多维数据可视化研究

周亮

北京大学 健康医疗大数据国家研究院
NATIONAL INSTITUTE OF HEALTH DATA SCIENCE AT PEKING UNIVERSITY

为什么研究多维数据可视化?

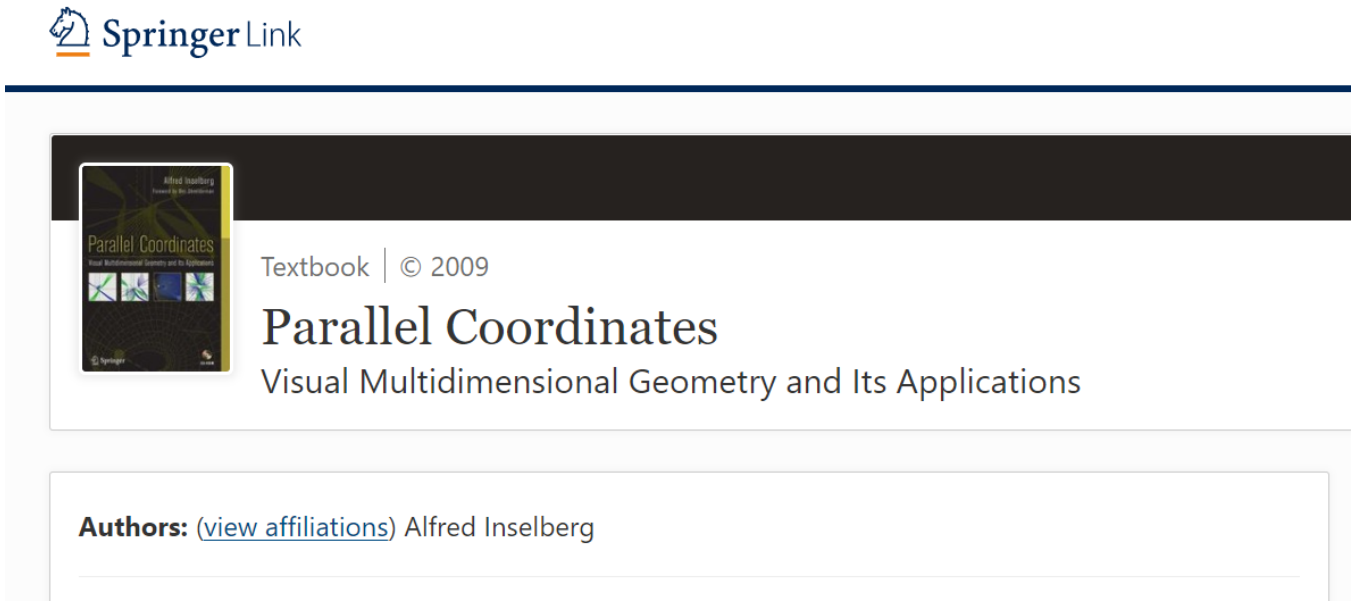
- 可视化研究中的基本问题之一
- 在信息可视化和科学可视化中都很常见
- 所在研究中心对包括多维可视化在内的各种基础问题有研究基础

最初的研究动机

- 观察1：多维数据分析和可视化多集中在
  - 单个数据点
  - 聚类
  - 全局统计信息

  但局部邻域信息分析和可视化研究不多
- 观察2：平行坐标系是常用多维数据可视化手段，但
  - 各种可视化变种多着重对线段跟踪做文章
  - 可视化领域只使用了平行坐标系专著中几页的内容

