KAN 矩阵：通过可视化非线性成对与多变量贡献获得物理洞见

Luis A. De la Fuente[1], Hernan A. Moreno[1], Laura V. Alvarez[1], 和 Hoshin V. Gupta[2]

[1]地球、环境与资源科学系 德克萨斯大学埃尔帕索分校，德克萨斯州埃尔帕索 79902，美国

[2]水文学与大气科学系 亚利桑那大学，亚利桑那州图森 85721，美国

通讯作者：

Luis A. De la Fuente（ladelafuenteco@utep.edu），开放研究者与贡献者身份识别码：0000-0001-6979-0547 Hernan A. Moreno（moreno@utep.edu），开放研究者与贡献者身份识别码：0000-0003-0408-6588 Laura V. Alvarez（alvarez@utep.edu），开放研究者与贡献者身份识别码：0000-0002-5047-5384 Hoshin V. Gupta（hoshin@arizona.edu），开放研究者与贡献者身份识别码：0000-0001-9855-2839

要点

- 输入输出关联包含两个组成部分：强度和函数形式。
- 与目标变量的成对和多变量关联有助于更深入地理解这种关联。
- 与传统方法相比，KAN 矩阵是一种更稳健且更简洁的工具。

关键词

关联、Kolmogorov-Arnold 网络、相关性、探索性数据分析、成对、多变量。

摘要

解读复杂数据集仍是科学家面临的一项重大挑战，尤其是变量的高维性和共线性所带来的困难。我们介绍了 Kolmogorov-Arnold 网络（科尔莫戈罗夫-阿诺德网络）的一种新颖应用，以提升可解释性和简约性，超越传统相关性分析所能达到的水平。

我们提出了两种可解释的彩色编码可视化工具：成对 *KAN* 矩阵（PKAN）和多变量 *KAN* 贡献矩阵（MKAN）。PKAN 用于刻画变量对之间的非线性关联，而 MKAN 则是一种非线性特征排名工具，用于量化输入变量在预测目标变量时的相对贡献。

这些工具支持模型开发工作流中的预处理（例如特征选择、冗余分析）和后处理（例如模型解释、物理洞见）。通过实验比较，我们证明，与皮尔逊相关和互信息相比，PKAN 和 MKAN 能够产生更稳健且信息更丰富的结果。通过捕捉关系的强度和函数形式，这些矩阵有助于发现隐藏的物理模式，并推动基于定义域知识的模型开发。

我们的时代以数据的海量增长以及机器学习模型在科学与工程领域的广泛应用为特征。这些资源为发现提供了机遇，但也带来了高维性、噪声、共线性和非线性关系等挑战，从而限制了可解释性。大型数据集还带来了巨大的计算需求，限制了分析并降低了对结果的信心。因此，各学科都需要能够高效管理复杂性，同时生成可解释且值得信赖的模型的工具[1]。

科学家和工程师在模型开发过程中面临着几项关键挑战：数据必须包含关于底层过程的充分信息[2]，质量良好[3,4]，并且输入变量与目标变量之间存在因果依赖[5,6]。如果不满足这些条件，模型往往会学到在实际应用中行不通的关系。能够支持严格数据检查的工具至关重要，它们能帮助研究人员在构建模型之前了解数据中的结构、局限性和因果信号。

在地球科学和其他领域，数据集通常包含许多变量，这些变量可以为预测另一个因变量提供信息。然而，由于相对重要性不明确，研究人员往往会在机器学习模型中使用所有输入变量，从而产生过于复杂、难以解读的结果。评估变量间关系的工具有很多，但最常见的方法（例如皮尔逊相关、主成分分析）假设关系是线性的，而其他方法（例如互信息、核方法、*t-SNE*）要么在大型数据集上计算成本过高，要么只能提供有限的洞见[7–9]。

许多工具只能揭示输入变量与目标之间的成对关联，忽略协同效应[10]，并且当未测量因素驱动这些关系时高估相关性[11]。虽然<样式 id='3'>偏相关分析</样式>有所帮助，但它仍然假设这些关系具有线性。因此，需要新的工具来理解复杂的数据生成过程（DGP）。

我们提出了一种方法，既能捕捉成对关联的强度（**图 1a**中的彩色线条，以及**图 1c**中的彩色阴影），还能捕捉它们的函数形式（ϕ 图 1a**中的函数，以及**图 1c**中的白色/灰色线条）。我们用**多变量关联矩阵（MKAN；**图 1b**和**图 1c**）补充成对分析。该矩阵综合考虑多个输入变量的关系强度和函数形式，有助于提炼出简单的关联。

成对关联（PKAN）和多变量关联（MKAN）矩阵基于近期提出的 *Kolmogorov-Arnold* 网络（KANs）[12]，用于刻画模型变量之间的关联。KAN 机器学习架构同时提升了可解释性和性能。尽管其相对于传统神经网络的性能仍在研究中，KAN 方法提供了易于解释的表示[13,14]，，与我们识别函数关系及其强度的目标一致。

在此，我们通过将 PKAN 和 MKAN 工具与两种常见的数据分析方法进行比较，展示它们的稳健性，并通过两项真实的

数据实验展示它们的可解释性。我们的目标是表明，这些工具能够促进可解释、可信赖的模型开发，以支持科学研究、提升利益相关者的信任度并辅助决策[16,17]。

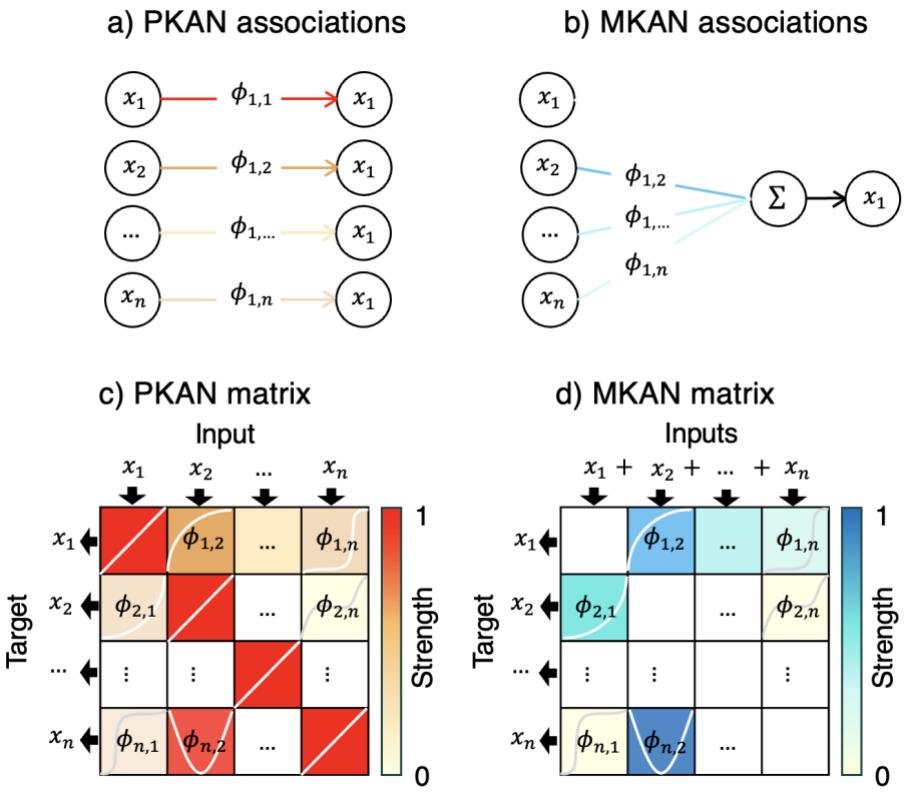

图 1| 概念示意图：成对 *KAN*（*PKAN*）和多变量 *KAN*（*MKAN*）矩阵。关联强度以背景颜色的深浅表示（0 = 无，1 = 最大），函数形式（ϕ）则由每个单元格中的白线或灰线表示。a）通过分析变量间所有可能的成对关联得出 PKAN。b）当所有非目标变量（列）用于预测一个唯一的目标变量（行）时，得到 MKAN。c）所有成对关联的矩阵表示。d）多变量关联的矩阵表示。箭头表示信息流的方向。

