、高优先级中断以及长期巡天目标的进一步制约，这些因素共同影响分配给每项决策的价值。此外，由于每项决策都会改变后续可行观测的集合，局部有利的决策并不一定能产生全局最优解 [22]。传统望远镜调度器[6, 7, 9, 20] 通过反复调节、基于环境和观测系统模型的启发式规则与优先级来应对这些挑战。它们通常会结合人类专家经验、物理模型，并对短期和长期科学目标进行明确加权。近期天文巡天提供了大量历史数据，记录了真实观测条件下的调度决策及其后果，为构建学习型调度器奠定了基础，从而减少对计算成本高昂的模拟和人类专家经验的依赖。

天文学调度中的机器学习（ML）方法越来越多地探索使用模拟训练环境，包括利用强化学习（RL）进行资源分配，以及在望远镜阵列之间开展协同调度 [27], 目标机会观测和后续观测调度 [25, 28], 以及自主目标选择 [4, 23, 26]。对于时空遗产巡天（LSST） [14], , Naghib等人 [17]探索了强化学习（RL）和全局优化策略，用于学习一组专家设计的基函数的最优权重。尽管这些研究展现了机器学习驱动调度的潜力，但它们并未同时处理本文所关注的大规模动作空间、较长的巡天时域、实时观测反馈以及迭代式天文台现场部署问题。

在本研究中，我们训练了一个模型，从由专家设计的算法obstac生成并执行的历史调度数据中直接学习调度行为 [20]。这些数据包含六年的暗能量巡天（DES； [5]）观测，观测使用美国国家科学基金会（NSF）的维克托·M·布兰科4米望远镜上的暗能量相机（DECam；[10]）获得，望远镜位于智利的托洛洛山美洲际天文台。该方法将我们的算法锚定于真实观测条件之中，减少了对完全合成且计算成本高昂的训练环境的依赖，但也将覆盖范围限制在历史数据中实际出现的轨迹上。我们使用行为克隆（BC） [2] 来评估仅凭模仿能够恢复哪些行为、在哪些情况下会失效，以及其性能在转入天文台现场部署时是否会发生变化。我们在三种日益接近真实情况的条件下评估调度器：（1）在训练期间未使用的专家状态上进行单步评估；（2）在留出的测试夜上进行闭环策略滚动评估，其中每次策略决策都会决定下一个模拟环境状态；（3）在布兰科望远镜上进行实时调度。为此，我们开发了一个模块化且与策略无关的框架，将学习型调度器连接至实时遥测数据和布兰科望远镜控制系统。

2 数据与方法

训练与评估数据。 该数据集汇集了 DES 六年的档案宽视场观测，覆盖南天 $\sim$5,000 度^2，历时 758 个夜晚，使用了 $\sim$1,700 个目标视场。每个视场在五种宽带滤光片（*grizY*）下各观测 $\sim$10 次，并采用预先定义的抖动策略，使每个视场的目标指向偏移 [1, 8]。运行元数据汇集自

山顶遥测数据库 [12]。数据集中剔除了仅包含时域观测（其调度策略不同于宽视场观测）的夜晚，最终剩余 689 个夜晚。由于每晚的状态–动作对高度相关，因此数据按夜晚随机划分，训练/验证/测试划分比例为 80%/10%/10%，对应 $553/68/68$ 个夜晚，共包含 $63{,}824/8{,}417/7{,}978$ 个状态–动作对。[1]

状态与动作空间。 每个动作 $a \in A$ ($|A| = 30{,}400$) 由一对 $(pointing, filter)$ 组成，其中 $pointing$ 是地平坐标系中的天球坐标（即方位角/高度角），而$filter$ 是五种 DES 宽带滤光片之一（$grizY$）。指向被离散化为 $N_{\mathrm{side}} = 32$个 HEALPix 像素 [11],[2]，我们称之为“分箱”；动作空间则限定为位于地平线以上的 6,080 个分箱。通过使用地平坐标而非赤道坐标，策略能够推广至 DES 覆盖区域之外的目标。状态向量 $s \in S$包含当前环境特征：望远镜的指向和滤光片、太阳和月球的位置、由 Skybright 预测的各波段天空亮度 [18], 、根据前 15 分钟内拍摄的曝光所测得的点扩散函数半高全宽预测的视宁度，以及各波段的巡天进度。状态还包括描述每个候选动作的逐分箱特征：气团质量、相对于当前指向的转向距离、与太阳和月球的距离、分箱落至地平线以下前的时间，以及局部巡天完成比例。

策略建模。 策略 $\pi_\theta(a|s)$ 是动作上的条件概率分布。我们使用共享权重多层感知器（MLP）对策略进行参数化。全局状态被编码为 128 维表示，并与每个候选分箱的特征拼接，然后输入由 128 维隐藏层、LeakyReLU 激活函数和层归一化组成的共享权重三层多层感知器。网络为每个候选动作输出一个分数，并在将分数向量传入 softmax层之前，对无效候选动作的分数进行掩码处理。选择概率最高的动作（即指向–滤光片对） $a^* = (b^*,\ f^*)$。不包含未完成视场的无效指向–滤光片组合会被掩码处理，从而形成状态依赖动作空间，即 $A(s) \subseteq A$。选择动作后，如果所选动作中包含多个目标视场，则仅保留需要所选动作关联的滤光片 f^* 的视场；随后，我们使用 SciPy 的 RBF插值器和五次核 [24]，将该滤光片对应的网络分数从 HEALPix 分箱中心插值到候选视场位置。调度器随后选择分数最高的视场。