平行坐标系研究

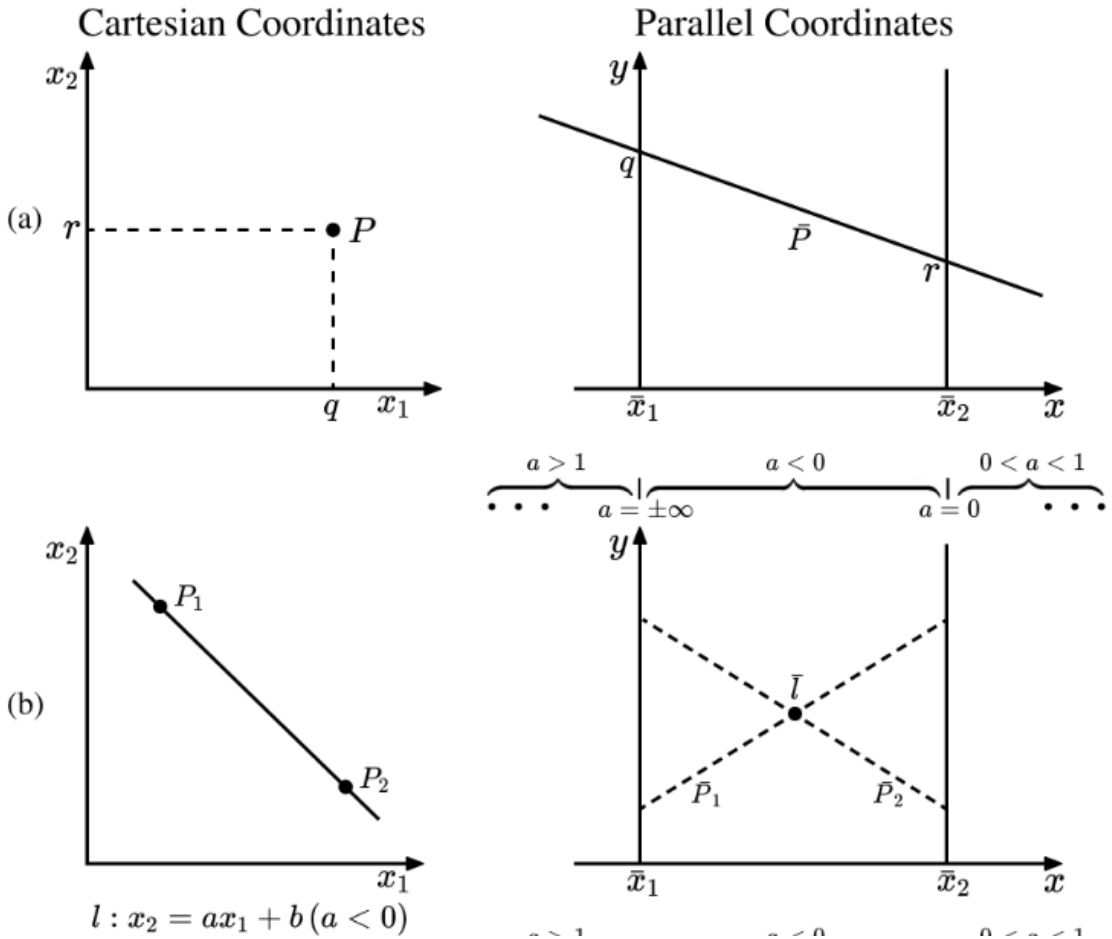


平行坐标系专著554页

可视化领域只基于书中很少一部分

平行坐标系基础：点线对偶

- 笛卡尔坐标系←→ 平行坐标系
- 绝大部分应用：
- 笛卡尔系点→平行坐标系线
- 平行坐标系点→笛卡尔系？



平行坐标系基础：点线对偶

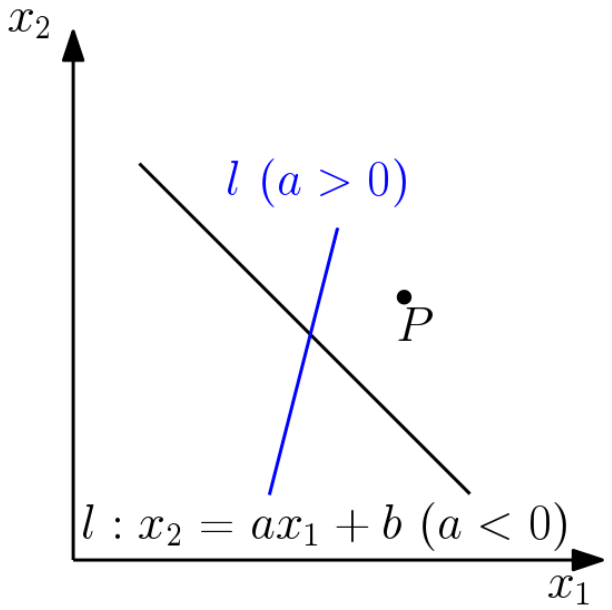
- **平行坐标系点（索引点）对应笛卡尔系中的线性关系 – 此对应未被很好利用**
- ***p*-flat**: a generalized flat surface of dimension *p*
- **Indexed Point**: a point representation of its corresponding *p*-flat within the space of the parallel coordinates