结果

量化函数形式在非线性关联中的作用

尽管数据编码了输入变量与目标变量之间的关联，但仅识别出关联是否存在还不够；由数学形式或函数形式定义的关系类型，对于模型开发和理解其局限性至关重要。现有方法虽然能够量化关系强度，但也存在局限。皮尔逊相关只能捕捉线性依赖[18]，，而互信息能够检测非线性关联，却无法指示其方向或形式。我们的第一项实验展示了在使用不同方法时，函数形式如何影响强度估计。

我们考察了由三个简单函数生成的数据：线性（$x_1 = x$）、二次（$x_2 = x^2$）和三次（$x_3 = x^3$），其中自变量x 从[至2,2]的均匀分布中抽取。将这三个变量x_1、 x_2 和x_3 汇集成一个“数据集”，并对它们之间的关联进行了研究。皮尔逊相关（**图 2a**）无法量化二次关系（黄色单元格）的强度；由于其对称性，它给出的线性相关强度为零。互信息（**图 2b**）不受这一局限影响，所有单元格均得到了非零值。不过，其强度会受关系形式影响：当关系单调非递减时，强度更高。

相比之下，PKAN 矩阵（**图 2c**）在使用 x_1 来预测x_2 或x_3 （二次 或三次；第一列）时，能正确识别强度（深红色）和函数形式；而使用x_3 （视为独立变量）来预测 x_1 或 x_2 （线性 或二次；第三列）时也能如此。在第二列中，由于此类映射需要满足单射性，PKAN 无法识别使用x_2 预测x_1 或x_3 时的强度和函数形式。若不将定义域划分为正区间或负区间，二次输入无法通过唯一函数确定线性或三次输出。

有趣的是，与皮尔逊相关和互信息方法不同，PKAN 矩阵（对角线）不具对称性，这表明变量之间存在非单射关系。这为所研究的关联提供了新的信息。

MKAN 矩阵（图 2d）报告了使用所有列输入变量（不包括目标变量）预测目标行变量时的特征贡献（强度）和函数形式。第一行和第三行分别通过 x_2 和x_3预测x_1 、通过 x_1 和x_2预测x_3 ，与 PKAN 的结果一致，表明x_2 （二次）对任一变量都不提供有用信息。第二行（预测 x_2）只使用 x_1 （线性），尽管可用的输入中还有 x_3 （三次），这反映了 KAN 倾向于选择最简单的函数形式。这是因为，与x_1 相比，x_3 不会为x_2 提供额外信息。由于与x_3 的关系会更复杂，因此只有在x_1 不可用时才会使用x_3 。

与传统方法相比，PKAN 和 MKAN 矩阵能够为非线性关系提供更稳健的强度估计。当关联不具单射性时，它们会揭示映射无法通过近似表示的情况。学习得到的函数形状突显了输入变量与目标变量之间敏感性较低和较高的区域，从而更细致地刻画数据集的行为。这一洞见对用户很有价值，因为其他所有经过评估的方法都无法提供这一信息。

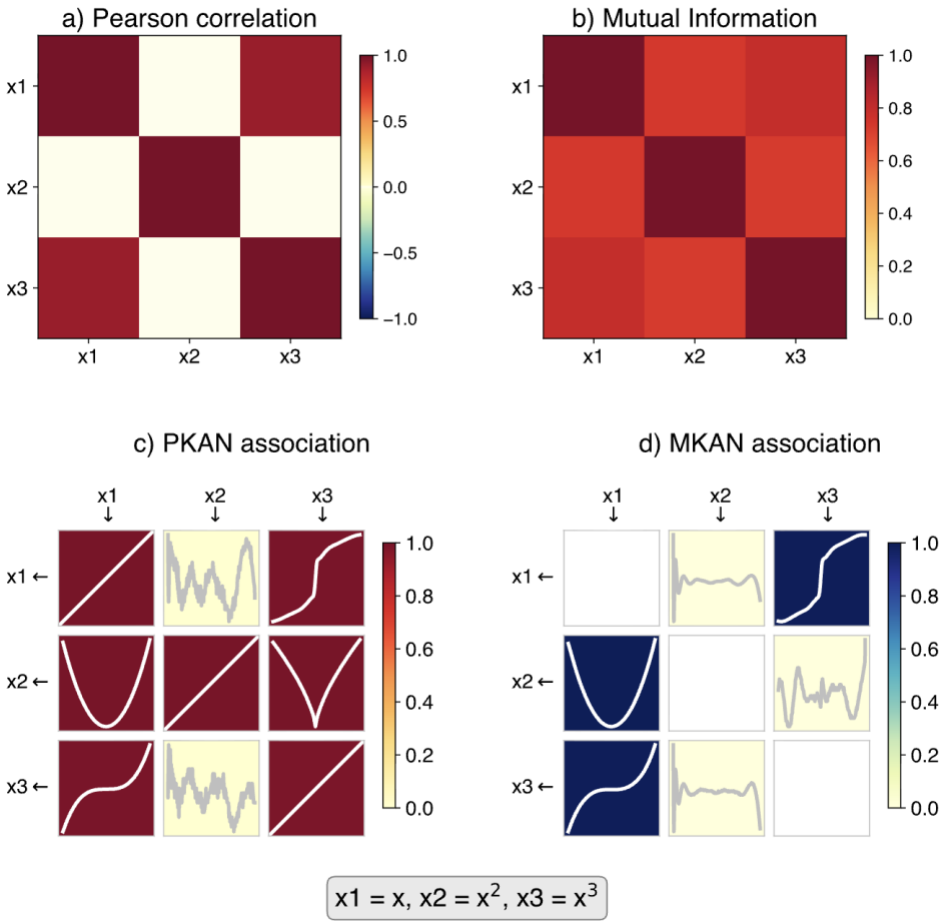

图 2| 非线性关系的各方法比较。a）皮尔逊相关无法捕捉二次结构。b）互信息能够检测非线性，但给出的强度估计不均匀。c）PKAN 和 d）MKAN 能够正确识别大多数函数形式，但由于相关关系具有非单射性质，一些二次模式未得到充分体现。

关联中的噪声效应

在许多现实场景中，测量误差会随变量的大小而增大。例如，在水文学中，洪水期间的河流流量更难测量，从而产生会扭曲关联的异方差误差。第二个实验评估了不同类型的噪声对这些方法的影响。

我们考察正态分布误差对线性关系的影响（$2x$），方法是添加标准差恒定的同方差噪声（$x_2 = 2x$ +噪声），或添加标准差可变且等于自变量的异方差噪声（$x_3 = 2x +$ 噪声（x））。随后，我们将这些变量汇总为一个数据集，并分析它们之间的关联（图 3）。

图3a显示，噪声类型（异方差与同方差）对皮尔逊相关的影响较小。相比之下，互信息（图3b）显著下降，因为误差分布的变化改变了底层数据分布（原本为均匀分布）。而 PKAN 矩阵所捕捉的强度（图3c）与皮尔逊相关相似，表明噪声类型的影响也相似。

当x_1 作为输入时（第一列），函数形式受噪声类型的影响很小。当自变量含有噪声时（第二列和第三列），退化程度更为明显。当x_2 （线性同方差）作为输入时（第二列），线性模式仍然足够清晰，可以恢复真实函数形式。然而，当 x_3 （线性异方差）作为输入变量时（第三列），恢复出的函数形式更像对数函数或平方根函数，而不是线性函数，表明在异方差误差下，关联发生了更严重的扭曲。

