arXiv:2301.00415v1 [astro-ph.IM] 1 Jan 2023

用于詹姆斯·韦布空间望远镜数据处理的基于PCA的数据降维与信号分离技术

Hatipo˘glu，G¨uray

2023年1月3日

摘要

基于主成分分析（PCA）的技术可将数据分离为不同的、不相关的成分，并作为预处理步骤促进统计分析。独立成分分析（ICA）可通过非参数迭代算法分离统计独立的信号源。非负矩阵分解是另一种类似PCA的方法，可将维度归类为具有物理可解释性的组。奇异谱分析（SSA）是一种与时间序列相关的类PCA算法。在介绍并回顾利用近红外相机（NIRCam）和中红外仪器（MIRI）处理JWST数据的相关文献之后，本文将讨论詹姆斯·韦布空间望远镜成像数据降维流水线中可能进行干预的部分。

1 方法论背景

1.1 主成分分析

读者可以阅读 Shlens (2014) 的教程以及 Everitt 和 Hothorn (2011) 的著作，以了解主成分分析背后的理论及其优缺点。以下是对主成分分析的简要介绍：

协方差矩阵的构建
我们有一个包含 N 个样本和 M 个变量的数据集。首先，对于每个样本，都要从其对应的变量值中减去该变量的均值。构建协方差矩阵并使用协方差的目的是观察初始变量共同变化的程度。协方差的计算公式如下：

$$cov_{xy} = \sigma_{xi}\sigma_{yi} = \sum_{i=1}^{n}(x_i - \mu_x) * (y_i - \mu_y)$$

共同变化的变量会具有较高的协方差值，这是因为它们各自偏离均值的数值相乘后，结果要高于一个变量发生变化而另一个变量接近其均值的情况。

完成此步骤后，构建相关矩阵即可检查变量间的多重共线性。可对该矩阵应用矩阵分解方法（例如特征值分解），以

将整个数据集从原始变量空间旋转到一个新空间，其中主成分 按数据方差从大到小排列。主成分是由具有不同权重的变量构成的线性组合。例如，第一个主成分为：

$$\alpha_1^{'}\mathbf{x} = \alpha_{11}x_1 + \alpha_{12}x_2 + ... + \alpha_{1p}x_p + = \sum_{i=1}^{m}(\alpha_{1i}x_i)$$

从数据集中提取 PC1 后，第二主成分应当是其中包含方差最大的主成分，并且应与 PC1 及所有其他主成分正交。更多详情请参见 Joliffe (2002)。特征值分解是生成主成分（特征向量）以及一组特征值的一种方法，这组特征值表示每个主成分所承载的总方差。一个方阵（例如我们在生成上述协方差矩阵后得到的矩阵）将按如下方式进行特征值分解：

$$A = Q * \lambda * Q^{-1}$$

上式中的 A 可以是我们生成的协方差矩阵。 λ 等式右侧中间位置的矩阵是一个对角矩阵，其对角线元素为特征值；Q 则是一个特征向量矩阵（有关主成分分析的教程，请参见 Shlens (2014)）。这就是经典主成分分析。它的一个问题是成分难以解释。例如，最大方差 旋转可能通过减少中等载荷的数量并增加 1、0、-1 的数量来提高可解释性（Kaiser，1958），但这通常会以牺牲成分的解释力为代价。除了下文概述的几种方法外，采用稀疏性 调整也可能改善这一问题。Zou、Hastie 和 Tibshirani (2006) 展示了稀疏主成分分析的一种构造方法。

1.2 独立成分分析（ICA）

独立成分分析（ICA）比主成分分析（或奇异值分解）更进一步，旨在生成统计独立的成分，同时实现零相关。Shlens (2014b) 提供了一份独立成分分析教程，包含理论、示例和示例代码，可用于运行高效的独立成分分析算法。ICA 分为两个步骤：首先通过奇异值分解（SVD）消除二阶相关性，然后运用统计技术以数值方式消除剩余的（假定存在的）相关性。SVD 如下：

$$A = U * \Sigma * V^T$$

U 是左奇异向量矩阵，V 是右奇异向量矩阵，而 Σ 是一个对角矩阵，其奇异值从左上到右下按降序排列。由于 U 和 V 矩阵是正交的，它们有助于按以下方式对矩阵 A 进行分解：

