

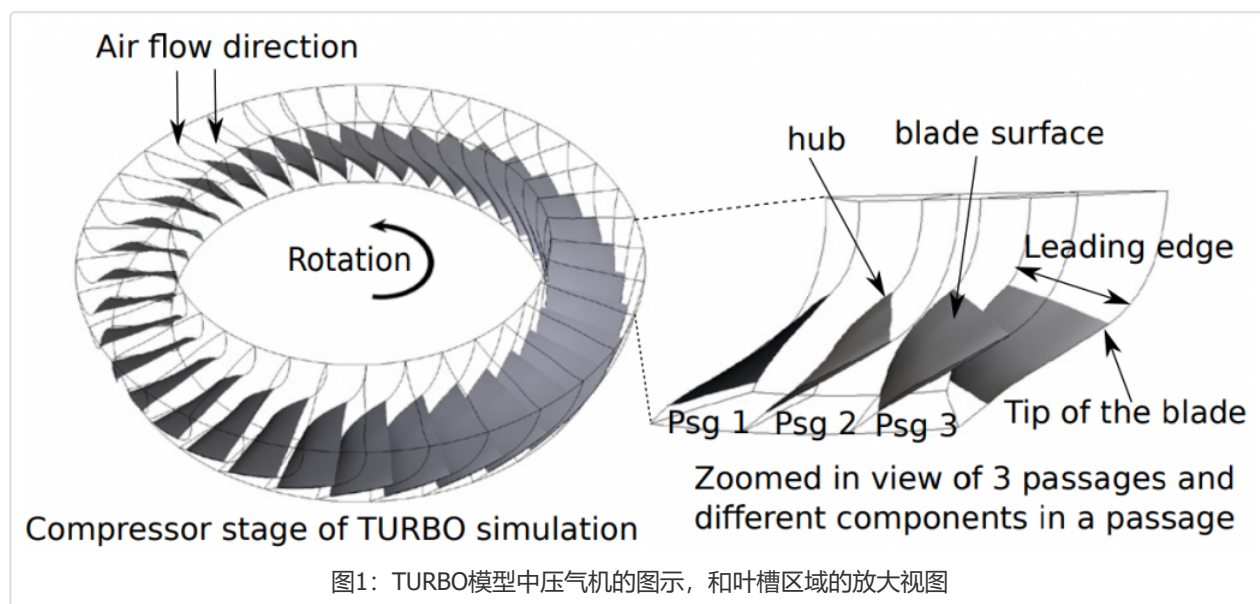
PKU Visualization Blog

北京大学可视化与可视分析博客

分布导向的喷气发动机模拟的原位分析和可视化 (In Situ Distribution Guided Analysis and Visualization of Transonic Jet Engine Simulations)