对于 MKAN（图 3d，第 1 行），使用 x_2 和 x_3 预测 x_1时，会结合两个含噪输入，并部分抵消它们的误差。第 2 行和第 3 行表明，MKAN 始终选择无噪声变量（x_1），这是一个理想特性。对于所有方法，尤其是 互信息，噪声都会削弱强度估计。输入中存在噪声时，噪声还会扭曲函数形式，但使用多个含噪输入可以减弱这一影响。总体而言，了解噪声类型和幅度至关重要，因为它们会影响关联学习。

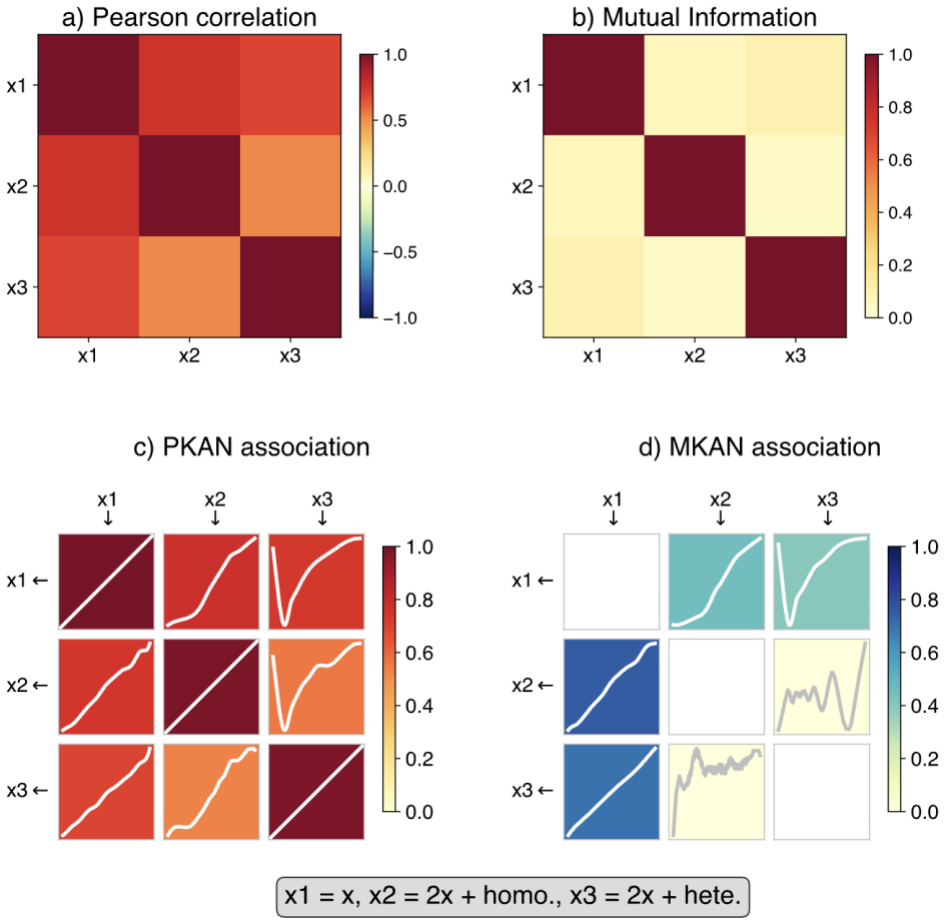

图 3| 同方差和异方差数据噪声对关联的影响。a）皮尔逊相关表现出中等程度的退化。b）互信息在误差分布发生变化时可能几乎完全消失。c-d）PKAN 和 MKAN 保持了关联强度和函数形式，但自变量中的噪声会扭曲恢复出的函数形式。

动态关联

存在滞后依赖的动力学过程十分常见，但量化起来颇具挑战。为了评估我们的方法在此类条件下的表现，我们考察了两个由均匀分布变量 x_1生成的滞后正弦示例。首先，我们生成了一个非线性但无记忆的变换（$x_2 = \sin(x_1)$）。其次，我们将 x_1 滞后 150 个单位后再进行正弦变换，以引入记忆（$x_3 = \sin(lagged\ x_1)$）。我们将 x_1、x_2、和 x_3 视为不同变量，并分析它们之间的关联（**图 4**）。

皮尔逊相关无法捕捉正弦关系的强度和滞后效应，而互信息的表现更好（**图 4a** 和**4b**）。相比之下，以x_1 为输入时，PKAN 能准确表征关系的强度（**图 4c**；第一列），并捕捉函数形式，仅在端点附近存在轻微偏差。由于正弦关系具有非单射性质，反向关联（使用x_2 或x_3 预测x_1）未能恢复；要表示这种关系，需要叠加多个单射函数。在完整的基于 KAN 的模型中，[19,20], 增加输入变量与目标变量之间的隐藏层节点数可以解决这一局限，但为了最大限度地提高可解释性，我们的简化实现为每个单元仅使用一种函数形式。MKAN（**图 4d**）同样难以在非单射情况下，根据反向关系（第一行）预测x_1；而第二行和第三行表明，在给定x_1时，模型能够较好地预测x_2和x_3。

即使存在滞后依赖，KAN 矩阵的表现仍优于传统方法。其主要局限在于逆关系具有非单射结构，因此每个单元格都需要多个函数形式。将这一能力纳入模型有待未来扩展。

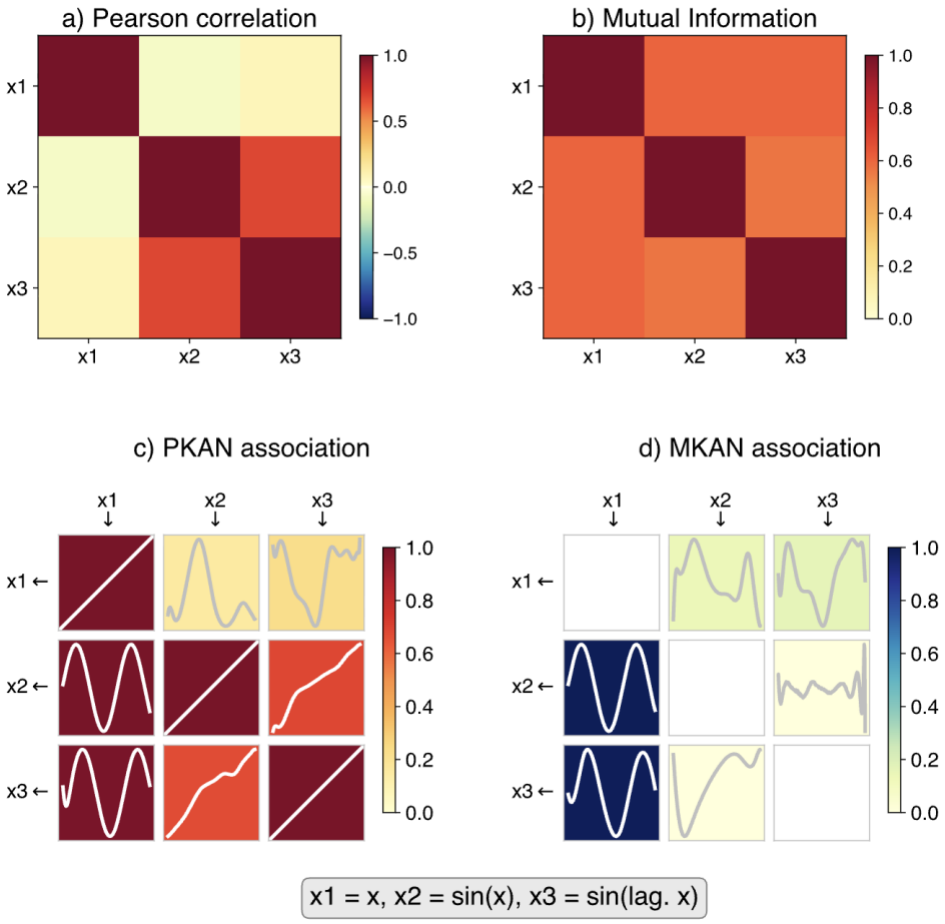

图 4| 循环和滞后关系实验。a）皮尔逊相关无法反映正弦关系或滞后正弦关系的真实强度。b）互信息仅显示出有限的改进。c-d）KAN 矩阵准确捕捉了正弦关系（第一列）的关联强度和函数形式，但当关系为非单射时则表现不佳。