$$A = U * \Sigma * V^T, A^T = V * \Sigma^T * U^T$$

$$A^T A = V * \Sigma^T * U^T U * \Sigma * V^T = V * \Sigma^T \Sigma * V^T$$

$$AA^T = U * \Sigma * V^T V * \Sigma^T * U^T = U * \Sigma^T \Sigma * U^T$$

此时，我们以类似特征分解的方式求解 $A^T A$ 和 AA^T，因为它们在结构上相似。此外，U 和 V 的转置分别与自身相乘，结果均为单位矩阵 I。根据 Shlens (2014b) 中的方法，分解的一半（U 和 Σ）通过类似主成分分析的方法求解，以消除二阶相关性，输入为数据的协方差。然后，通过数值方法求出剩余的 V 矩阵，以确保奇异分量之间的依赖性最小。统计独立性：

$$I(y) = \int P(y) \log_2 \frac{P(y)}{\prod_i P(y_i)} dy$$

这是多信息的定义：非负的 $I(y)$ 值当且仅当所有 y_i 彼此独立时，才达到最低值（0）。经过第一步主成分分析进行去相关，并乘以奇异值（归一化）后：

$$\hat{s} = V x_w$$

其中，$\hat{s}$ 是源矩阵，x_w 是经过前序变换后的矩阵，V 是剩余的旋转矩阵。这里的关键问题如下：通常，按 U、Σ、V 的顺序相乘，我们会还原出 A。然而，构造 V 以最大化所得成分之间的独立性，会使所得运算的列的逆（W^{-1}）中出现不同的成分：

$$W = V * D^{\frac{-1}{2}} * E^T$$

其中，W 是解混矩阵。到目前为止，我们尚未作出噪声假设，即这些成分是数据中的实际成分，例如鸡尾酒会上某个人的声音和背景音乐。通过这种方式，独立成分分析试图生成彼此统计独立的成分。它的一种扩展——依赖成分分析（Li、Li 和 Wang，2010）——能够找到在统计上相互独立的成分簇，其中每个簇内部的成分彼此依赖。

1.3 非负矩阵分解-NMF

非负矩阵分解（NMF）是一种将矩阵分解为非负成分的方法，有助于提高可解释性。

$$M = UV^T$$

M 是具有 n 行和 p 列（m x n）的原始矩阵，U 为（m x r），V 为（n x r），目标为：

$$\arg \min_{\{U,V\} \geq 0} ||M - UV^T||_F$$

有许多计算高效算法可在不同假设下从矩阵 M 中提取非负元素。可分非负矩阵分解就是其中一例

（SNPA）。其额外的可分性假设可迭代定义，直至r 秩（Nadisic 等人，2006），其J 索引如下：

$$H^*(:,j) = \arg\min_{h \epsilon \Delta} f(M(:,j) - M(:,J)h)$$

$$R(:,j) = M(:,j) - M(:,J)H^*(:,j)$$

1.4 奇异谱分析

奇异谱分析（SSA）适用于时间序列数据。它不会对数据作出先验假设，并且是一种非参数方法。它有许多不同的分支，其中主要形式（一维）包括以下步骤（Hassani，2010）：1-]轨迹矩阵：

对于长度为**L** 的窗口，一维时间序列数据会转换为如下多维汉克尔矩阵：

$$X = \begin{bmatrix} y_1 & y_2 & y_3 & \cdots & y_K \\ y_2 & y_3 & y_4 & \cdots & y_{K+1} \\ y_3 & y_4 & y_5 & \cdots & y_{K+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_L & y_{L+1} & y_{L+2} & \cdots & y_T \end{bmatrix}$$

在该汉克尔矩阵中，反对角线元素相等。
2-] 对XX进行奇异值分解 ^{T}：
其形式如下：$XX^T = P\Lambda P^T$
3-]特征向量选择：
这一步是将一维序列分解为不同成分的过程。处理这一步的方法有很多，其中一种是将相关元素归为同一个簇，并将相关性最低的元素分到不同的簇中（Golyandina 和 Ko- robeynikov，2013），同时考察重构元素之间的相关性（加权相关矩阵）。此外，如果事先了解某个特定信号或噪声的频率，将L 窗口设为该频率的倍数，有助于分离该成分。关于选择L 窗口长度的建议，详见 Golyandina、Nekrutkin 和 Zhigljavsky（2001）。