- Equation of 1-flats

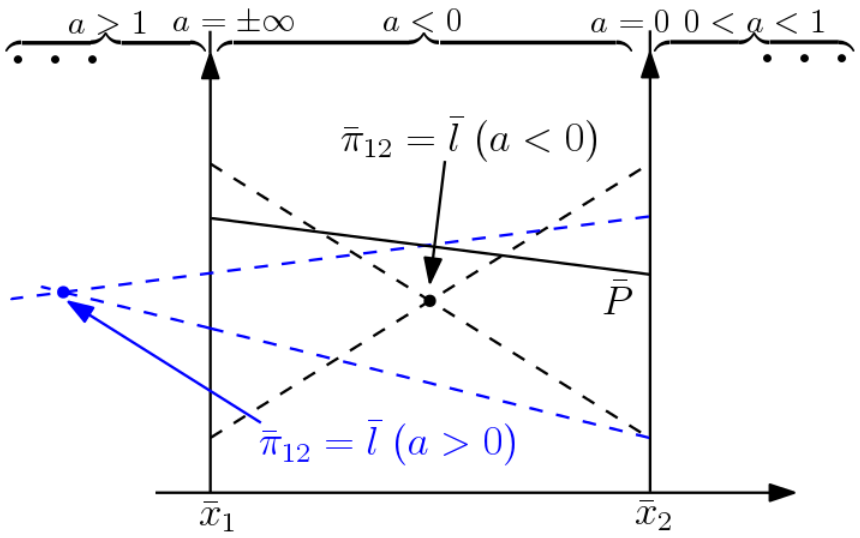
$$l : x_2 = ax_1 + b$$

- The indexed point

$$\bar{l} = \left(\frac{1}{1-a}, \frac{b}{1-a} \right), a \neq 1 .$$



Data Domain



Parallel Coordinates

平行坐标系索引点对应多维笛卡尔系中的线性结构

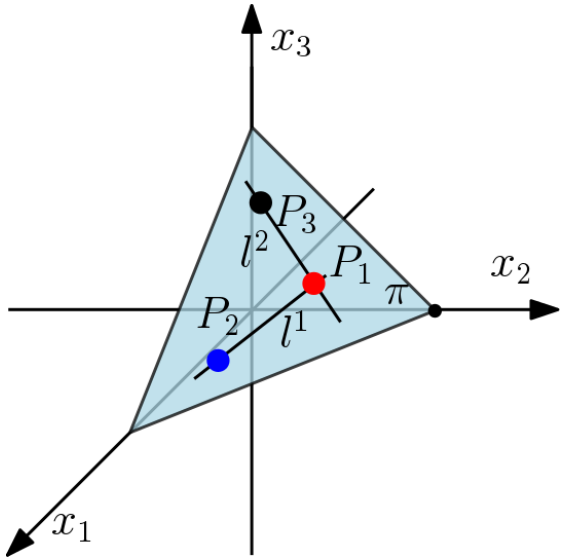
- The implicit equation of 2-flats