水文建模中的多变量贡献强度和最佳预测变量

接下来，我们使用真实的地球科学数据，展示更准确地估计和理解强度与函数形式所带来的益处。在此，我们评估数据集中每个变量的多变量强度贡献（MKAN），以及这如何提升模型的简洁性和性能。

在第一个示例中，我们使用 CAMELS 数据集[21]，，该数据集提供了美国本土 671 个流域的动态和静态流域属性数据。这个与水文学相关的数据集被广泛用于构建和测试河川流量预测模型，并研究机器学习模型的可解释性[22–26]。静态属性分为六类：地形、气候、植被、土壤、地质学和水文学（57 项属性）。

鉴于静态属性数量众多，水文分析通常会寻求一组精简且有意义的属性，用于河川径流预测。然而，对于哪些属性最能提供信息、适合用于精简的模型开发，目前仍不明确。以往研究表明，可能只需一小部分属性[27]，，而增加属性可能导致数据驱动模型出现空间过拟合[28]。在此，我们采用多变量 MKAN 方法，识别对可预测性贡献最大的属性，并利用关联强度指导设计更精简、稳健的模型。其他属性的 KAN 矩阵见补充信息（S1）。

我们针对三个关键径流属性计算了多变量贡献（MKAN）：均值（q_{mean}）、5th 百分位数（q_5）和95th 百分位数（q_{95}）。我们分别将每个属性作为目标变量应用 MKAN，并根据贡献强度对其他属性进行排名。我们对各目标变量对应的强度取平均值，以估算每个属性对河川径流预测的重要性（基于局部 MKAN 分析的多目标分析）。我们使用 皮尔逊相关 和 互信息 生成类似的排名。为评估这些排名的实际价值，我们训练了一个机器学习模型，仅使用排名靠前的属性来联合预测这三个径流属性。我们采用了随机森林方法[29]，，该方法在小型数据集上表现良好[30]。由于百分位数值的范围跨越多个数量级，我们对数据进行了对数变换。

表 1报告了使用不同排名方法所确定的排名靠前属性（数量从 2 到 14 个）训练的随机森林模型的性能。对于任意固定数量的属性，基于 *MKAN* 的排名均优于皮尔逊相关和互信息。 基于皮尔逊相关的排序（最常见的方法）需要额外增加 2–6 个属性，才能达到与 MKAN 相同的性能。这体现了*MKAN*的非线性关系捕捉能力，说明其具有更强的表征简约性。使用排名前 12 的 MKAN 属性即可获得大部分性能提升，而其他方法仍需要更多属性。互信息的性能排名第二，这可能是因为相较于皮尔逊相关，它能更好地处理非线性关系。

表 1| 采用不同方法排名靠前的属性数量训练的随机森林模型性能（0 表示较差，1 表示完美）。

排名方法	R^2 技巧评分（水文学中的 Nash-Sutcliffe 效率）						
	前 2 名	前 4 名	前 6 名	前 8 名	前 10 名	前 12 名	前 14 名
皮尔逊相关	0.620	0.721	0.719	0.722	0.750	0.755	0.754
互信息	0.652	0.788	0.804	0.801	0.798	0.814	0.819
MKAN 贡献	**0.768**	**0.802**	**0.812**	**0.828**	**0.836**	**0.840**	**0.837**

表2报告了用于基于 MKAN 的排序的三个流量特征的属性贡献（强度）。有趣的是，贡献强度最高的属性各不相同，具体取决于流量属性。例如，无降水日频率和平均坡度对q_5~~有~~贡献，但对q_{mean} 或q_{95}, 没有贡献；而平均降水量和森林覆盖率

对q均值和q_{95}的贡献更大，但对q_5的贡献较小。此外，q_{mean}和q_{95}所需的特征数量少于q_5（分别为3和5个，而后者为12个），这表明预测较低径流值的复杂度更高。“最佳”CAMELS属性的选择取决于具体目标；因此，任何特征选择任务都应考虑整个径流分布的变化。

表 2| 用于预测径流属性的 MKAN推导的属性贡献（CAMELS 数据集）。

属性	q_5	q_{mean}	q_{95}	平均值	排名
日平均降水量	0.109	0.623	0.526	0.419	1
降雪降水占比	0.080	0.182	0.144	0.135	2
森林覆盖率	0.033	0.129	0.077	0.080	3
无降水日频率（<1 毫米/天）	0.129	0.000	0.000	0.043	4
降水的季节性与时间分布	0.000	0.001	0.104	0.035	5
流域平均坡度	0.097	0.000	0.000	0.032	6
砂粒含量	0.068	0.000	0.000	0.023	7
流域平均海拔	0.023	0.000	0.042	0.022	8
月平均值最大值与最小值之差 叶面积指数	0.058	0.000	0.000	0.019	9
地下渗透率	0.055	0.000	0.000	0.018	10
黏土含量	0.054	0.000	0.000	0.018	11
绿色植被覆盖率的月平均值最大值	0.050	0.000	0.000	0.017	12
月平均值的最大值与最小值之差 绿色植被覆盖率	0.050	0.000	0.000	0.017	13
土壤深度	0.035	0.000	0.000	0.012	14
总强度贡献	**0.841**	**0.935**	**0.893**	**0.890**	

这一结果展示了 MKAN 多变量贡献分析在变量选择方面更出色的性能和更高的简约性。它还有助于更好地理解河川径流预测所涉及的复杂性，并促进复杂任务中的降维。

强度和函数形式：理解流体动力学关系的工具

流体动力学难以理解，即使已知其控制偏微分方程[31,32]也是如此。因此，全面了解主导变量及其过程关系至关重要。本文采用分阶段的方式应用基于KAN的框架，通过评估并移除变量，逐步揭示不同层次的特征重要性。

本实验分析的模型生成数据，源自*España*等人[33] 提出的基准情景，其中模拟数据由一个针对美国亚利桑那州大理石峡谷内科罗拉多河河段的数值模型生成。具体而言，该模型采用大涡模拟（LES）框架解析湍流场[33–35]。尽管

求解了纳维–斯托克斯方程，但连接静态数据（x、y、z、时间）与动态模型输出（速度场（U_x、U_y、U_z）、压力（p）和湍流（涡）黏度（nut）之间的函数关系并不明确。此外，还纳入了动态变量此前 50 个时间步的滞后值，以考察系统过去的动态如何影响当前行为。

该*PKAN* 矩阵（**图 5a**）显示许多变量具有较高的强度。例如，动态变量（U_x、U_y、U_z、p、nut）与坐标（x,y）、滞后数据以及变量自身存在关联，这是由变量的内部依赖所致。然而，无法区分各项关联的贡献。相比之下，MKAN 矩阵（**图 5b**）显示，纳入滞后模拟数据后，大多数成对强度都会减弱，这表明当前状态与过去状态高度相似。变量与其滞后对应变量之间呈现的线性函数形式（例如，$U_x = f(U_{x_lagged})$）表明存在持续效应，因此需要移除滞后模拟数据，以揭示更多关联。

如**图 5c**所示，移除滞后模拟数据后，x 和 y 坐标对U_x 和U_y的贡献占主导地位。由于这些坐标缺乏物理意义且保持不变，因此，均值等空间均匀分量可能有助于表征这些关系。随后，从速度、压力和湍流黏度中减去均值（从而创建新变量：U_x''、U_y''、U_z''、p''和nut''）。