4-] 一维时间序列重构：
这是通过在步骤 3 中选择上述 X 的成分后，使用 $PP^T\mathbf{X}$ 实现的。

2 文献综述

2.1 近红外相机-NIRCam

詹姆斯·韦布空间望远镜近红外相机的调试与首次在天观测结果，以及近红外相机本身的特性，详见Girard等人（2022年）。
Schlawin等人（2022）利用詹姆斯·韦布空间望远镜对HAT-P-14系统的观测数据，研究了近红外相机（NIRCam）短波长时间序列测光的性能。他们报告称，**反射镜倾斜事件** 带来的额外误差在理论光子

噪声的50%以内。这些事件的发生频率逐渐降低，但尚未完全消失。除其他几种方法外，他们还利用主成分分析，通过近红外相机的棱栅和精细制导传感器（FGS）检测倾斜事件，但结果几乎处于噪声水平。他们效果最好的方法是噪声加权差分点积：

$$D_i = ((\vec{A}_i - \vec{R})/\vec{V}) \cdot (\vec{M} - \vec{R})$$

$\vec{A}_i$ 是单张FGS校准数据图像，$\vec{R}$ 是倾斜后的参考图像，$\vec{M}$ 是倾斜前的参考图像，$\vec{V}$ 是方差图像。此方法还需要从近红外相机数据中获取倾斜事件发生时间的额外信息。这些倾斜可能存在于第1周期的所有观测中。

Ahrer 等人（2022）首先利用下一代凌星巡天（NGTS）和凌日系外行星巡天卫星（TESS）研究 WASP-39，以确认 WASP-39b 适合作为詹姆斯·韦布空间望远镜的观测目标。此步骤之后，使用近红外相机进行了 8.2 小时的观测，采用 R棱栅 + F322W2滤光片的长波（LW，2.420-4.$025\ \mu$m）通道，以及 WLP8弱透镜和 F210M滤光片的短波（SW，2.0-2.2 μm）通道。两种模式均使用 SUBGRISM256子阵列模式和 SHALLOW4读出模式。最终，凌星观测共获得 366 次积分。他们使用了**jwst** Python 软件包中的标准数据处理流程，仅作少量调整。在短波光度测量中，使用了 Eureka! 和 tshirt 流水线。在 Eureka! 中进行孔径光度测量后，目标孔径半径设为 65 像素，相对于中心，背景环位于 70 至 90 像素范围内。**tshirt** 采用逐行及按放大器奇偶行（ROEBA）扣除方法来降低 1/f噪声。光度提取的源半径为 79 像素，背景环范围为 79 至 100 像素。在长波光谱学中，使用了 Eureka!、大气透射光谱分析代码（HANSOLO）、t shirt 和色度拟合流水线。作为参考，调试计划1076使用了行星状星云 IRAS 05248-7007 的彭德氢线系和布拉开氢线系所提供的三次多项式波长解。Eureka!第1和第2阶段的跳变检测设置为6σ，与标准**jwst** 流水线相同。经过数个筛选步骤后，他们用高斯轮廓拟合光谱迹线。背景扣除使用双重迭代 7σ 异常值剔除阈值。HANSOLO以 jwst 第1阶段的校准输出为起点。根据 LACOSMIC算法去除了宇宙射线效应，并利用莫法特函数识别光谱迹线并对每一列进行拟合。对每列拟合线性趋势并将其扣除后，去除了天空背景，同时排除了迹线中心两侧各 20 像素的区域。通过互相关对齐不同图像的光谱。在 tshirt第3阶段中，最优光谱提取将像素间协方差用作权重；更多空间排除细节见论文的方法部分。他们提醒，边缘处的误差偏差较大，因此建议使用 2.420-4.$025\ \mu$m 波长范围。他们认为略作修改的 Eureka! 表现最佳，并开展了 ATMO、PHOENIX 和 PICASO 3.0 的大气辐射—对流平衡模型。所有正向模型均显示，大气具有太阳至超太阳金属丰度（太阳丰度的 1-100 倍）、次恒星碳氧比（$\leq$0.36），且云层贡献显著。他们还对 H_2O，CH_4和CO_2 丰度进行了进一步约束。