训练。该策略采用行为克隆 [2], 这一最大似然监督学习技术进行训练。使用档案数据拟合策略 π_θ 的多层感知器参数 θ。由于我们的动作空间是离散的，因此我们最小化专家动作与策略之间的交叉熵 $\pi_\theta(a|s)$，其中 π_θ 在可行动作集合 $A(s) \subseteq A$上归一化。我们使用 Adam 进行训练，批量大小为 1024，初始学习率为 10^{-3}，采用余弦退火调度器，并根据验证损失进行早停。模型检查点根据验证集划分选取。在 H100 GPU[3]上，训练耗时 $\sim$30 分钟。

3 结果与讨论

作为数据质量的汇总指标，我们使用有效曝光时间比例因子，

$$t_{\mathrm{eff}} = \eta^2 \left(\frac{0.9''}{\mathrm{FWHM}} \right)^2 \left(\frac{b_{\mathrm{dark}}}{b} \right) \tag{1}$$

其中 η 为大气透射率（到达望远镜的光线比例），FWHM 为点扩散函数的半高全宽（即视宁度），b 和 b_{dark} 分别表示观测到的和名义参考暗夜天空亮度 [19, 20]。曝光的 t_{eff} 描述了相较于名义观测条件下预期灵敏度所实现的灵敏度（例如， t_{eff} 提升 10% 相当于运行效率提升 $\sim$10%）。

专家状态评估。 在基于专家留出状态的单步预测中，学习策略有 61.4% 的时间匹配专家的滤光片选择，约为 3$\times$ 跨五个滤光片进行均匀选择的更高比例。天空亮度和月球高度角的逐滤光片边际分布（这两个因素是 DES 调度器算法确定滤光片选择的关键因素）都与专家的分布高度一致

[1]训练和评估代码可在 https://github.com/rachelhur/skyway 获取，分析和绘图脚本可在 [13] 获取。[2] http://healpix.sourceforge.net

（详见附录 A）。学习策略有 24.3% 的时间复现专家的指向分箱，$\sim 10^3\times$ 比跨分箱均匀选择的频率更高；在 81.3% 的状态转移中，其落点与专家视场的距离在 5 度以内，月球角距中位数为 2.08 度。在相同状态转移上取平均，学习策略与专家的气团质量（1.189 对 1.190）、月球角距（92.7 度对 92.8 度）、t_{eff}（0.672 对 0.672）一致，并且每次曝光的转向显著更少（1.65 度对 4.72 度）。在相似的观测条件下，DES 调度器经常交替选择 g 波段和 r 波段滤光片，以及 z 波段和 Y 波段滤光片。指向选择中也存在类似的退化现象，即邻域内的多个分箱被认为同样有利。因此，即使是训练良好的调度器，我们也不期待其能够准确复现 DES 的动作。

离线闭环行为。学习策略在全部 68 个测试夜上进行展开，每晚从第一个状态开始；其中一个夜晚的专家曝光数量过少，因此未纳入以下比较。由于模拟观测计划没有实测数据质量，我们根据当晚测得的条件，估算专家观测计划和离线闭环展开观测计划的 t_{eff}（详情见附录 C 和 [19, 20]）。所有测试夜的两种观测计划都在夜晚早些时候出现分歧，之后从未重新同步，因此无法直接逐次比较曝光。我们因此按夜晚进行比较，并使用刀切重采样估算不确定性。平均而言，学习型调度器的单次曝光 t_{eff} 比专家低 $6.9 \pm 3.4\%$。这一差距会随着夜晚的推进而扩大（图 1，左）：夜晚开始时，差异与零相符；到夜晚末尾则达到 $\sim 13\%$。这种差距与观测逐渐转向落下目标和更高气团质量相关，而这种现象只出现在闭环滚动评估中。在专家状态上，该策略能非常接近地复现专家的气团质量和时角分布（附录 B）。这一模式表明可能存在累积误差 [21]：部署策略偏好较短转动的倾向在专家状态上已经显现，并可能使调度器持续观测一个正在落下的区域，导致观测气团质量高于专家在同一夜晚所观测到的水平（进一步讨论见附录 B）。

DES 调度器会优先考虑累积的 t_{eff}，同时也注重随时间推移在巡天覆盖区域内实现均匀覆盖。我们在图 1（中间面板）中，将 DES 专家与学习策略在测试夜上的闭环滚动评估，以及非学习基线启发式进行比较。为衡量深度和均匀度，我们将覆盖区域内每个视场 i 在所有测试夜曝光中的 t_{eff} 求和，$T_{\mathrm{eff},\,i} = \sum t_{\mathrm{eff}}$，并将未观测的视场设为 $T_{\mathrm{eff},\,i} = 0$。纵轴为所有视场的均值 $\langle T_{\mathrm{eff}} \rangle$，横轴为相对离散度 $\sigma(T_{\mathrm{eff}})/\langle T_{\mathrm{eff}} \rangle$，其中 $\sigma(T_{\mathrm{eff}})$ 是 $T_{\mathrm{eff},\,i}$ 在各视场间的标准差。相对离散度越低，天空覆盖就越均匀，因此较好的观测计划位于左上方。“随机”启发式会随机选择可观测视场，而“最小转向”启发式则只选择最近的未完成视场。这些基线代表了在优化 $\langle T_{\mathrm{eff}} \rangle$ 与均匀覆盖天空之间权衡的两个极端。行为克隆调度器累积的 $\langle T_{\mathrm{eff}} \rangle$ 比随机选择多 $\sim$50%，且对巡天覆盖区域的覆盖远比最小转向启发式均匀。该调度器累积的 $\langle T_{\mathrm{eff}} \rangle$ 与专家相当，但在将观测分散到整个覆盖区域方面逊于专家。