去除均值（**图 5d**）后，除了压力之外，几乎所有残余关联都消失了。压力增量场（p''）的行为值得关注，因为它与时间变量的强关联超过了整体空间分布的影响。尽管所有变量都受纳维–斯托克斯方程约束，这种时间交互作用却是压力增量所独有的。这些结果表明，速度场主要由空间结构表征，而压力则表现出更显著的时间演化效应。以序列方式通过 KAN 矩阵可视化并理解非线性强度，有助于更深入地认识如何控制关联。

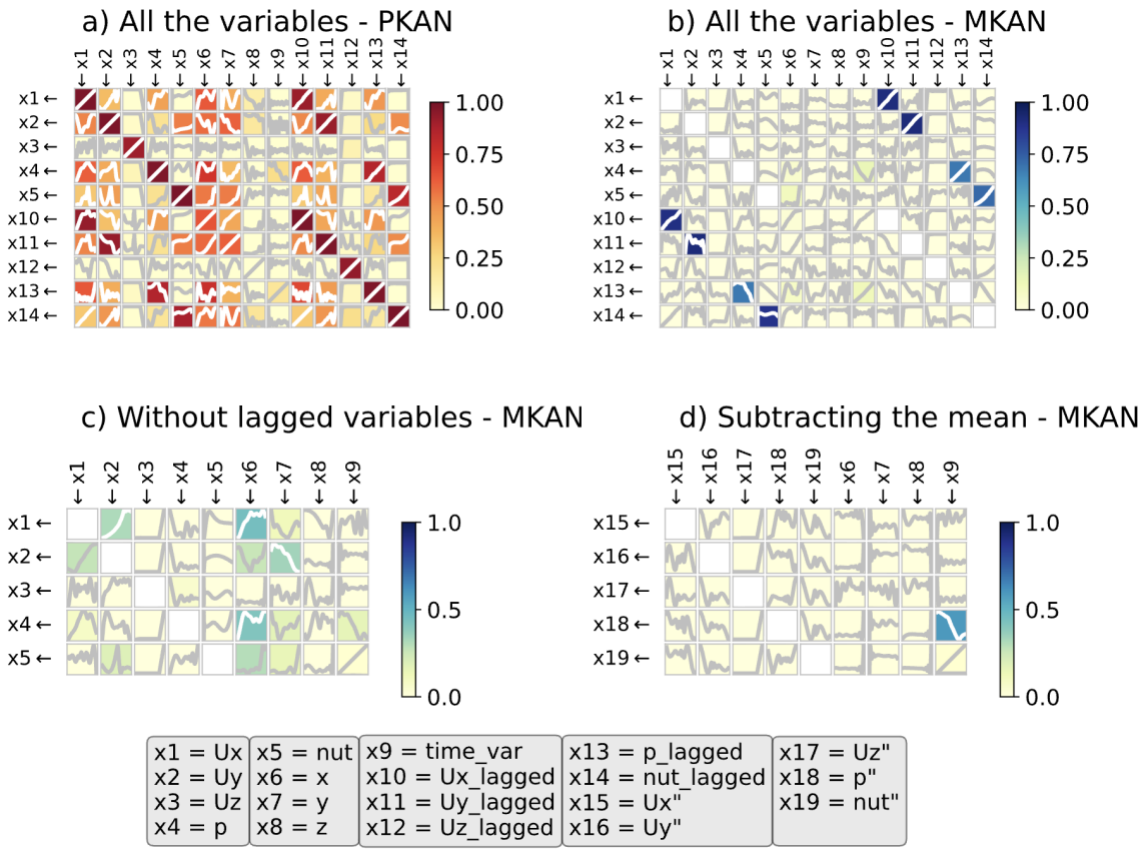

图 5| 用于计算流体动力学的 KAN 矩阵（非因果变量 x 、y 、 z 、和 时间_v a r 未作为目标变量）。a）PKAN 揭示了多个变量之间的强关联。b）MKAN 表明，滞后的模拟数据通过一种平凡解主导了关联。c）MKAN 突显了 x 和 y 坐标对速度分量的重要贡献。d）MKAN 显示，经过均值减法后，大多数与速度相关的贡献都消失了，这反映出底层空间模式的持续存在。

讨论与结论

机器学习模型可能缺乏可解释性，并呈现"黑箱"行为。这促使人们开发能够更清晰地呈现数据-模型关系的工具。然而，在预处理和后处理过程中，复杂的非线性、噪声和多变量关联仍可能对可解释性构成挑战。KAN架构能够提取并可视化变量之间的函数形式。我们利用这一特性提出成对（PKAN）和多变量（MKAN）矩阵，它们使用简单的 KAN网络来表征变量之间的关系强度和函数形式。

皮尔逊相关和互信息等传统方法广泛用于评估依赖关系，但无法提供关系形式的信息，而且往往难以应对非线性关系和噪声。我们的 KAN 方法在此类

条件下更为稳健。此外，它还能通过交互强度的减弱识别非单射关系，这是其他方法所不具备的能力。

PKAN 和 *MKAN* 从互补的角度审视同一组数据。构建数据驱动模型时，会收集各种信息来源，以便与目标变量建立较强的关联（在成对矩阵中以红色像素表示）。然而，并非每种交互都能为预测提供独特信息。*MKAN* 会基于特定输入变量找出主导预测的简单交互。这种方法不同于典型的机器学习；后者通常会对所有关联赋予权重（不进行正则化），以实现最佳性能。

因此，基于 KAN 的分析有助于在高维空间中确定特征优先级，正如 CAMELS属性优化实验所示。除了简化模型和提升可解释性之外，剪枝还能提高计算效率，因此在大规模或资源受限建模中很有价值。因此，*KAN* 矩阵可作为预处理工具，引导科学家构建更简单、更透明的数据驱动表示。此外，正如计算流体动力学实验所示，KAN 矩阵有助于对主导机制往往不明确的复杂系统开展诊断性后处理分析，从而增进理解。

最后，要理解任何数据生成过程，对关联强度和函数形式进行双重刻画至关重要。稳健地从数据中估计关系形式和强度，同时揭示信息流的方向与结构，是一项普遍的科学挑战，其适用范围远不止水文学和流体动力学。此外，了解函数形式有助于揭示隐藏知识，同时还能与现有领域知识进行一致性检验。基于这些原因，KAN矩阵分析在提升机器学习的可解释性、深化不同定义域中的科学理解方面展现出巨大潜力。

方法

KAN架构

Liu 等人[19] 将 KAN 架构作为*Kolmogorov-Arnold*定理（公式 1[36]）的扩展提出。该定理指出，任意复杂函数 $y = f(x_1, \ldots x_n)$ 都可以分解为其输入变量 x_j 的非线性单变量变换 $\phi_{i,j}$ 的加性组合，再对这些分量进行进一步的非线性单变量变换 ψ_i 并作加性组合。

$$f(x) = \sum_{i=1}^{2n+1} \psi_i \left[\sum_{j=1}^{n} \phi_{i,j}(x_j) \right], \quad (1)$$

该定理并未规定单变量转换函数 $\phi_{i,\ j}$ 和 ψ_i的具体函数形式，只说明它们必须存在。*Liu* 等人 [19] 提出使用样条近似来学习这些转换，并通过引入分层组合结构扩展了该方法的能力，从而构建更深的网络。 *Liu* 等人[20]

进一步通过添加乘法选项和用于科学发现的其他特征，增强了该方法。

KAN 矩阵

我们使用一种简单的 KAN架构，仅实现第一层转换（即输入与目标层之间只有一个 $\phi_{i,\ j}$ 函数）。这种表示将输入变量视为在信息上相互独立，同时允许输入变量与目标变量之间存在非线性关系。