Carter 等（2022）以及下文 MIRI 小节中概述的 MIRI，均使用詹姆斯·韦布空间望远镜的近红外相机对超级木星系外行星 HIP 65426 b 进行高对比度成像（HCI）。他们采用了角度差分成像（ADI），并使用 HIP 68245 进行参考差分成像（RDI）。这些参考观测设置了九个不同的抖动位置，以构建点扩散函数库，捕捉恒星与

日冕仪掩模在每次科学曝光中的不同失准。spaceKLIP Python 软件包生成了经过点扩散函数扣除的图像、对比度曲线以及伴星测光和天体测量结果。近红外相机日冕成像使用圆形日冕仪掩模 MASK335R，并针对 F250M、F300M、F410M、F356W 和 F444W 滤光片依次采用两个转动角度。作者提醒，NIRCam MULTIACCUM 探测器读出中可能出现错误的“跳变”检测。对于近红外相机和 MIRI 数据，点扩散函数均采用三种不同的基于主成分分析的方法进行扣除：角差成像、参考差分成像以及 ADI+RDI。

2.2 中红外仪器-MIRI

Bouwman 等（2022）利用詹姆斯·韦布空间望远镜中红外仪器（MIRI）的数据，对凌星系外行星 L 168-9 b 进行了中红外时间序列观测。此次时间序列观测（TSO）持续 4.14 小时，共有 9371 次积分（数据存档于 Mikulski 太空望远镜档案（MAST）中的 PID 1033，观测 5）。观测模式为无缝低分辨率光谱（LRS），使用 F1000W 滤光片，且目标捕获步骤成功。根据精细制导传感器（FGS）和高增益天线（HGA）的工程遥测数据，指向稳定性极佳。其原始数据处理采用的最佳校准路径使用了*jwst* 处理流程的 Detector1 部分，步骤如下：数据质量初始化、饱和、复位、线性度、最后一帧、暗电流、采用 5.0 修正阈值的跳变步骤，以及斜坡拟合；相关细节已在论文中阐述。在第二阶段，即光谱提取与时间序列分析中，他们采用了两种主要基于默认**jwst** Spec2Pipeline 的方法。校准和参考文件使用了 JWST 校准参考数据系统（CRDS）流水线映射版本 859。**分配世界坐标系** 和**平场** 为光谱图像中的每个像素指定了赤经和 DEC。背景校正采用随时间变化的方法，并针对每次积分分别扣除中值背景。中值背景取自色散方向垂直方向上的第 10 至 17 列及第 57 至 72 列探测器列。他们使用“使用 Casual Data 进行凌星光谱校准”（CASCADe）软件包中的 CASCADe 滤波来搜索宇宙线击中和坏像素。随后，CASCADe 抖动定位光谱迹线，并在 Canny 边缘滤波器（Canny，1986）的实现中计算 3×3 Scharr 算子（Scharr，2000）的雅可比矩阵和海森矩阵。依据 Canny 边缘滤波器，此步骤沿最大特征值对应特征向量的方向，对光谱迹线相关像素的海森矩阵进行二阶泰勒展开。一维光谱由**extract1d** 处理流程步骤提取，使用的自定义参数来自前一步生成的迹线拟合。之后，CASCADe 凌星光谱软件包通过半同胞回归方法校准时间序列数据，并提取 L 168-9 b 的透射光谱。他们将光谱重分箱，在 7.5 μm 处的光谱分辨率为 50；由于时间序列数据开头 30 分钟（对应 1019 次积分）存在较强的响应漂移，因此予以跳过。根据 L 168-9 b 的发现（Astudillo-Defru 等，2020）设定偏心率和倾角，并使用数据中的观测凌星中点时间及相对半长轴；借助 ExoTETHyS 软件包（Morello 等，2020），按照 Claret（2000）采用非线性临边昏暗校正。光变曲线使用 Batman 软件包计算（Kreidberg，2015）。他们通过包含 600 个样本的自助抽样估计凌星深度拟合和系统误差的误差，更多细节见 Carone 等（2021）。除 CASCADe 流程外，他们还使用了 0.5 版本的**Eureka!**处理流程，共分为 5 个阶段。第 1 阶段的输出与 CASCADe 相同；第 2 阶段使用 11.16.5 版本的**crds** 软件包和 1.6.0 版本的**jwst** 处理流程。为避免其对时间序列数据造成劣化，跳过了“photom”步骤。第 3 阶段的噪声去除窗口设为 140–393 个 y 行和 13–64 个 x 行。经过比较 8–16 后，选定源孔径半宽为 4 、背景排除区域半宽为 10。第 4 阶段将光谱分为 48 个通道，通道宽度为 0.149 μm，范围为 4.86–1.86 μm，并设置一个白光通道。由 HGA 引起的