天文台现场部署。学习策略用于在2026年6月24日后半夜自主控制布兰科望远镜，从协调世界时4:48运行至10:00，共计 ~ 5.2 小时。它在 $griz$ 中执行了149次曝光，并产生了平均 $t_{\mathrm{eff}} = 0.75$，标准差为0。26。调度器可以执行长距离转动以及转向高气团质量区域，但始终选择具有较高有效曝光时间的合理目标。为了比较这是否不同于预期的离线评估结果，我们使用相同的检查点、候选视场和实测视宁度轨迹，对同一测试夜进行离线回放。两次运行观测到的视场数量相同，且离线回放观测到了天文台现场部署所观测视场—滤光片组合的92%。离线回放和天文台现场部署的每次曝光平均 t_{eff} 均为0。75。对于这一单独的半夜观测，从离线回放转向天文台现场部署时，我们未检测到额外的性能退化，这表明离线评估与实天性能一致。这种一致性符合以下解释：测试夜中的闭环性能退化反映的是专家克隆保真度下降，而非模拟器误差。

局限性。每个分箱的大小大致相当于 DECam 视场的大小（$\sim$3 度^2），因此我们的候选动作不包含相邻视场的信息，而这些信息对于学习最优铺砖策略是必要的。未来的工作将利用图神经网络等架构或多层分箱方案来编码有关邻域的汇总统计量，从而纳入这些信息。有效曝光时间比例因子 t_{eff} 并不能完整描述调度器的科学效用；深度均匀性和滤光片平衡等其他因素同样重要，今后也将进一步加以考虑。最后，天文台现场部署仅覆盖一个夜晚

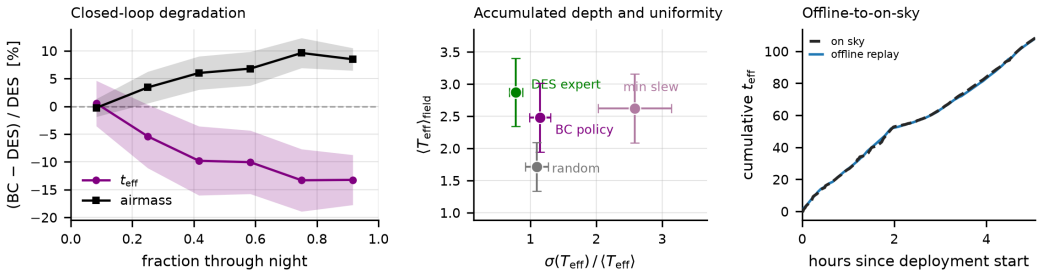

图 1：（左）在留出夜晚中，闭环展开期间学习得到的行为克隆调度器与 DES 调度器之间平均 $t_{\rm eff}$（紫色）和平均气团质量（黑色）的百分比差异。平均值是在每晚分别按分箱计算后，再对各夜晚取平均。色带表示对测试夜晚进行逐晚删除刀切法得到的 95%区间。（中）DES 调度器、行为克隆调度器以及两种启发式方法（均匀随机和最小转向距离）的累积深度与均匀度之间的关系。 $T_{\rm eff}$是所有留出夜晚中每个视场的曝光总和。纵轴表示其在覆盖区域视场中的平均值，横轴表示各视场之间的标准差除以该平均值，因此数值越小，整个覆盖区域的深度越均匀。未访问的视场计为 $T_{\rm eff} = 0$。误差条表示对测试夜晚进行刀切重采样得到的 95%区间。（右）自部署开始经过的小时数与所有先前观测的累积 $t_{\rm eff}$ 之间的关系。天文台现场部署结果（黑色虚线）与估计的离线回放结果（蓝色实线）高度一致。

且当晚的天气状况在该季节中尤其良好。要评估其在不同夜晚、季节和天气条件下的稳健性，还需要在这些条件下进行测试，并增加天文台现场部署次数。不过，我们发现，对于这次特定的部署夜晚，模拟器和 $t_{\rm eff}$ 估计器能够可靠地预测实天性能。

4 结论与未来工作

我们分析了一个使用六年 DES观测通过行为克隆训练的调度器，并逐步在更贴近现实的条件下进行评估。基于专家状态的单步预测，能够很好地复现归档调度在滤光片选择、气团质量、月球角距和 $t_{\rm eff}$方面的分布，但预测的指向间平均转向距离较短。在闭环滚动评估中，随着策略逐渐偏向更高的气团质量，平均 $t_{\rm eff}$ 比专家低 6.9%。初步实天性能与离线预期一致，表明在这段 5.2 小时、天气良好的观测时段内，仿真到现实差距很小。

这一真实数据基线旨在支持正在开展的工作：使用经反事实轨迹增强的归档观测夜，通过强化学习训练调度器；这些轨迹的状态和奖励可利用经真实数据验证的解析投影进行估计。通过偏离专家推导的轨迹、但仍与之接近的轨迹进行学习，可以避免在庞大的动作空间中盲目探索；强化学习还应通过使用策略自身决策所产生的状态训练策略，减少此处观察到的闭环漂移。将这些未来策略与行为克隆基线进行比较，有助于我们评估引入反事实经验后性能如何变化，包括离线和实天表现。

致谢与资金披露

我们衷心感谢 NSF-西蒙斯天空人工智能研究所（SkAI）通过 NSF AST-2421845 和西蒙斯基金会 MPS-AI-00010513 提供的支持。本研究得到了芝加哥大学人工智能倡议的资金支持。本研究还部分得到了美国能源部科学办公室高能物理办公室通过创世纪计划提供的资金支持。