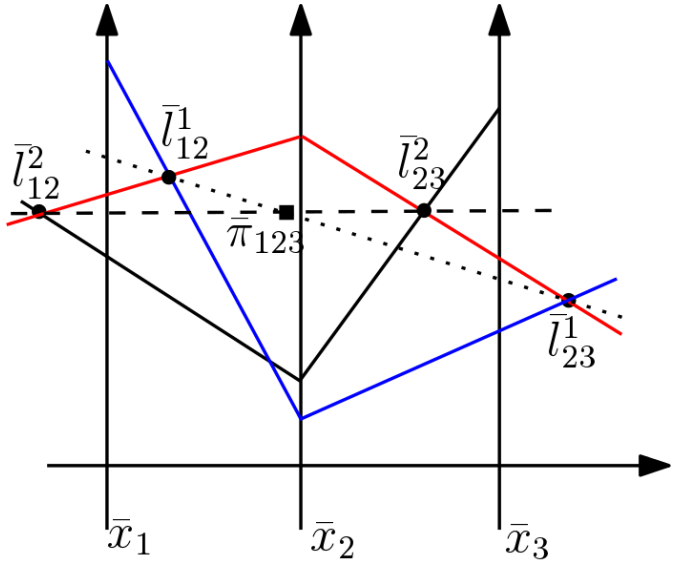
$$\pi : c_1x_1 + c_2x_2 + c_3x_3 = c_0$$

- The first indexed point

$$\bar{\pi}_{123} = \left(\frac{c_2 + 2c_3}{c_1 + c_2 + c_3}, \frac{c_0}{c_1 + c_2 + c_3} \right)$$

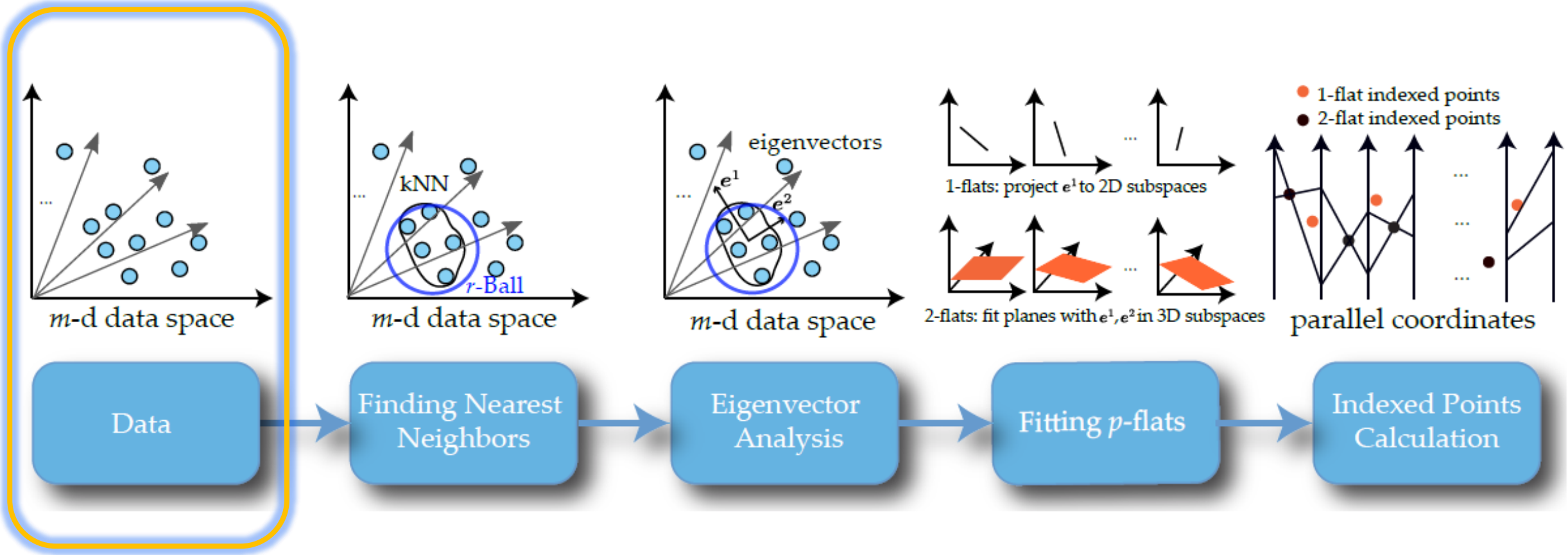
- Recursive construction from lower dimensional *p*-flats





多变量局部关系与平行坐标系索引点结合?