作者: Fan Hong 日期: 2016年11月19日

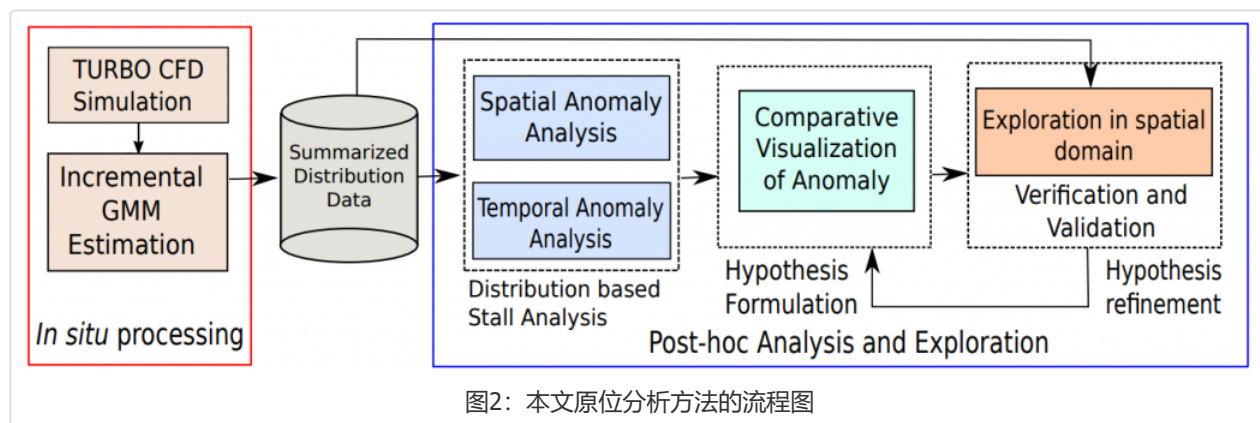
这篇文章关注的是一个科学应用领域的问题：喷气发动机的旋转失速 (rotating stall) 现象。喷气发动机中的压气机在运转过程中，可能由于气流的不稳定，阻碍压气机中叶片的旋转，从而导致叶片逐渐停止转动，发动机失去动力。这是一种十分危险的现象。领域科学家们希望借由模拟来研究旋转失速这种现象，尤其是希望能发现这种现象的早期征兆，从而预防其发生。最近，NASA开发了名叫TURBO的模型，能够以非常高的精度来模拟压气机的行为，从而能够极大地帮助领域专家对旋转失速的研究。然而，其模型的高精度特性也使得产生的数据量非常巨大，存储所有数据带来的I/O花费变得难以承受，从而导致传统的后处理 (post-processing) 分析难以直接应用。基于这些背景，本文提出了分布 (distribution) 数据导向的原位分析和可视化框架，用以解决这个问题。这篇文章来自于今年的IEEE VIS科学可视化部分，并获得了最佳论文的提名奖。



领域专家对TURBO模拟的分析有四个需求：

1. 能显著减少所产生的数据规模，但也要保留原始数据中的重要信息。对于一个完整的TURBO模型，模拟4转将产生20TB的数据。随着数据规模的增加，数据的I/O和移动都将成为后处理的瓶颈。
2. 偏好于基于区域的分析和时变分析，这能帮助及早发现异常可能产生的时间和空间区域。而已有的领域分析方法通常无法提供基于时空的探索分析。
3. 能够比较多个变量作为异常指示器时的表现。绝大部分已有的领域方法通常只用压强来进行分析，领域专家希望了解其他变量的表现。
4. 约减后的数据应该仍能支持空间的渲染，从而能直观地支持结果的探索和验证。

针对领域专家的这些要求，本文提出了分布数据导向的原位分析流程，如图2所示。



在原位处理部分，作者提出使用分布数据来近似表示原始数据，从而实现对数据的约减，也同时保留重要信息。具体来说，原始模拟的空间域被划分为若干个小块 (block)，对于每个时间步里的每个块，用分布来表示原始数据。这里，作者选用高斯混合模型 (Gaussian Mixture Model, GMM) 来表示分布数据，即用若干个不同权重的高斯核来拟合数据的分布。

$$p(x) = \sum_{j=1}^K \omega_j * \mathcal{N}(x|\mu_j, \sigma_j)$$

图3: 高斯混合模型

作者之所以选择高斯混合模型，主要是基于其在存储上的高效率。一些无参数的模型，例如核密度估计，需要高昂的计算的存储代价；直方图能很快的计算，但存储代价高；有参数模型，如使用单个高斯核，则可能导致不精确。

然而，传统高斯混合模型的估计 (estimation) 需要执行EM算法，需要耗费大量的时间，对模型本身的模拟影响太大。因此，作者采用了一种增量式的估计算法，即对于每一个块来说，其在一个时刻的高斯混合模型可以从前一时刻的结果以很小的时间代价推算得到。这个方法在作者去年的IEEE VIS工作[2]上就有使用。具体来说，对于某个数据块，假设在前一时刻已经得到K个高斯核来表示其分布，那么对于当前时刻，数据块中的数据，依次判断是否被这K个高斯核“覆盖”。这里的“覆盖”指的是数据点在一个高斯核均值的2.5个标准差之类。如果有“覆盖”，则更新它们的参数；如果没有，则将权重最小的高斯核替换掉，代之以当前数据点为均值、方差大，而权重小的一个新高斯核。最后，再根据数据点距离高斯核的远近 (“覆盖”程度)，来更新每个高斯核的权重。需要说明的是，第一时间步的高斯混合模型，仍然通过EM算法得到。在实践中，数据块的大小为5x5x5，K值被固定为4。