本研究使用了 DeltaAI 先进计算与数据资源，该资源由国家科学基金会（资助编号 OAC 2320345）和伊利诺伊州提供支持；还使用了 Delta 先进计算与数据资源，该资源由国家科学基金会（资助编号 OAC

2005572）和伊利诺伊州提供支持。Delta 和 DeltaAI 是伊利诺伊大学厄巴纳-香槟分校及其国家超级计算应用中心的联合项目。
本研究由费米前沿探索集团有限责任公司根据与美国能源部科学办公室高能物理办公室签订的第 89243024CSC000002 号合同完成。出版方确认，美国政府依据能源部公共访问计划授予其提供公共访问的许可。（http://energy.gov/downloads/doe-public-access-plan）。

本项目使用了暗能量巡天（DES）的公开档案数据。DES 项目的经费由美国能源部、美国国家科学基金会、西班牙科学与教育部、英国科学技术设施委员会、英格兰高等教育拨款委员会、伊利诺伊大学厄巴纳-香槟分校国家超级计算应用中心、芝加哥大学卡弗里宇宙学物理研究所、俄亥俄州立大学宇宙学与天体粒子物理中心、德州农工大学米切尔基础物理与天文学研究所、巴西研究与项目融资机构、里约热内卢州卡洛斯·查加斯·菲略研究支持基金会、国家科学技术发展委员会和科学、技术与创新部、德国研究联合会以及暗能量巡天合作机构提供。

合作机构包括阿贡国家实验室、加利福尼亚大学圣克鲁兹分校、剑桥大学、马德里能源、环境与技术研究中心、芝加哥大学、伦敦大学学院、DES-巴西联盟、爱丁堡大学、苏黎世联邦理工学院、费米国家加速器实验室、伊利诺伊大学厄巴纳-香槟分校、空间科学研究所、高能物理研究所、劳伦斯伯克利国家实验室、慕尼黑大学及其关联的卓越集群“宇宙”、密歇根大学、国家光学天文台、诺丁汉大学、俄亥俄州立大学、OzDES成员联盟、宾夕法尼亚大学、朴次茅斯大学、SLAC国家加速器实验室、斯坦福大学、萨塞克斯大学和德州农工大学。

部分依据在托洛洛山美洲际天文台的观测。该天文台隶属国家光学天文台，由天文学研究大学联盟（AURA）根据与国家科学基金会签订的合作协议运营。

本文部分结果使用 healpy 和 HEALPix 软件包推导得出。

参考文献

[1] T. M. C. Abbott 等。《暗能量巡天数据发布2》。《天体物理学杂志增刊》，255(2):20，2021年8月。doi: 10.3847/1538-4365/ac00b3。[2] Michael Bain 和 Claude Sammut。《行为克隆框架》。载于《机器智能 *15*，智能代理》 *[*牛津圣凯瑟琳学院，牛津，*7*月 *1995]*, 第103–129页，GBR，1999年。牛津大学。ISBN 0198538677。[3] Brett Bode、Gregory Bauer、Laura Herriott、Volodymyr Kindratenko 和 William Gropp。《DeltaAI：人工智能/机器学习研究国家资源》。载于《天文高级软件与控制 *II 2025*：协作的力量》，PEARC ’25，美国纽约州纽约市，2025年。计算机协会。ISBN 9798400713989。doi: 10.1145/3708035. 3736062。URL https://doi.org/10.1145/3708035.3736062。[4] Hai Cao、Shaoming Hu、Junju Du、Xu Chen、Shuqi Liu、Shuai Feng、Bo Zhang 和 Yuchen Jiang。《利用深度强化学习高效调度地基天文观测》。《天文杂志》，170(2):88，2025年8月。doi: 10.3847/1538-3881/ade3dc。[5] 暗能量巡天合作组、T. Abbott 等。《暗能量巡天：不止于暗能量——概述》。《皇家天文学会月刊》，460(2):1270–1299，2016年8月。doi: 10.1093/mnras/stw641。[6] Francisco Delgado 和 Michael A. Reuter。《LSST调度器：从设计到建造》。载于 Alison B. Peck、Robert L. Seaman 和 Chris R. Benn（编），《天文台运行：战略、流程与系统 *VI*》，第9910卷国际光学工程学会（*SPIE*）会议丛书，第991013页，2016年7月。doi: 10.1117/12.2233630。