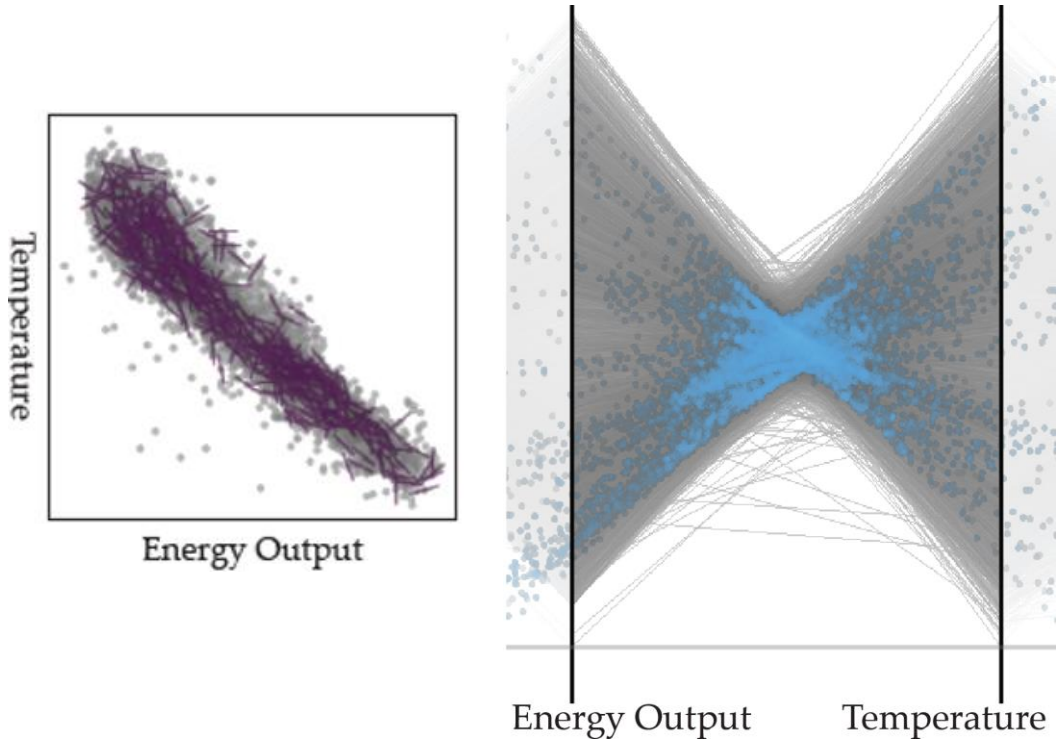
- The work flow of our method



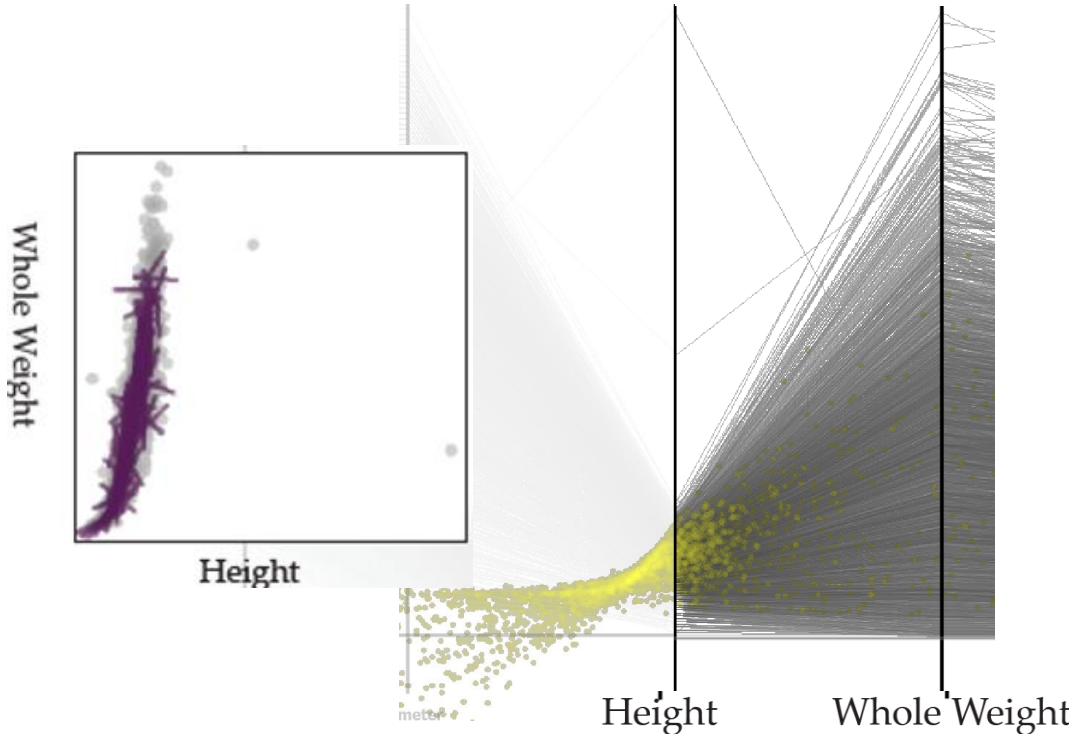
L. Zhou and D. Weiskopf, "Indexed-Points Parallel Coordinates Visualization of Multivariate Correlations," in *IEEE Transactions on Visualization and Computer Graphics*, doi: 10.1109/TVCG.2017.2698041.