通过前述算法之后，就得到了每一个数据块在每一时间步的高斯混合模型估计结果，也就是数据分布数据。接下来，作者基于分布数据，进行后处理和分析。

作者首先进行空间异常和时间异常的分析，即要找到可能引发旋转失速的时间和位置。基于空间的检测主要基于叶片间的对称特性，如图4所示。即相对于各个叶片的相同的位置 (例如图中的红色区域)，它们的行为应该是一样的，从而各个变量的数值也应接近，其分布也应近似。如果有叶片附近位置的分布与其他叶片对应位置的分布差异很大，则这个位置被判定为异常。具体在计算中，作者用所有相同位置的分布的均值表示期望的分布，计算计算各个叶片对应位置分布与这个期望分布的EMD距离 (Earth Mover Distance)，如果距离超过指定的阈值，则被认为是异常。基于时间的检测则基于压气机正常运行下应该有的稳定性质，如图5所示。即如果压气机正常运行，对于一个空间位置，其各个变量的数值不应该有剧烈变化。如果有某时刻，变

量的分布产生了突然变化，即相邻两时刻分布的EMD距离超过某个阈值，则被认为是异常。这两种方法实际上是一种互补的关系。绝大部分情况下，两者检测出来的异常所发生的时间和位置应该相同，但对于某些情况下，两者能检测出对方检测不到的异常。例如，当气流的不稳定从单个叶片传播到大部分叶片是，此时的平均分布已经不能代表正常状态了，那么基于空间的检测就会失效；另一方面，如果某个位置的异常已经完全发展起来 (fully developed) 了，那么其分布可能不会再剧烈变化了，而是稳定在一个异常的状态，此时基于时间的检测就会失效。