由于 KAN 表示输入输出关系，因此可以任意定义用于预测目标的输入变量数量，从而评估每个输入变量与目标变量之间任何可能的局部关联。我们通过构建一个由多个元素组成的矩阵来直观呈现这一点，其中每个元素表示一个输入变量与一个目标变量之间成对关联的性质（强度和函数形式）（不含 ψ_i 函数）。在这种情况下，**方程 1** 可简化为**方程 2**。绘制每个输入变量 x_j 与目标变量 x_i之间的函数$\phi_{i,\ j}$，并利用 KAN 库[37], 提供的关联强度，即可刻画一种独特的成对关联。对输入变量和目标变量的所有可能组合执行此分析，即可生成一个包含所有可能关联的矩阵（**图 1c**）。每一行与每一列的交点刻画了关系的函数特性$\phi_{i,\ j}$，而背景颜色的深浅则表示关联强度。

$$f_i(x_j) = \phi_{i,j}(x_j), \tag{2}$$

图 1c突显了两个关键性质。首先，对角线具有最大强度（与完全关联一致），因为输入变量与目标相同。其次，矩阵不一定是对称的（$\phi_{i,\ j} \neq \phi_{j,\ i}$），因为函数关系可能不是单射。在极端情况下，函数 $\phi_{i,\ j}$ 可能存在，但函数 $\phi_{j,\ i}$可能不存在。

成对关联矩阵简单易懂，也便于可视化，但没有考虑变量之间的冗余信息。一个自然的扩展是，在函数表示中纳入所有相关变量，以预测所选目标。在这种情况下，**图 1d**中除目标之外的所有列输入变量都会预测行目标。因此，**方程 1**可简化为**方程 3**。

$$f_i(x_{i\neq j} j = 1, \ldots, n) = \sum_{j=1}^{n} \phi_{i,j}(x_j), \tag{3}$$

在此可视化中，函数形式和强度的含义与**图 1c**相同，但它们是在所有输入变量在相互独立（不包括自身）的条件下预测目标时获得的。在这种情况下，对角线不会填入恒等函数，因为未必能实现完美关联。相反，它表示观测目标与使用**方程 3**预测的目标之间的关系。只有在完美拟合时才会出现恒等函数。我们将此矩阵称为多变量贡献 *KAN*（或*MKAN*），因为它刻画了多个变量在与目标的学习关联中的强度。

关联的强度（以背景颜色的深浅表示）由*Liu* 等人[20]提出的属性得分定义。他们考察边激活结果的标准差与连接节点的标准差之比。该比值定义了边对节点的贡献。然而，它并未考虑整个网络。因此，他们通过迭代计算该比值，从输出层向输入层移动，并统计特定边前后的有效连接性。与*Liu* 等人[19]使用的原始 L1 范数相比，该方法能更准确地估算特征贡献。

本文将*Liu* 等人[20] 的属性得分归一化，确保其取值范围为 0–1，以便在 KAN 矩阵中绘图。然而，这并未考虑所得关联的预测强度。为此，我们将属性得分乘以一个用于表征预测强度的性能指标。我们测试了两种指标：技能评分版本的*Kling-Gupta* 效率（KGE[38]）[39]，以及R^2 技巧评分（在水文学中称为*Nash- Sutcliffe* 效率，NSE[40]）。用户可以选择自己偏好的指标，从而使关联强度更易于解释。在我们的实验中，两种指标都经过缩放，使得 0.0 表示始终使用目标均值作为预测值，1.0 表示始终通过一个或多个输入变量实现完美预测。我们还采用最大-最小归一化，以避免输入变量因其量级而在两个矩阵中占据不成比例的重要性。

实验详情

为增进对所设计实验的理解，本文介绍了有关实验构思的一些细节。

- 在可视化互信息（MI）时，用于绘制背景颜色的归一化版本不一定具有对称性。将互信息除以目标熵或输入熵，能够揭示变量的复杂性。例如，在维恩图中，如果一个集合 X 包含于另一个集合Y之中，那么对于被包含的集合，归一化互信息为 $NMI(X,Y) = MI(X,Y)/H(Y) = 1$；对于包含其他集合的集合，则为 $NMI(X,Y) = MI(X,Y)/H(X) < 1$，因此这种归一化并不对称。本文采用输入熵与目标熵的平均值（$H(X) + H(Y))/2$ 作为归一化基准，由此得到对称的归一化互信息 $NMI(X,Y) = 2 \cdot MI(X,Y)/(H(X) + H(Y))$。也可以采用非对称归一化，但这些结果并未为本文讨论增添价值，因此将其纳入补充信息。

- 这些实验中用于生成合成数据的数学表达式列于 **表 3**。

表 3| 用于模拟合成数据中不同类型行为的表达式（注意，服从正态分布的随机样本表示为 随机(μ, σ)，其中 μ 表示均值，σ 表示标准差）。

行为实验	表达式
非线性关系	x = (-2,2) 均匀分布 y = x^2 + 随机(0, 0.1) z = x^3 + 随机 (0, 0.1)
异方差性	x = (0,10) 均匀分布 y = 2x + 随机 (0, x) z = 2x + 随机 (0, 5.0)
滞后关系	x = (0,10) 均匀分布 y = sin(x) + 随机 (0,0.1) z = sin(shift = 150) + 随机 (0,0.1)

- 随机森林模型所用的数据划分为：80%用于训练，20%用于测试。所有排序方法均采用相同的模型架构，并使用固定的随机种子进行随机化。这些超参数是该架构所采用的标准值；由于本研究旨在比较方法而非追求最佳性能，因此未对这些超参数进行优化。超参数见 **表4**。

表 4|随机森林模型所用的超参数。

超参数	数值
估计器数量	100
最小样本分裂数	4
叶节点最小样本数	3
最大特征数	0.7
最大样本数	0.7
自助采样	True
随机种子	42

数据可用性

用于生成图表和表格的 Jupyter 笔记本可在 GitHub（https://github.com/ldelafue/KAN_matrix）上免费下载。Python 代码可从同一代码仓库克隆。

致谢

本材料基于以下项目资助的研究成果：美国国家科学基金会（NSF）授予的第 2239550 号 CAREER 奖、美国陆军研究办公室授予的 W911NF-24-1-0296 号资助，以及美国国家海洋和大气管理局（NOAA）合作科学中心教育伙伴计划面向少数族裔服务机构（MSI）的资助