[7] Arjun Dey 等。《DESI遗产成像巡天概述》。《天文杂志》，157(5):168，2019年5月。doi: 10.3847/1538-3881/ab089d。[8] H. T. Diehl。《暗能量巡天及其运行：第六年——终章》。技术报告 FERMILAB-TM-2720-AE，SLAC国家加速器实验室（SLAC），美国加利福尼亚州门洛帕克；费米国家加速器实验室（FNAL），美国伊利诺伊州巴达维亚，2020年1月。URL https://www.osti.gov/biblio/1596042。[9] A. Drlica-Wagner 等。《DECam本地体积探索巡天：概述与首次数据发布》。《天体物理学杂志增刊》，256(1):2，2021年9月。doi: 10.3847/1538-4365/ac079d。[10] B. Flaugher 等。《暗能量相机》。《天文杂志》，150(5):150，2015年11月。doi: 10.1088/ 0004-6256/150/5/150。[11] K. M. Górski、E. Hivon、A. J. Banday、B. D. Wandelt、F. K. Hansen、M. Reinecke 和 M. Bartelmann。《HEALPix：球面分布数据的高分辨率离散化与快速分析框架》。《天体物理学杂志》，622:759–771，2005年4月。doi: 10.1086/427976。[12] K. Honscheid、T. Abbott、J. Annis、E. Buckley-Geer、F. Castander、J. Eiting、M. Gladders、M. Haney、I. Karliner、D. Kau、K. Kuehn、S. Kuhlmann、T. Qian、M. Selen、J. Thaler、D. Tucker 和 A. Zhao。《DES相机的读出与控制系统（SISPI）》。载于 Alan Bridger 和 Nicole M. Radziwill（编），《天文高级软件与控制 *II*》，第7019卷国际光学工程学会（*SPIE*）会议丛书，第701911页，2008年8月。doi: 10.1117/12.787926。[13] Rachel Hur、Paul Chichura、Alex Drlica-Wagner、Eric Neilsen、Guillermo Damke 和 Aleksandra Ciprijanovic。《“从历史观测计划到在轨控制：4米布兰科望远镜上的自主调度”的补充分析代码、模型检查点、闭环推演数据和部署数据》，2026年11月。URL https://doi.org/10.5281/zenodo. 24401148。[14] Željko Ivezi´c 等。《LSST：从科学驱动因素到参考设计及预期数据产品》。《天体物理学杂志》，873(2):111，2019年3月。doi: 10.3847/1538-4357/ab042c。[15] Bernhard Korte 和 Jens Vygen。《组合优化：理论与算法》，《算法与组合学》第21卷。施普林格，第6版，2018年。ISBN 978-3-662-56039- 6。doi: 10.1007/978-3-662-56039-6。URL https://link.springer.com/book/10.1007/ 978-3-662-56039-6。[16] E. Morganson、R. A. Gruendl、F. Menanteau、M. Carrasco Kind、Y.-C. Chen、G. Daues、A. Drlica-Wagner、D. N. Friedel、M. Gower、M. W. G. Johnson、M. D. Johnson、R. Kessler、F. Paz-Chinchón、D. Petravick、C. Pond、B. Yanny、S. Allam、R. Armstrong、W. Barkhouse、K. Bechtol、A. Benoit-Lévy、G. M. Bernstein、E. Bertin、E. Buckley-Geer、R. Covarrubias、S. Desai、H. T. Diehl、D. A. Goldstein、D. Gruen、T. S. Li、H. Lin、J. Marriner、J. J. Mohr、E. Neilsen、C.-C. Ngeow、K. Paech、E. S. Rykoff、M. Sako、I. Sevilla-Noarbe、E. Sheldon、F. Sobreira、D. L. Tucker，以及 W. and Wester。《暗能量巡天图像处理流水线》。《太平洋天文学会刊》，130(989):074501，2018年5月。ISSN 1538-3873。doi: 10.1088/1538-3873/aab4ef。URL http://dx.doi.org/10.1088/1538-3873/aab4ef。[17] Elahesadat Naghib、Peter Yoachim、Robert J. Vanderbei、Andrew J. Connolly 和 R. Lynne Jones。《望远镜调度器框架：及其在大面积同步巡天望远镜中的应用》。《天文杂志》，157(4):151，2019年4月。ISSN 0004-6256, 1538-3881。doi: 10.3847/1538-3881/ aafece。URL http://arxiv.org/abs/1810.04815。arXiv:1810.04815 [astro-ph].[18] Eric Neilsen。《Skybright》，2019年1月。

[19] Jr. Neilsen、Eric、Gary Bernstein、Robert Gruendl 和 Stephen Kent。《暗能量巡天第一年中的极限星等、τ、 t_{eff}和图像质量》。技术报告 FERMILAB-TM-2610-AE-CD，费米国家加速器实验室（FNAL），美国伊利诺伊州巴达维亚，2016年3月。网址：https://www.osti.gov/biblio/1250877。

[20] Jr. Neilsen、Eric、James T. Annis、H. Thomas Diehl、Molly E.C. Swanson、Chris D’Andrea、Stephen Kent 和 Alex Drlica-Wagner。《暗能量巡天的观测策略、战术与曝光调度器》。技术报告 FERMILAB-TM-2714-AE-CD-PPD，科学与技术信息办公室（OSTI）、费米国家加速器实验室，2019年11月。网址：http://dx.doi.org/10.2172/1574836。[21] Stephane Ross、Geoffrey Gordon 和 Drew Bagnell。《将模仿学习和结构化预测转化为无遗憾在线学习》。载于 Geoffrey Gordon、David Dunson 和 Miroslav Dudík 编，第十四届人工智能与统计学国际会议论文集，机器学习研究论文集第15卷，第627–635页，美国佛罗里达州劳德代尔堡，2011年4月11–13日。PMLR。网址：https://proceedings. mlr.press/v15/ross11a.html。[22] Daniel Rothchild、Christopher Stubbs 和 Peter Yoachim。《ALTSched：利用 LSST 改进时域科学的调度》。*PASP*，131(1005):115002，2019年11月。doi: 10.1088/1538-3873/ab3300。[23] Franco Terranova、M. Voetberg、Brian Nord 和 Amanda Pagul。《自动驾驶望远镜：利用离线强化学习自主调度天文观测活动》。*arXiv* 电子预印本，文章编号 arXiv:2311.18094，2023年11月。doi: 10.48550/arXiv.2311. 18094。[24] Pauli Virtanen、Ralf Gommers、Travis E. Oliphant、Matt Haberland、Tyler Reddy、David Cournapeau、Evgeni Burovski、Pearu Peterson、Warren Weckesser、Jonathan Bright、Stéfan J. van der Walt、Matthew Brett、Joshua Wilson、K. Jarrod Millman、Nikolay Mayorov、Andrew R. J. Nelson、Eric Jones、Robert Kern、Eric Larson、C J Carey、 ̇Ilhan Polat、Yu Feng、Eric W. Moore、Jake VanderPlas、Denis Laxalde、Josef Perktold、Robert Cimrman、Ian Henriksen、E. A. Quintero、Charles R. Harris、Anne M. Archibald、Antônio H. Ribeiro、Fabian Pedregosa、Paul van Mulbregt 和 SciPy 1.0 贡献者。《SciPy 1.0：Python科学计算基础算法》。自然·方法，17:261–272，2020年。doi: 10.1038/s41592-019-0686-2。[25] Xuhang Wang 和 Haiyan Wu。《面向高频动态到达的天文卫星机会目标任务调度方法》。中国空间科学学报，45:1376，2025年1月。doi: 10.11728/cjss2025.05.2024-0125。[26] Alex Zhang、Jackson Craig 和 Lara Waldrop。《在 NASA Carruthers地冕观测站任务中，利用深度强化学习实现快速长时域运行调度》。国际自动规划与调度会议论文集，36:538–546，2026年6月。doi: 10.1609/icaps. v36i1.42871。[27] Yajie Zhang、Ce Yu、Chao Sun、Yi Hu、Zhaohui Shang、Jizeng Wei 和 Xu Yang。《GRRIS：分布式巡天望远镜阵列的实时站内观测调度方案》。《天文杂志》，168(5):214，2024年10月。doi: 10.3847/1538-3881/ad77ab。网址：https://doi.org/10.3847/1538-3881/ad77ab。[28] Yajie Zhang、Ce Yu、Chao Sun、Jizeng Wei、Junhan Ju 和 Shanjiang Tang。《解决天文后续观测中的在线资源约束调度问题：一种强化学习方法》。未来一代计算机系统，169(C)，2025年8月。ISSN 0167-739X。doi: 10.1016/j.future.2025.107781。网址：https://doi.org/10.1016/j.future.2025.107781。