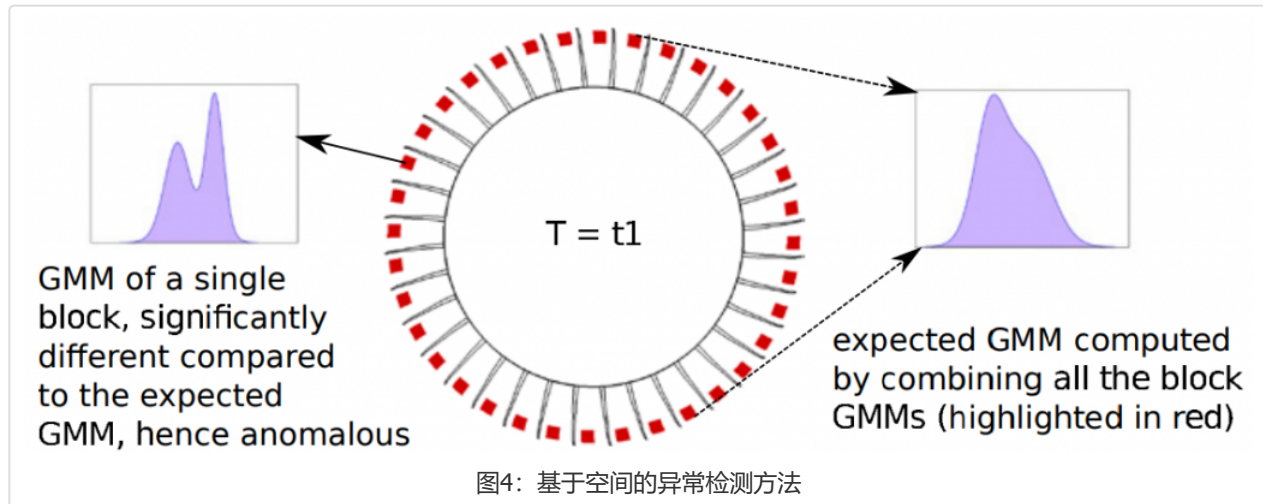


图4：基于空间的异常检测方法

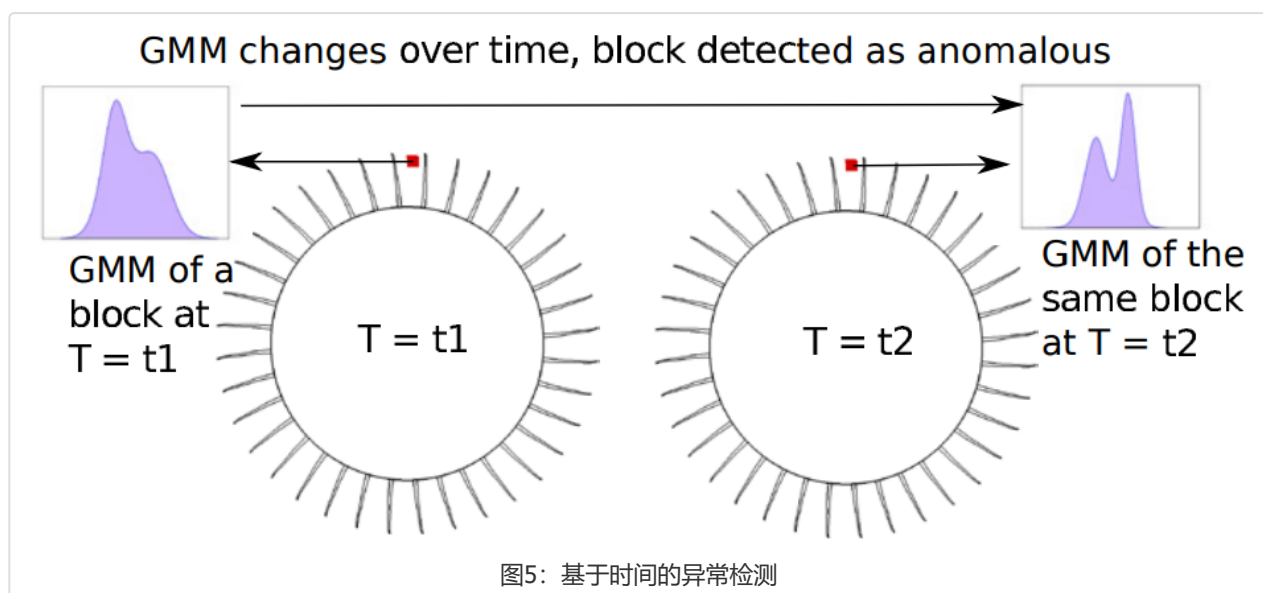


图5：基于时间的异常检测

当两种检测方法检测出异常发生的位置和时间之后，作者提出了异常图 (Anomaly Chart) 对其进行可视化，如图6所示。异常图实际是一个二维的散点图，横轴纵轴分别表示异常发生的时刻和位置。需要指出的是，领域专家更关心异常是如何在叶槽间传播的，而不是在叶槽内传播，因此异常图并没有对异常的具体位置进行可视化，而只是指出了异常发生的叶槽编号。颜色的深浅表示对应时刻对应叶槽内检测到的异常块数量。为了比较多个变量作为异常指示时的表现，作者将两个散点图重叠，将公共的散点用额外的颜色进行高亮。