异常也在此步骤中移除：使用宽度为 500 次积分的箱形滤波器，并通过迭代西格玛裁剪最多迭代十次。第 5 阶段使用**dynesty** 根据**batman** 凌星模型拟合观测数据；有关白光曲线和临边昏暗相关处理的更多细节见论文。最终，CASCADe 和 Eureka! 两种流程所得结果彼此一致，且均与 Patel & Espinoza（2022）的结果相符。

除近红外光谱仪外，Miles等人（2022）也使用詹姆斯·韦布空间望远镜上的MIRI研究了褐矮星VHS 1256 b。MIRI覆盖了全部4个积分视场单元通道，波长范围为4.98至28.1 μm，使用短、中、长三种光栅观测了约2小时。读出模式为FASTR1。我们没有使用jwst 流水线中的背景扣除，而是在经过校准的科学数据立方体中，将一个参考孔径放置在目标之外。与近红外光谱仪类似，在进行孔径光度测量之前，先进行了塌缩和二维高斯拟合。ResidualFringeStep检测并去除了MIRI光谱部分区域中残留的干涉条纹。整个通道4（17.71至20.94 μm）被塌缩为一个测光点。他们报告称，近红外光谱仪与MIRI的重叠波长范围内结果一致。最后，他们报告发现了由湍流垂直混合导致的非平衡化学、能够形成云的小型硅酸盐颗粒，以及 H_2O、CH_4、 CO和 CO_2的存在。

Carter 等（2022）也将 MIRI 与近红外相机日冕仪观测结合起来，使用四象限相位掩模日冕仪观测 HIP 65426 b；每个象限对应不同的滤光片，并使用不同滤光片（F1140C 和 F1550C）重复采集数据。鉴于跳变检测阈值设得非常低，他们将 F1140C 滤光片的阈值从 4 提高到 8。本研究通过去除每次曝光的第一个积分，大幅规避了探测器复位异常。MIRI 日冕仪视场中的“荧光棒”杂散光特征，则针对每个滤光片，利用专门背景观测中每 4–5 次积分生成的中值背景图像进行扣除。他们尝试使用 Karhunen-Loeve图像处理（KLIP）和局部优化图像组合（LOCI），将荧光棒提取为单独的主成分，但要么残余通量也受到损害，要么无法捕捉积分之间的变化。尽管如此，他们的中值帧背景扣除效果非常好，几乎没有改进空间。对于恒星绝对位置，他们尝试了几种方法，但均未能确定其位置，因此以日冕仪掩模中心作为位置的替代值。对于近红外相机和 MIRI 数据，他们使用三种不同的基于主成分分析的方法，即角差成像、参考差分成像和角差成像+参考差分成像，来扣除点扩散函数。仅在 MIRI 的 F1550C 滤光片中使用角差成像时，未能探测到伴星 HIP 65426 b；其余滤光片使用角差成像和参考差分成像均成功探测到该伴星。

3 詹姆斯·韦布空间望远镜数据处理——成像提纲

以下表格和信息编译自空间望远镜科学研究所 (STScI) 网站的网络研讨会资料部分，该网站专门介绍 $JWebbinars$ (STScI, 2022年)，以及 JWST 用户文档中的$calwebb\ detector1$ (JWST 用户文档，2022年)。

-

阶段	Step	近红外相机	MIRI
第1阶段	数据质量初始化	✓	✓
第1阶段	饱和度检查	✓	✓
第1阶段	误差初始化（EI）	✓	✓
第1阶段	EI－复位异常校正		✓
第1阶段	EI－首帧校正		✓
第1阶段	EI - 末帧校正		✓
第1阶段	EI - 线性度校正		✓
第1阶段	EI - RSCD校正		✓
第1阶段	EI - 暗电流扣除		✓
第1阶段	EI - 超偏置扣除	✓	
第1阶段	参考像素校正 (RPC)	✓	✓
第1阶段	RPC - 线性度校正	✓	
第1阶段	RPC - 持久性校正	✓	
第1阶段	RPC - 暗电流扣除	✓	
第1阶段	跳变检测	✓	✓
第1阶段	斜率拟合	✓	✓
第 2 阶段	WCS信息	✓	✓
第 2 阶段	背景扣除	✓	✓
第 2 阶段	平场校正	✓	✓
第 2 阶段	通量标定	✓	✓
第 2 阶段	整流二维产品	✓	✓
第3阶段	精化世界坐标系	✓	✓
第3阶段	移动目标世界坐标系	✓	✓
第3阶段	背景匹配	✓	✓
第3阶段	异常值检测	✓	✓
第3阶段	图像组合	✓	✓
第3阶段	源目录	✓	✓
第3阶段	更新后的曝光级产品	✓	✓