单步专家状态评估

我们使用月球高度角的边缘分布来评估滤光片选择保真度。月球高度角是影响 DES 调度器滤光片选择的一个重要因素（图 2 左图）。对于每种滤光片，在分位数和分布整体形状方面，两种调度器的月球高度角分布都较为相似。在图 2 右图中，分歧主要集中在 g/r 和 z/Y 这两组滤光片上，DES 调度器通常会交替选择其中任意一种滤光片。

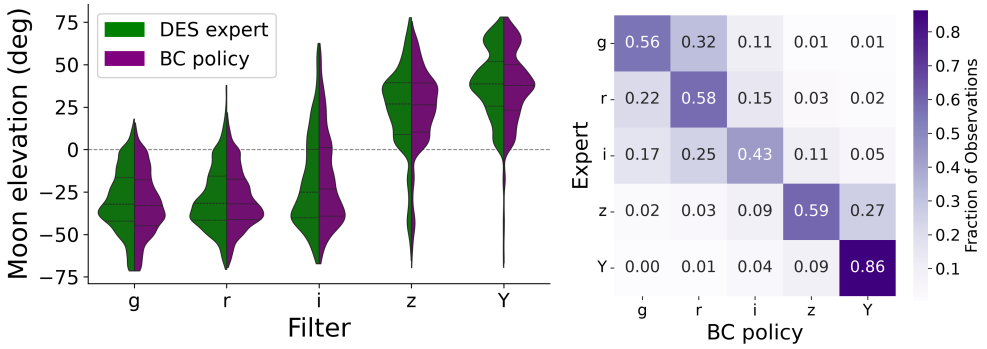

图 2：留出专家状态下的滤光片选择。（左）每次状态转移时的月球高度角小提琴图，按所选滤光片划分；绿色表示专家 DES 调度器，紫色表示行为克隆策略。（右）相同状态下行为克隆策略与 DES 调度器的滤光片选择对比，按行归一化。分歧主要集中在 g/r 和 z/Y 这两组滤光片上，DES 调度器通常会交替选择其中任意一种滤光片。

由于天空指向分布在 6,080 个分箱中，且许多分箱具有相同的取值，因此我们不依赖前 k 项一致性，而是通过转向距离、时角、气团质量和月球角距的边缘分布（图 3 顶部和中部各行）以及 t_{eff}（图 3 底部）来评估指向选择保真度。对于时角、气团质量、月球角距和 t_{eff}，行为克隆策略（黑色虚线）与 DES 专家（绿色）高度匹配。然而，对于转向距离，尽管行为克隆策略和 DES 专家的大多数曝光都集中在小于 $\sim$2 度的转向距离上，且累积分布的整体形状相似，但 DES 调度器会将曝光分布在更大的转向距离范围内。

B 闭环离线展开

协议。68 个测试夜均从第一个专家状态开始展开，即从 DES 调度器执行的第一次曝光开始。历史数据包含因天气和其他科学项目导致的中断，而我们的环境模拟器并未对这些情况进行建模。我们将这些中断识别为持续时间为> 5 分钟的间隔（布兰科望远镜的最长转动时间为 100 度转动时的 $\sim$4 分钟）。为避免对天气相关图像质量的比较造成偏差，我们移除了这些时段内行为克隆调度器的观测。