NA22SEC4810016。本文档中的观点和结论仅代表作者意见，并不反映 NSF、美国陆军研究办公室或 NOAA 的观点。

参考文献

1. Reichstein, M. 等 面向数据驱动地球系统科学的深度学习与过程理解。《自然》《自然》 **566**, 195–204 (2019). 2. Karpatne, A., Ebert-Uphoff, I., Ravela, S., Babaie, H. A. & Kumar, V. 地球科学机器学习：挑战与机遇。《IEEE 知识与数据工程汇刊》《*IEEE* 知识与数据工程汇刊》 **31**, 1544–1554 (2019). 3. Venkat, G., Apon, A. & Ding, J. 大数据和机器学习的数据质量考量：超越数据清理与转换。《软件进展国际期刊》《软件进展国际期刊》 **10**, 1–20 (2017). 4. Mohammed, S. 等 数据质量对表格数据机器学习性能的影响。《信息系统》《信息系统》 **132**, 102549 (2025). 5. Schölkopf, B. 机器学习中的因果关系。载于《概率与因果推断》《概率与因果推断》(编者 Geffner, H., Dechter, R. & Halpern, J. Y.) 765–804 (美国计算机协会, New York, NY, USA, 2022). doi:10.1145/3501714.3501755. 6. Guo, R., Cheng, L., Li, J., Hahn, P. R. & Liu, H. 基于数据学习因果关系综述：问题与方法。《ACM计算调查》《*ACM*计算调查》 **53**, 1–37 (2021). 7. Kraskov, A., Stögbauer, H. & Grassberger, P. 互信息估计。《物理评论E》《物理评论*E*》 **69**, 066138 (2004). 8. Zhu, Y. & Ting, K. M. 利用隔离核提高随机邻域嵌入的有效性与效率。《人工智能研究杂志》《人工智能研究杂志》 **71**, 667–695 (2021). 9. Meyer, B. H., Pozo, A. T. R. & Nunan Zola, W. M. 面向大规模应用、保持全局和局部结构的GPU t-SNE方法。《应用中的专家系统》《应用中的专家系统》 **201**, 116918 (2022). 10. Stramaglia, S., M Cortes, J. & Marinazzo, D. 动力网络格兰杰因果分析中的协同与冗余。《新物理学杂志》《新物理学杂志》 **16**, 105003 (2014). 11. Fan, J., Han, F. & Liu, H. 大数据分析的挑战。《国家科学评论》《国家科学评论》 **1**, 293–314 (2014). 12. Liu, Z. 等 KAN：Kolmogorov–Arnold 网络。载于《第十三届国际学习表征会议》第十三届国际学习表征会议 (2025). 13. Panczyk, N. R., Erdem, O. F. & Radaideh, M. I. 揭开人工智能黑箱：利用柯尔莫哥洛夫–阿诺德网络进行符号回归，应用于先进能源应用。《能源与人工智能》《能源与人工智能》 **22**, 100595 (2025). 14. Barašin, I., Bertalanič, B., Mohorčič, M. & Fortuna, C. 探索用于可解释时间序列分类的Kolmogorov-阿诺德网络。预印本，网址：https://doi.org/10.48550/ARXIV.2411.14904 (2024).

15. Panczyk, N. R., Erdem, O. F. & Radaideh, M. I. 揭开黑箱：利用柯尔莫哥洛夫-阿诺德网络进行符号回归，应用于能源领域。预印本，网址：https://doi.org/10.48550/ARXIV.2504.03913 (2025). 16. Rudin, C. 停止为高风险决策解释黑箱机器学习模型，改用可解释模型。《自然·机器智能》《自然·机器智能》 **1**, 206–215 (2019). 17. Lipton, Z. C. 模型可解释性的神话。《美国计算机协会通讯》《美国计算机协会通讯》 **61**, 36–43 (2018). 18. Kinney, J. B. & Atwal, G. S. 公平性、互信息与最大信息系数。《美国国家科学院院刊》《美国国家科学院院刊》 **111**, 3354–3359 (2014). 19. Liu, Z. 等 KAN：柯尔莫哥洛夫-阿诺德网络。预印本，网址：https://doi.org/10.48550/ARXIV.2404.19756 (2024). 20. Liu, Z., Ma, P., Wang, Y., Matusik, W. & Tegmark, M. KAN 2.0：柯尔莫哥洛夫-阿诺德网络与科学相遇。预印本，网址：https://doi.org/10.48550/ARXIV.2408.10205 (2024). 21. Addor, N., Newman, A. J., Mizukami, N. & Clark, M. P. CAMELS数据集：面向大样本研究的流域属性与气象数据。《水文与地球系统科学》《水文与地球系统科学》 **21**, 5293–5313 (2017). 22. Kratzert, F., Klotz, D., Brenner, C., Schulz, K. & Herrnegger, M. 使用长短期记忆（LSTM）网络进行降雨-径流建模。《水文与地球系统科学》《水文与地球系统科学》 **22**, 6005–6022 (2018). 23. Feng, D., Fang, K. & Shen, C. 利用结合大陆尺度数据的长短期记忆网络增强流量预报并提取洞见。《利用大陆尺度数据整合的长短期记忆网络增强流量预报并提取洞见》。《水资源研究》《水资源研究》 **56**, e2019WR026793 (2020). 24. Feng, D., Liu, J., Lawson, K. & Shen, C. 可微、可学习、区域化且具有多物理量输出的过程模型可达到最先进的水文预测精度。《具有多物理量输出的可微、可学习、区域化过程模型可达到最先进的水文预测精度。《水资源研究》 **58**, e2022WR032404 (2022). 25. De La Fuente, L. A., Ehsani, M. R., Gupta, H. V. & Condon, L. E. 探索水文系统的可解释LSTM建模。《水文与地球系统科学》《水文与地球系统科学》 **28**, 945–971 (2024). 26. Wang, Y. & Gupta, H. V. 利用质量守恒感知器进行可解释的物理概念型流域尺度水文建模。《水资源研究》《水资源研究》 **60**, e2024WR037224 (2024). 27. Yang, Y. & Chui, T. F. M. 利用潜因子模型对流域和水文模型进行画像与配对。《水资源研究》《水资源研究》 **59**, e2022WR033684 (2023). 28. Heudorfer, B., Gupta, H. V. & Loritz, R. 水文学中的深度学习模型具备实体感知能力吗？《地球物理研究快报》《地球物理研究快报》 **52**, e2024GL113036 (2025). 29. Breiman, L. 随机森林。《机器学习》《机器学习》 **45**, 5–32 (2001). 30. Biau, G. & Scornet, E. 随机森林导览。预印本，网址：https://doi.org/10.48550/ARXIV.1511.05741 (2015).

31. Pierrehumbert, R. T. 流体力学：典型的复杂系统。《流体力学杂志》 **938**, F1（2022年）。32. Sharma, S., Marcucci, G. & Mahmud, A. 流体力学的复杂性视角。预印本：https://doi.org/10.48550/ARXIV.2212.00153（2022年）。33. España, R. E., Alvarez, L. V. & Samarasinghe, J. T. 将网格独立性研究应用于使用分离涡模拟（DES）技术，对科罗拉多河大理石峡谷河段进行的现场尺度计算流体动力学（CFD）模型研究。《地表过程与地貌》 *50*, e70030（2025年）。34. Alvarez, L. V., Schmeeckle, M. W. & Grams, P. E. 一种分离涡模拟模型，用于研究大型峡谷束缚河流沿线的侧向分离区。《地球物理研究杂志：地表》**122**, 25–49（2017年）。35. Alvarez, L. V. & Grams, P. E. 一种涡解析数值模型：利用分离涡模拟研究现场尺度侧向分离区中的湍流、沉积物与河床演变。《地球物理研究杂志：地表》 **126**, e2021JF006149（2021年）。36. Kolmogorov, A. 关于通过较少变量的连续函数的叠加表示多个变量的连续函数。美国数学会（1961年）。37. Liu, Z. PyKAN。GitHub（2024年）。38. Gupta, H. V. & Kling, H. 关于均方误差和纳什–萨特克利夫效率类指标的典型取值范围、敏感性与标准化。《水资源研究》 **47**, 2011WR010962（2011年）。39. Knoben, W. J. M., Freer, J. E. & Woods, R. A. 技术说明：基准与否？纳什–萨特克利夫效率与 Kling–Gupta 效率评分的比较。《水文与地球系统科学》*23*, 4323–4331（2019年）。40. Nash, J. E. & Sutcliffe, J. V. 通过概念模型进行河流流量预报，第一部分——原理讨论。《水文学杂志》 **10**, 282–290（1970年）。

作者贡献

L.D. 设计了这项研究，编写了工具代码，开展了分析，并撰写了论文初稿。H.G.、H.M. 和 L.A. 提出了基础性构想，并积极参与了论文的编辑和修订。L.A. 提供了湍流模型输出模拟，L.A. 和 H.M. 负责管理该研究的资助文件。