以下概述这些步骤的内容，以及上述数据转换方法可以在哪些环节发挥作用：

第*1*阶段：
Data 质量初始化：像素将具有数据质量标志和不确定度，这些标志和不确定度将upd在必要时通过流水线进行。
饱和：此步骤旨在找出因来自极亮源的光子过多而发生饱和的像素；当超过某一阈值后，入射光子数量的增加不会改变探测器的 数字量 。在确定探测器对这些光子的响应之前，此步骤会移除饱和像素。

复位异常校正：这一针对 MIRI 的特定校正与 MIR斜坡中前 12 个组因暗场复位而出现的异常有关。复位异常是一种系统性的加性瞬态偏移。

首帧校正：MIRI 的首帧仍具有尚未表征的瞬态特征。因此，目前正将其从分析中移除。

末帧校正：与首帧类似，最后一帧目前也存在异常
在处理的后续步骤中将其舍弃。
线性度校正（MIRI）：MIT探测器的非线性响应特性已得到充分表征，呈现低阶多项式结构。在 ¿21 μ m 附近，该响应具有波长相关性。针对特定探测器和波长的校准参考文件会对这一误差进行补偿。

ResET 开关电荷衰减 - RSCD：一种特征明确的衰减指数函数，其大小与

前一次积分中的计数成正比，用于描述这种积分次数相关瞬态。借助探测器专用校准参考文件，此校正会从第二次及之后的积分中减去该指数函数。

暗场扣除（MIRI）：MIRI响应包含与积分编号相关的分量，使用积分专用校准参考文件进行逐组暗场扣除，可将其去除。

超偏置扣除：特定偏置会使每个像素和每组的数字量发生偏移，且该偏置与斜坡中的非线性成正比。校准参考文件有助于消除这一误差。

线性度校正（近红外相机）：校正超偏置后，使用探测器专用校准参考文件，通过低阶多项式拟合来表征近红外探测器响应中的非线性。

持久性校正：近红外探测器中，先前曝光的残像会持续出现在后续曝光中。此外，残像还会在不同曝光之间呈指数衰减。使用探测器专用校准参考文件可校正这一误差。暗电流扣除（NIRam）：使用探测器专用校准参考文件，逐组扣除暗场曝光中近红外探测器的过量响应。

参考像素校正（RPC）：通过一个对光不敏感的等效像素对像素进行校准，从而显现并扣除电子相关贡献对响应的影响。

跳变检测：当两个连续的探测器组之间的响应差异大于其他连续探测器组时，通常是宇宙射线击中相应像素所致。判定某个值为跳变 时所采用的阈值，以高于噪声的标准差数量来设定。雪团和簇射 噪声检测也属于此步骤的范围。