闭环漂移。我们考察专家状态上存在的逐步偏差在闭环展开过程中如何演变。行为克隆调度器在闭环运行中的转向分布，与其在单步专家状态上评估时的分布几乎相同，但二者进行的长距离转动都远少于专家（图 3，左上）。另一方面，其气团质量分布在单步预测中遵循专家的分布，但在闭环展开中并非如此（图 3，左中）。策略的气团质量过量和 t_{eff}不足都从接近零开始，随着夜间观测逐渐增大，到夜末分别达到 $\sim$9% 和 $\sim$13%（图 1，左），而闭环 t_{eff} 分布整体发生了偏移，而不仅仅是尾部发生变化（图 3，底部）。在专家状态下，策略复现了专家的时角分布；但在闭环运行中，它转向落下目标（图 3，右上）。综合来看，这些结果与累积误差相符：由于很少进行长距离转动，策略可能无法随着其目标区域落下而重新定位，从而使其处于比专家在同一夜晚观测到的目标更高的气团质量下。初步检查表明，始终选择概率最高的动作会放大这一偏差（图 4）。在专家状态下，由策略自身概率所隐含的转向分布与专家的分布高度一致，而通过选择概率最高的动作所执行的转动则要短得多。尽管策略为长距离转动分配了不可忽略的概率，但它无法确定确切的动作，因此，a

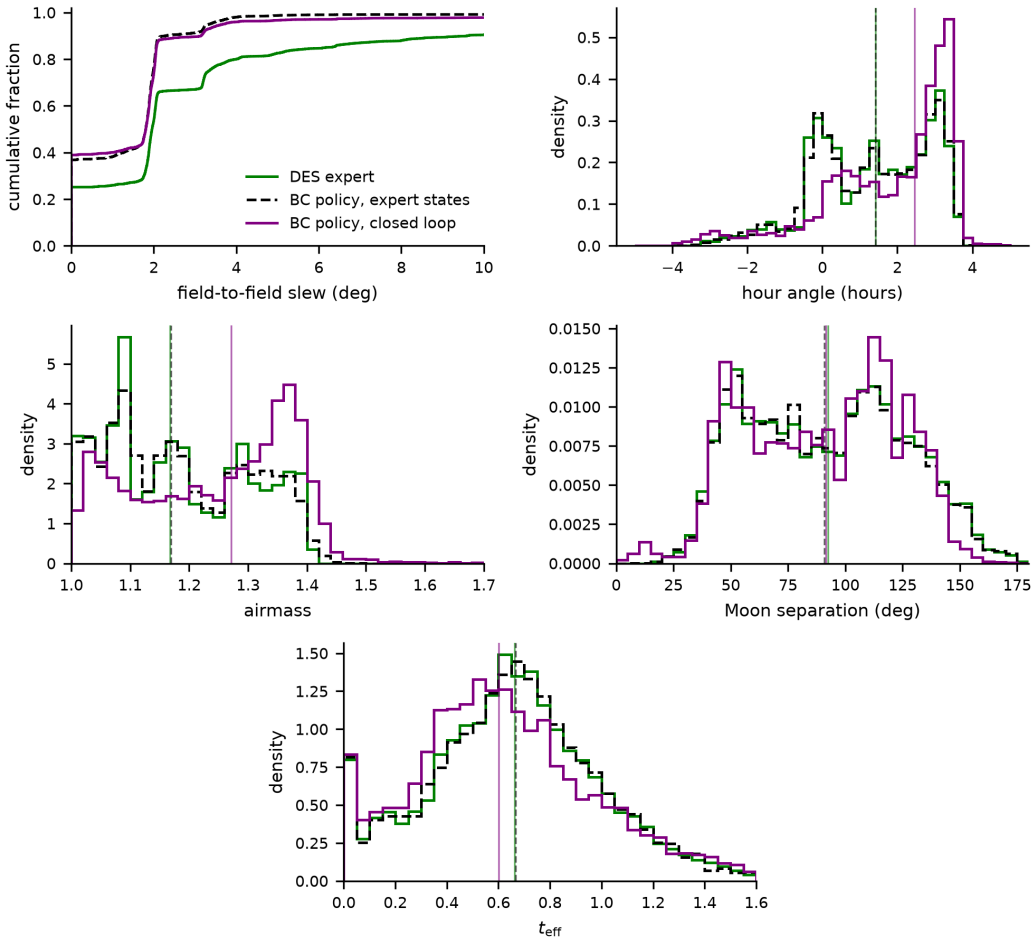

图 3：在留出夜晚中，DES 调度器（绿色）和行为克隆调度器在专家状态（黑色虚线）及闭环展开（紫色实线）中的指向选择，以及它们所实现的 $t_{\rm eff}$。（左上）视场间转向距离的累积分布。（右上）每次曝光的时角（负值为东向、上升；正值为西向、下降）。（左中）气团质量。（右中）月球角距。（底部）估计的 $t_{\rm eff}$（附录 C）。竖线标示每个分布的中位数。

相邻分箱几乎总是唯一概率最高的选择。此外，这一概率分散在许多相距较远的分箱中，进一步加剧了对较长距离转动的影响。

基线构建。这两种非学习启发式方法仅获得当晚的观测窗口以及每个视场所需的剩余观测次数。在每个决策步骤中，它们从*grizY* 中随机抽取一个滤光片，并从仍需在该滤光片下进行观测且气团质量低于 1.8 的视场中进行选择。“随机”启发式方法假设所有视场服从均匀分布并据此抽取一个视场，而“最小转向”则选择距离上一个指向最近的视场。每种启发式方法都使用与闭环调度器相同的模拟器，在整个观测夜晚进行演化。

C 估算有效温度 $t_{\rm eff}$

历史调度方案记录了决定有效曝光时间比例因子（公式 1）以及 $t_{\rm eff}$ 本身的各项测量值。为了在留出夜晚直接比较学习型调度器与专家的表现，我们使用归档的数据质量测量值、一组解析表达式[19, 20], 以及一个拟合天空亮度超额的模型 [20]，为历史调度方案和学习型调度器重建 $t_{\rm eff}$。对于 $t_{\rm eff}$ 的每个组成部分（视宁度、天空亮度和云层衰减），我们根据当晚所有