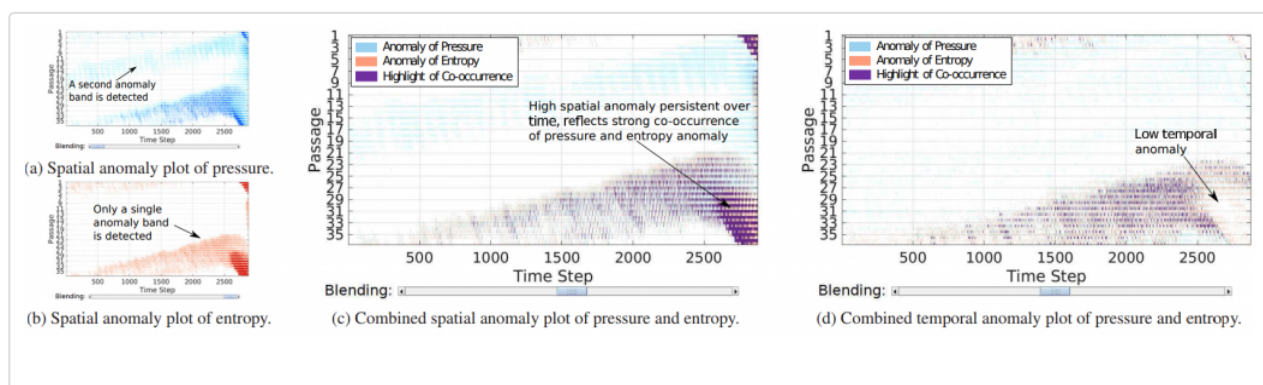


图6: (a)(b)使用压强和熵值检测得到的异常的可视化; (c)同时展示两种变量检测得到的结果; (d)基于时间的检测结果可视化

同时, 基于所得到的分布数据, 作者也提供了等值面渲染来帮助探索和验证结果, 如图7所示。作者使用了两种等值面方法: 均值等值面和不确定等值面。均值等值面就是计算每个数据块的均值, 然后计算等值面得到。不确定等值面是计算每个数据块与真实等值面相交的概率, 然后将其渲染。