斜率拟合：也称为斜坡拟合，其本质是在考虑跳变标记像素及处理流程前序步骤的基础上，求出响应／光子斜率。

第 *2* 阶段：

世界坐标系（WCS）信息：仪器和探测器专用校准参考文件提供这些WCS信息及畸变模型，用于将像素坐标转换为天文学中的赤经（RA）和赤纬（DEC）坐标。

背景扣除：当科学目标曝光有对应的背景目标时，会从目标曝光中扣除其合成背景图像。

平场校正：此步骤利用仪器和探测器专用校准参考文件，校正像素间响应差异及响应中的大尺度变化。

通量标定：通量标定（或测光校准）会对数据应用一个转换因子，将每秒计数转换为兆扬斯基每球面度；此外，还会应用另一个转换因子，将每秒计数转换为微扬斯基每平方角秒。

整流二维产品：生成整流后的二维图像，以供目视检查。

第*3*阶段：

优化相对 WCS - Tweakreg：**AstroDrizzle** 利用不同图像中共同源的位置，优化其相对世界坐标系信息。

移动目标 WCS：对于移动目标，此步骤会将输出帧的最终世界坐标系中心设在移动目标的平均位置。

背景匹配 - Skymatch：当背景存在重叠时，**AstroDrizzle** 会通过此步骤匹配不同图像中的背景响应。

异常值检测：这是对异常值的第二次检查 , 现在使用 **AstroDrizzle** 代码。

图像组合 - 重采样：此步骤将关联文件中同一波段的多张图像拼接成马赛克，并排除“坏”像素。

源目录：此步骤使用 Photutils 源提取功能，对给定图像中的不同光源进行编目。此外，还会执行去混叠程序，以分离重叠源。更新曝光级产品：在处理流程完成后生成质量最高的产品。

3.1 通过数据变换方法进一步检查的步骤

用于减少原始JWST数据的现有方法不断改进。这些流水线中的若干步骤采用基于PCA的数据变换方法后，可能会取得更好的效果。

持久性校正（第*1*阶段参考像素校正）：图像包含来自先前曝光的实际曝光和响应。可以尝试根据曝光顺序，使用主成分分析和独立成分分析将这些曝光彼此分离，而先验成分数量将简单地设为曝光顺序。

背景扣除（第*2*阶段）：背景扣除步骤对主成分分析结果的影响；即，背景扣除步骤前后主成分结果的差异可能揭示潜在的过度扣除。

背景匹配 - *Skymatch*（第*3*阶段）：在此步骤之前，算法处理的是包含目标以及其他点源、弥散源和天空背景信号的单幅或多幅图像。这本质上是一个源分离问题。除了宇宙射线等造成的误差外，每幅图像中都应当始终存在不相关且统计独立的信号源。特定区域帧的不同曝光可能会被视为不同维度进行分箱，并且应尝试将同一像素的所有像素值降至更少的成分。这可能有助于背景扣除。

源目录（第*3*阶段）：此步骤通常在 Python 中借助 Astropy 的 Photutils 包，通过设置阈值来查找来自单个源的像素值。考虑不同积分之间时间变化的基于像素的类PCA方法（例如奇异谱分析）可能会揭示更多源特有信号，或者解混，以替代或补充基于分割的方法。

4 总结

总而言之，先进的数据变换方法所涵盖的众多不同步骤，或许能带来更多由数据驱动的见解。我们需要探索这些潜在的改进空间，同时不能忽视可能得到增强的处理流程在计算效率和稳健性方面的表现。

5 参考文献

A. Claret. LTE恒星大气模型的新型非线性边缘昏暗定律。针对-5.0 ¡= log[M/H]¡= +1、2000 K ¡= Teff ¡= 50000 K，在多个表面重力下进行计算。《天文学与天体物理学》*363*。（2000）。

A. L. Carter、S. Hinkley、J. Kammerer、A. Skemer、B. A. Biller、J. M. Leisenring、M. A. Millar-Blanchaer、S. Petrus、J. M. Stone、K. Ward-Duong、J. J. Wang、J. H. Girard 等。JWST早期发布科学计划：系外行星系统直接观测 I：对系外行星 HIP 65426 b 的高对比度成像，波长范围为 2-16 μm。（2022）。arXiv：2208.14990