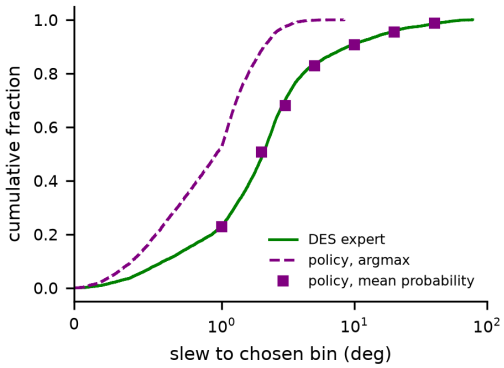

图4：专家状态下转向的累积分布，测量的是从当前指向到所选天空分箱中心的距离，分别对应DESexpert（绿色）和选择其最可能动作（最大值参数；紫色虚线）的 BC 策略。紫色方块表示由 BC 策略的动作概率预测的累积转向分布，该分布在所有专家状态上取平均，并在选定的转向距离处进行评估。

DECam 曝光的测量值对观测条件随时间进行插值，然后在每次计划曝光的时间、指向和滤光片处进行评估。档案曝光采用相同的处理方式，但使用留一测量。

天空亮度 由 skybright模型重建， $\Delta\mu_{\texttt{skybright}}$ 加上当晚曝光的实测值减去模型值的残差，并在时间上进行插值， $\Delta\mu_r$该残差吸收了模型与 DECam 测量值之间缓慢漂移的偏移量。当条件允许时，残差按滤光片使用一小时内相同滤光片曝光的数据进行插值；否则使用所有滤光片的数据，并拟合每个滤光片的偏移量

$$\Delta\mu(t,\alpha,\delta,f) = \Delta\mu_{\texttt{skybright}}(t,\alpha,\delta,f) + \Delta\mu_r(t,f) \tag{2}$$

其中， t 表示时间， α， δ 分别表示赤经和赤纬天球坐标， f 表示滤光片。随后即可计算天空亮度因子：

$$\frac{b_{\mathrm{dark}}}{b} = 10^{-0.4\,\Delta\mu} \tag{3}$$

视宁度 重建遵循 [20] 中所述的解析表达式。首先，从每次曝光测得的半高全宽中以平方和方式移除仪器项 $0.45''$，并将其投影为i波段的天顶等效值（即观测者正上方）。对于每次计划曝光，将插值得到的数值重新投影到计划使用的波段和气团质量，并与仪器项以平方和方式重新合并。气团质量缩放为 $(X_f/X_i)^{3/5}$，波段缩放为 $(\lambda_f/\lambda_i)^{-0.34}$，其中指数根据 DECam 各波段的比值进行拟合。我们采用 $0.9''$（λ_f/λ_i）$^{-1/5}$的基准半高全宽。

大气透射率遵循 DES约定 [16]。衰减建模为

$$\Delta m = k_f(X-1) + c(t) \tag{4}$$

其中， k_f 是各波段的消光系数， X 是气团质量， $c(t)$是当晚插值得到的云层衰减。低于 0.2 等星阈值的衰减视为晴空

$$\eta^2 = \begin{cases} 1, & \Delta m \le 0.2, \\ 10^{-0.8\,(\Delta m - 0.2)}, & \Delta m > 0.2. \end{cases} \tag{5}$$

由于云层衰减仅在档案中的指向位置测量，因此我们假设云层在整个天空中均匀分布。

验证。我们将留一法估计与测试夜期间测得的 t_{eff} 进行比较（图 5，左），发现该估计存在 $\Delta t_{\mathrm{eff}} = -0.003$的偏差；与其归一化中位数绝对偏差相比，这一偏差很小

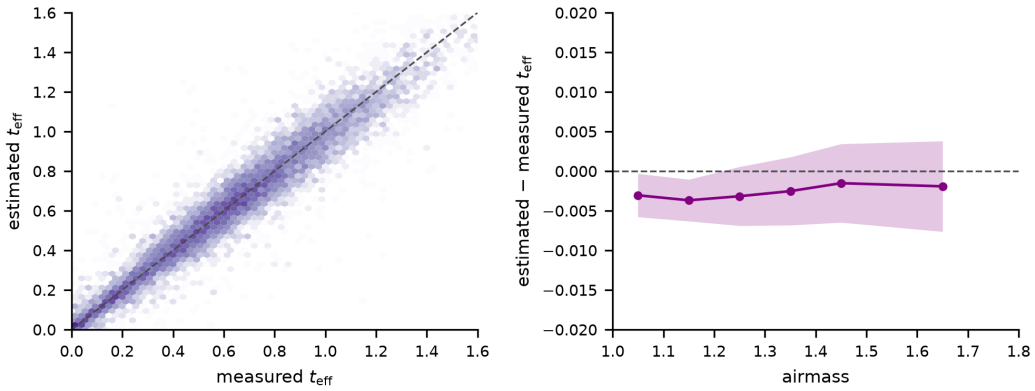

图5：（左）估算值与测量值 t_{eff} 的对数密度，并标出一一对应线（黑色虚线）。（右）按气团质量分箱后，估算值减去测量值所得 t_{eff} 的平均残差；阴影带表示通过刀切重采样得到的 95%区间。

散布为 0。064。该偏差与气团质量无关。由于假定云层在整个天空中均匀分布，指向局部云层较多（较少）区域的观测计划不会受到惩罚（获得奖励）。