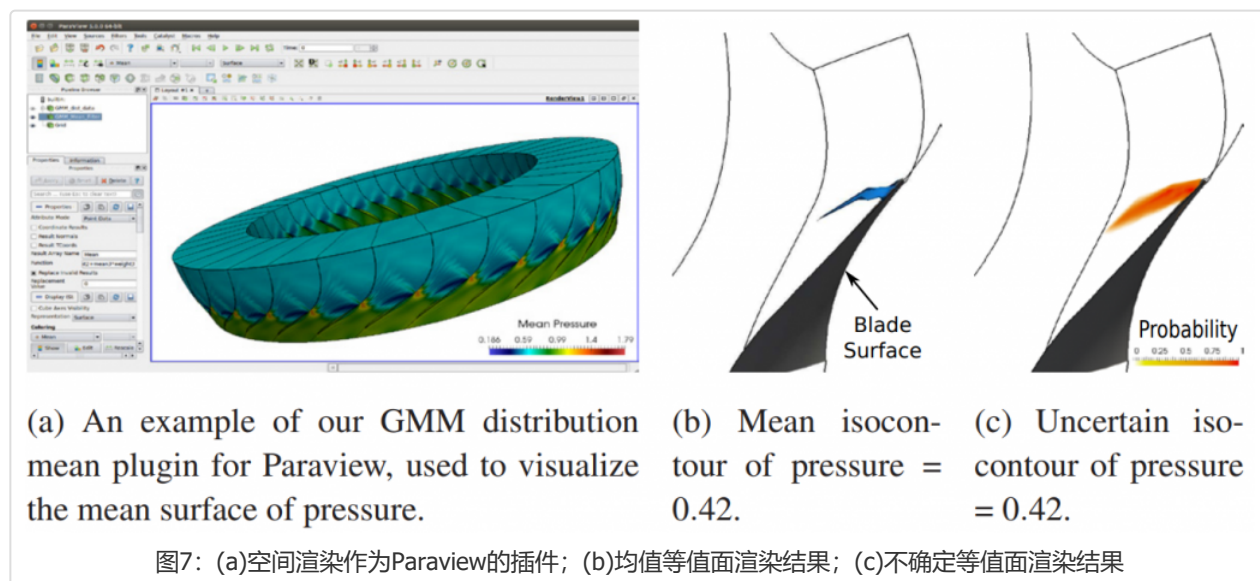


图7: (a)空间渲染作为Paraview的插件; (b)均值等值面渲染结果; (c)不确定等值面渲染结果

最后, 作者进行了案例分析和性能分析。图6是其中一个分析的结果。我们能看到, 在2540时间步左右, 出现了连续的异常位置, 并且是两个属性(压强和熵)同时检测得到。而异常的征兆最早出现在500时间步, 已有的方法难以发现这个征兆。使用压强来检测时, 实际能检测到两个异常带((c)中蓝色的带), 但最终只有一条导致了旋转失速的发生。这说明基于压强的结果并不总是精确, 作者建议和熵一同来使用。领域专家还有许多发现, 可以参看论文, 这里就不赘述。然后, 作者也展示了空间渲染的结果, 如图8所示。可以容易地看到, 基于压强和熵值的检测得到的异常区域基本重合, 并都位于叶片尖端。根据领域专家的先验知识, 旋转失速现象最早起源于叶槽间的stall cells结构。它们原本具有较低的压强, 但由于气流的不稳定, 导致如果用等值面可视化时, 它们的边界十分的不平滑; 同时, 他们的熵值也很高。因此, 图9中分别展示了压强等于0.42和熵值等于1时的等值面。可以看到, 所得到的区域满足先验知识, 并与本文方法检测得到的异常区域一直, 因此也支持了本文异常检测方法的有效性。文中, 作者还用了一个稳定运行的例子, 表明本文方法不会产生误检。