补充信息

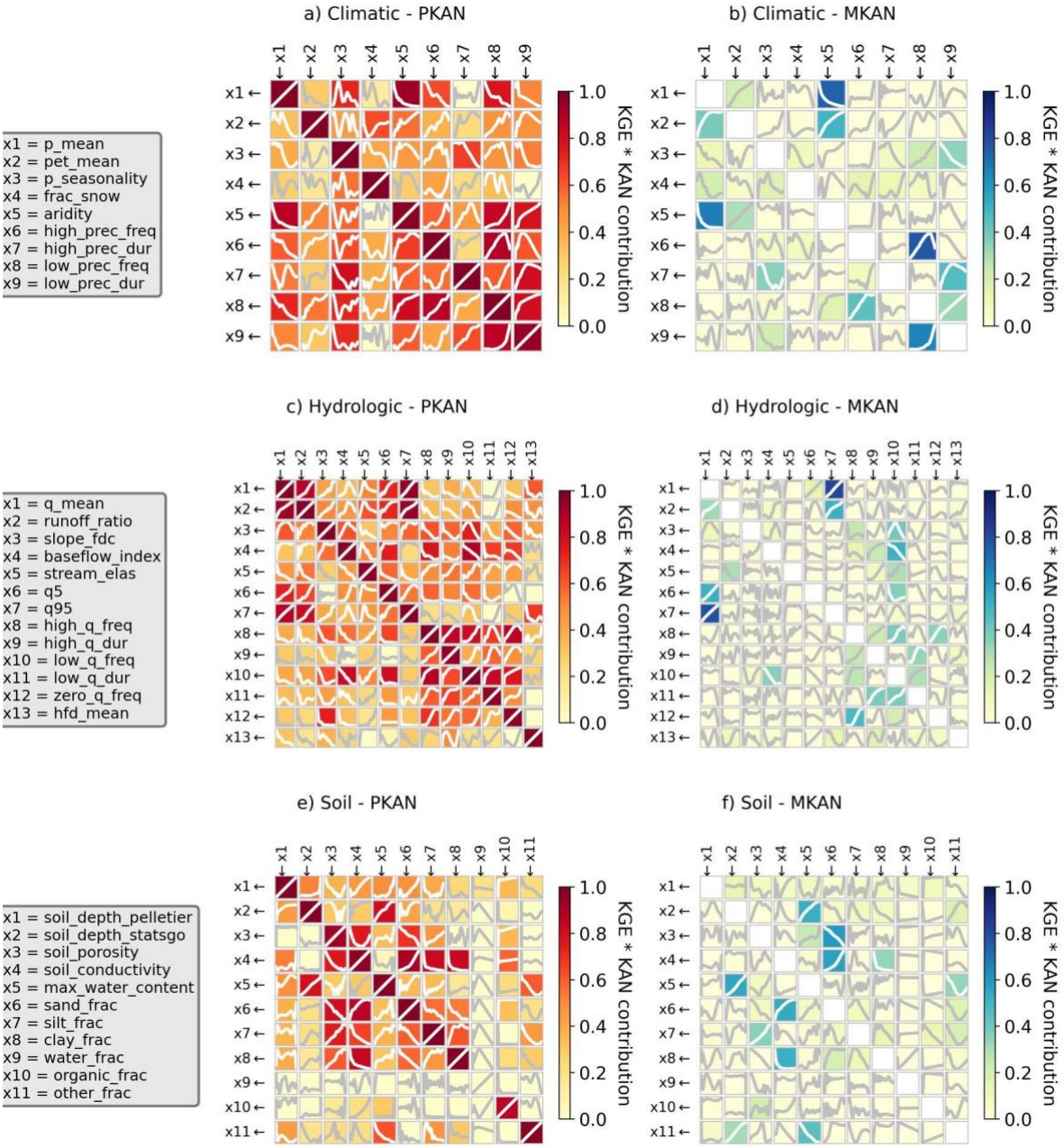

图 S1| P针对 CAMELS 各属性类别的 KAN 和 MKAN。PKAN（左列）显示出许多强成对关联，表明存在大量冗余。MKAN（右列）则揭示，只有少数属性在多变量环境中具有显著贡献，凸显出信息量最大且无冗余的预测变量。

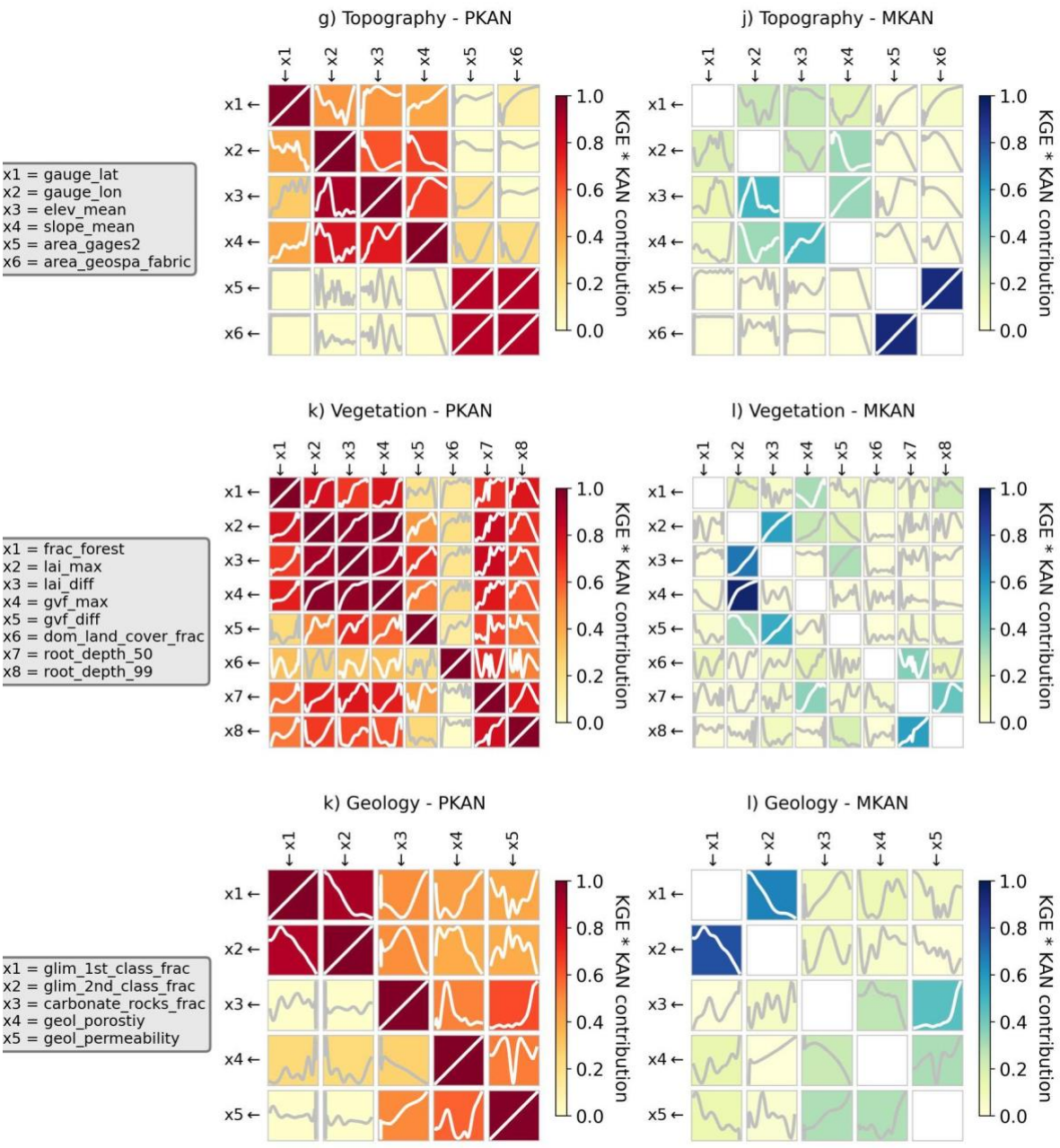

图 S1（续）| P针对 CAMELS 的各个属性类别的 KAN 和 MKAN。PKAN（左列）显示出许多较强的成对关联，表明存在大量冗余。MKAN（右列）揭示，只有一小部分属性在多变量环境中具有实质性贡献，凸显了信息量最大且冗余最少的预测变量。