B. Everitt、T. Hothorn。《R应用多元分析导论》。*Springer*。（2011）。E. M. Ahrer、K. B. Stevenson、M. Mansfield、S. E. Moran、J. Brande、G. Morello、C. A. Murray、N. K. Nikolov、D. J. M. Petit dit de la Roche、. Schlawin、P. J. Wheatley 等。利用 JWST 近红外相机对系外行星 WASP-39b 进行早期发布科学观测。（2022）。arXiv：2211.10489 E. Schlawin、T. Beatty、B. Brooks、N. K. Nikolov、T. P. Greene、N. Espinoza、K. Glidic、K. Baka、E. Egami、J. Stansberry、M. Boyer 等。JWST 近红外相机离焦成像：测光稳定性表现及其感知镜面倾斜的能力。（2022）。arXiv：2211.16727 G. Morello、A. Claret、M. Nartin-Lagarde、C. Cossou、A. Tsiaras、P.-O. Lagage。《ExoTETHyS 软件包：主星周围系外行星凌星的工具》。《天文学杂志》*159*（2020）。doi：10.3847/1538-3881/ab63dc H. F. Kaiser。因子分析中用于解析旋转的最大方差准则。《心理计量学》*23*。（1958）。doi：10.1007/BF02289233 H. Hassani。奇异谱分析简要介绍。获取自：https: // ssa.cf.ac.uk/ssa2010/a_brief_introduction_to_ssa.pdf（2010）。H. Scharr。《数字图像处理中的最优算子》。博士论文。（2000）。doi：10.11588/heidok.00000962 H. Zou、T. Hastie、R. Tibshirani。稀疏主成分分析。《计算与图形统计学杂志》*15*。（2006）。doi：10.1198/106186006x113430 I. T. Joliffe。主成分分析。*Springer*。第2版。（2002）。J. A. Patel、N. Espinoza。基于凌日系外行星巡天卫星数据的已知系外行星经验临边昏暗系数与凌星参数。《天文学杂志》*163*。（2022）。doi：10.3847/1538-3881/ac5f55 J. Bouwman、S. Kendrew、T. P. Greene、T. J. Bell、P. Lagage、J. Schreiber、D. Dicken、G. C. Sloan、N. Espinoza、S. Scheithauer、A. Coulais、O. D. Fox 等。JWST 中红外仪器光谱时间序列的性能。（2022）。arXiv：2211.16123 J. Canny。一种边缘检测的计算方法。《*IEEE*模式分析与机器智能汇刊》。*PAMI-8* **679**。（1986）。doi：10.1109/TPAMI.1986.4767851 J. H. Girard、J. Leisenring、J. Kammerer、M. Gennaro、M. Rieke、J. Stansberry、A. Rest、E. Egami、B. Sunnquist 等。JWST/近红外相机日冕仪成像：调试与首次天空观测结果。（2022）。arXiv：2208.00998 J. Shlens。独立成分分析教程。（2014b）。arXiv：1404.2986 J. Shlens。主成分分析教程。（2014）。arXiv：1404.1100 JWST 用户文档。calwebb detector1。https://jwst-docs.stsci.edu/jwst-scienc

詹姆斯·韦布空间望远镜数据处理阶段/calwebb_Detector1。(2022年)。L. Carone、P. Molli`ere、Y. Zhou、J. Bouwman、F. Yan、R. Baeyens、D. Apai、N. Espinoza、B. V. Rackham、A. Jord´an。《炽热的类地行星》WASP-117b 的高金属丰度迹象及甲烷缺失。《天文学与天体物理学》*646*。(2021年)。doi：10.1051/0004-6361/202038620

L. Kreiderg. batman：Python 中的基本凌日模型计算。《太平洋天文学会刊》*127*。(2015年)。doi：10.1086/683602 N. Astudillo-Defru、R. Cloutier、S. X. Wang、J. Teske、R. Brahm、C. Hellier、G. Ricker、R. Vanderspek、D. Latham、S. Seager、J. N. Winn、J. M. Jenkins、K. A. Collins、K. G. Stassun 等。凌日系外行星巡天卫星揭示了绕明亮 M 型矮星 L 168-9 运行的炽热类地行星。《天文学与天体物理学》*636*。(2020年)。doi：10.1051/0004-6361/201937179 N. Golyandina、A. Korobeynikov。奇异谱分析基础与使用 R 进行预测。(2013年)。arXiv：1206.6910 N. Golyandina、V. Nekrutkin、A. Zhigljavsky。《时间序列结构分析：奇异谱分析及相关技术》。Chapmann&Hall/CRC。(2001年)。N. Nadisic、A. Vandaele、J. E. Cohen、N. Gillis。稀疏可分非负矩阵分解。(2006年)。arXiv：2006.07553 R. Li、H. Li、F. Wang。依赖成分分析：概念与主要算法。《计算机学报》*5*。(2010年)。doi：10.4304/jcp.5.4.589-597 STScI。通过 JWebbinars 学习詹姆斯·韦布空间望远镜数据分析。材料与视频，*3-Pipeline*：成像模式（*2021*年*5*月）。https://stsci.app.box.com/s/gz5q1r8ijb1hvimmeh8fi7xlb3b

368ds。

(2022年)。T. J. Bell、E. M. Ahrer、J. Brande 等。《Eureka!：用于韦布空间望远镜时间序列观测的端到端流水线》。(2022年)。arXiv: 2207.03585