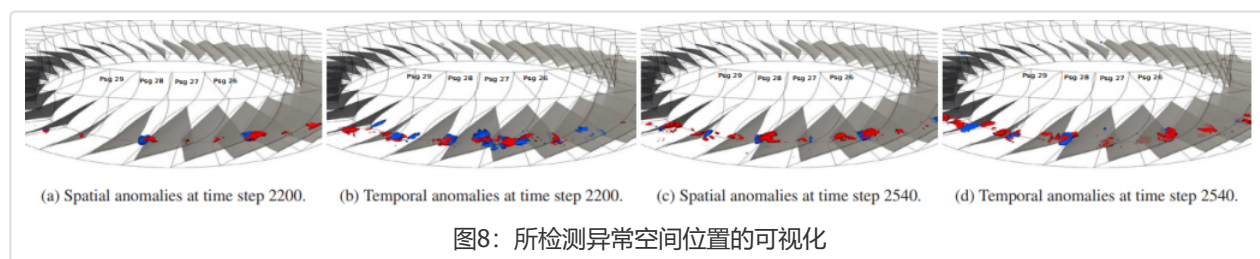


图8: 所检测异常空间位置的可视化

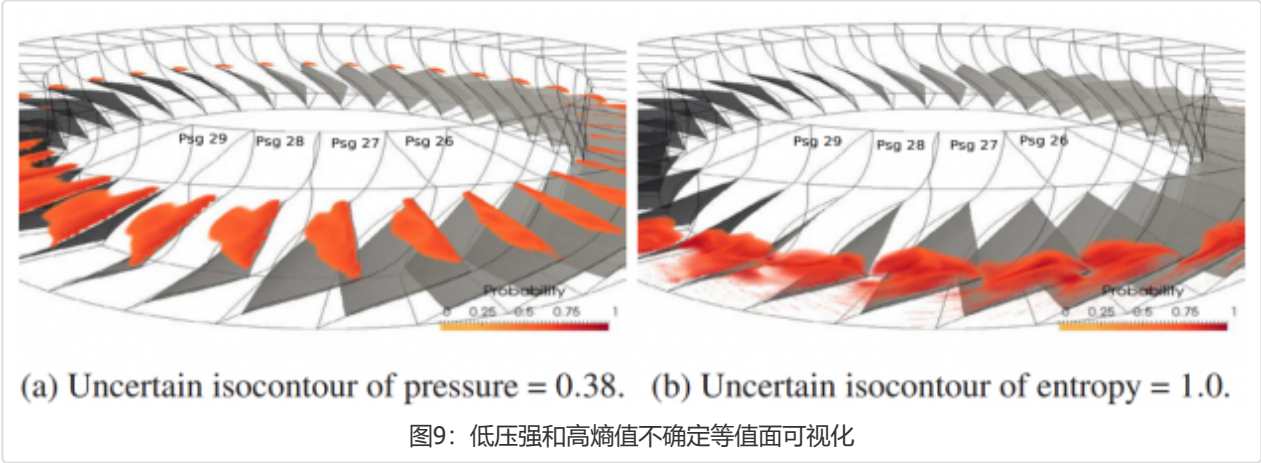
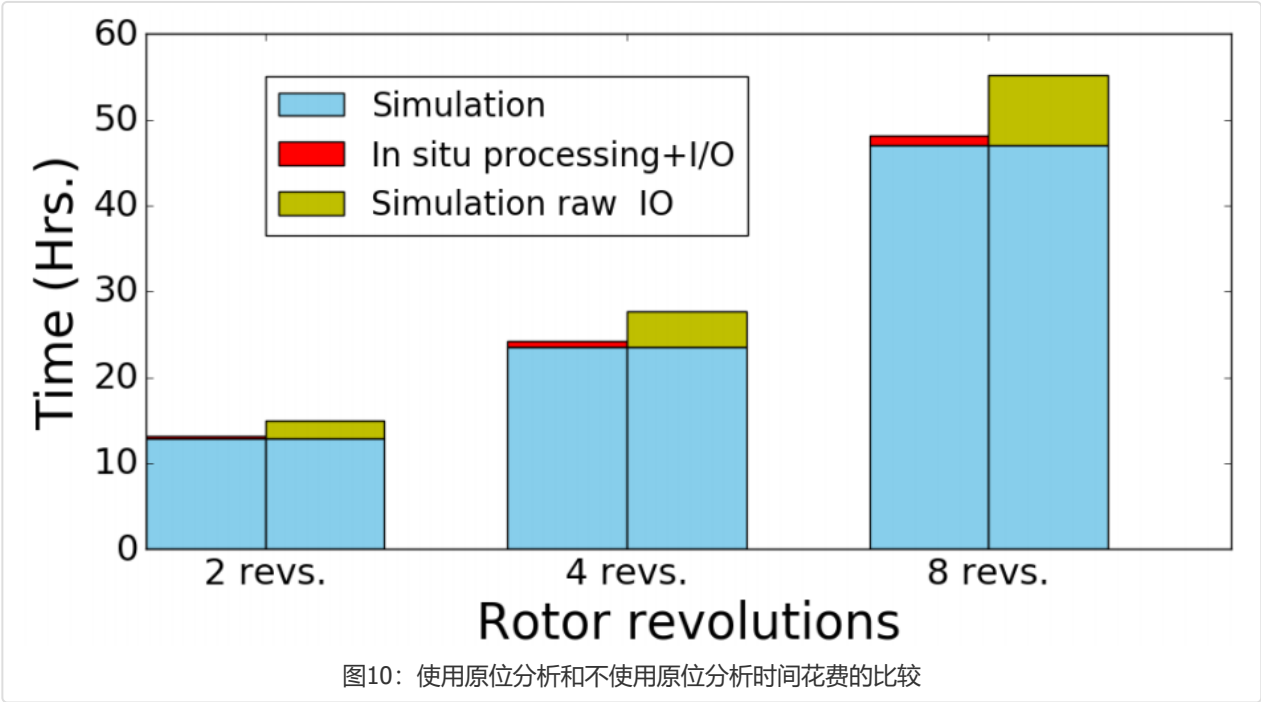


图10展示了使用本文原位处理和不使用原位处理的时间花费。可以看到原位处理极大地节省了数据I/O的时间。作者也测试了原位处理在不同数据规模下的时间花费，高斯混合模型的增量计算带来的额外开销是非常小的。



总结起来，本文提出了针对喷气发动机旋转失速现象的分布数据导向的原位分析流程。对原位分析的原始需求进行了详细的总结，然后分别提出了应对方案。对于原位分析的结果也进行了详细的分析与验证，从而保证原位分析能提供准确的结果。

[1] Soumya Dutta, Chun-Ming Chen, Gregory Heinlein, Han-Wei Shen, and Jen-Ping Chen, In Situ Distribution Guided Analysis and Visualization of Transonic Jet Engine Simulations, IEEE VIS 2016 and to appear in IEEE Transactions on Visualization and Computer Graphics.

[2] Dutta, Soumya and Shen, Han-Wei, "Distribution Driven Extraction and Tracking of Features for Time-varying Data Analysis", IEEE VIS 2015, also in IEEE Transactions on Visualization and Computer Graphics, vol. 22, no. 1, 2016

论文报告