Visualization of Indexed Points

- Visual signatures of negative/positive correlations



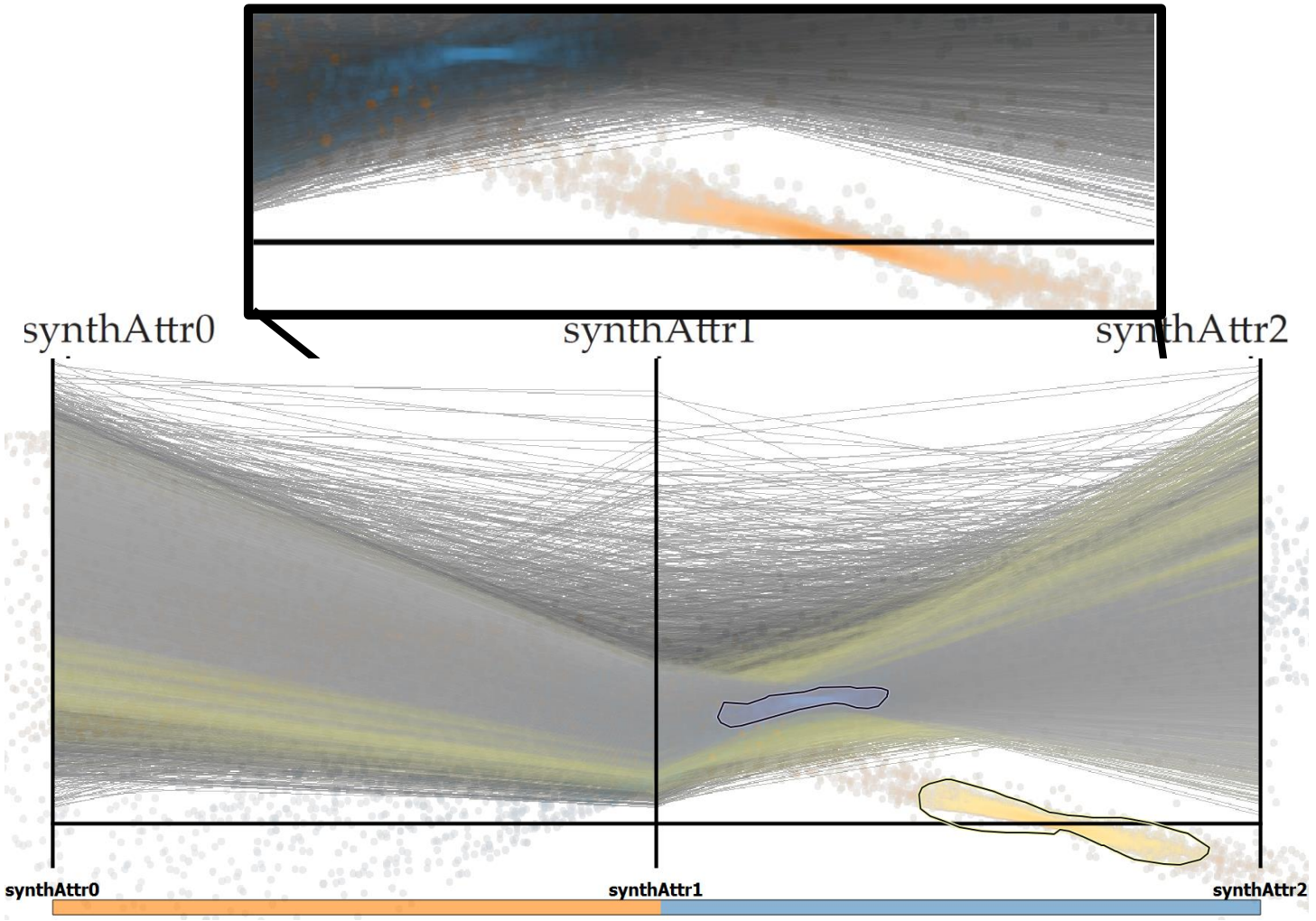
Negative Correlation

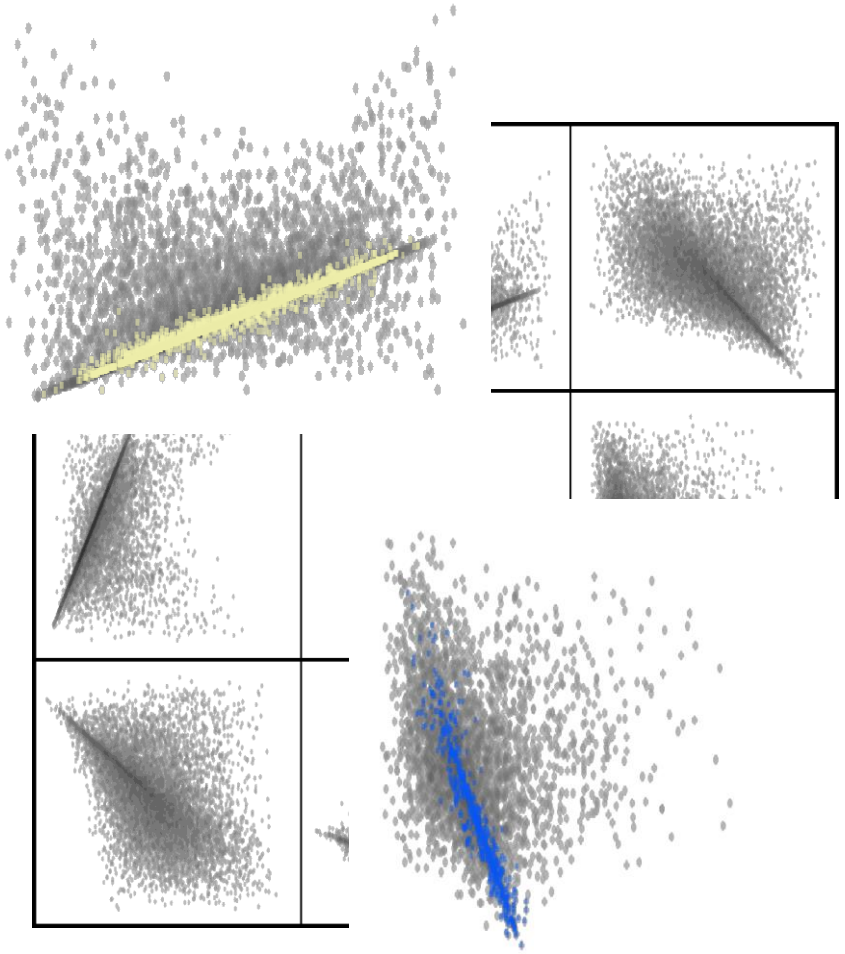


Positive Correlation

L. Zhou and D. Weiskopf, "Indexed-Points Parallel Coordinates Visualization of Multivariate Correlations," in *IEEE Transactions on Visualization and Computer Graphics*, doi: 10.1109/TVCG.2017.2698041.

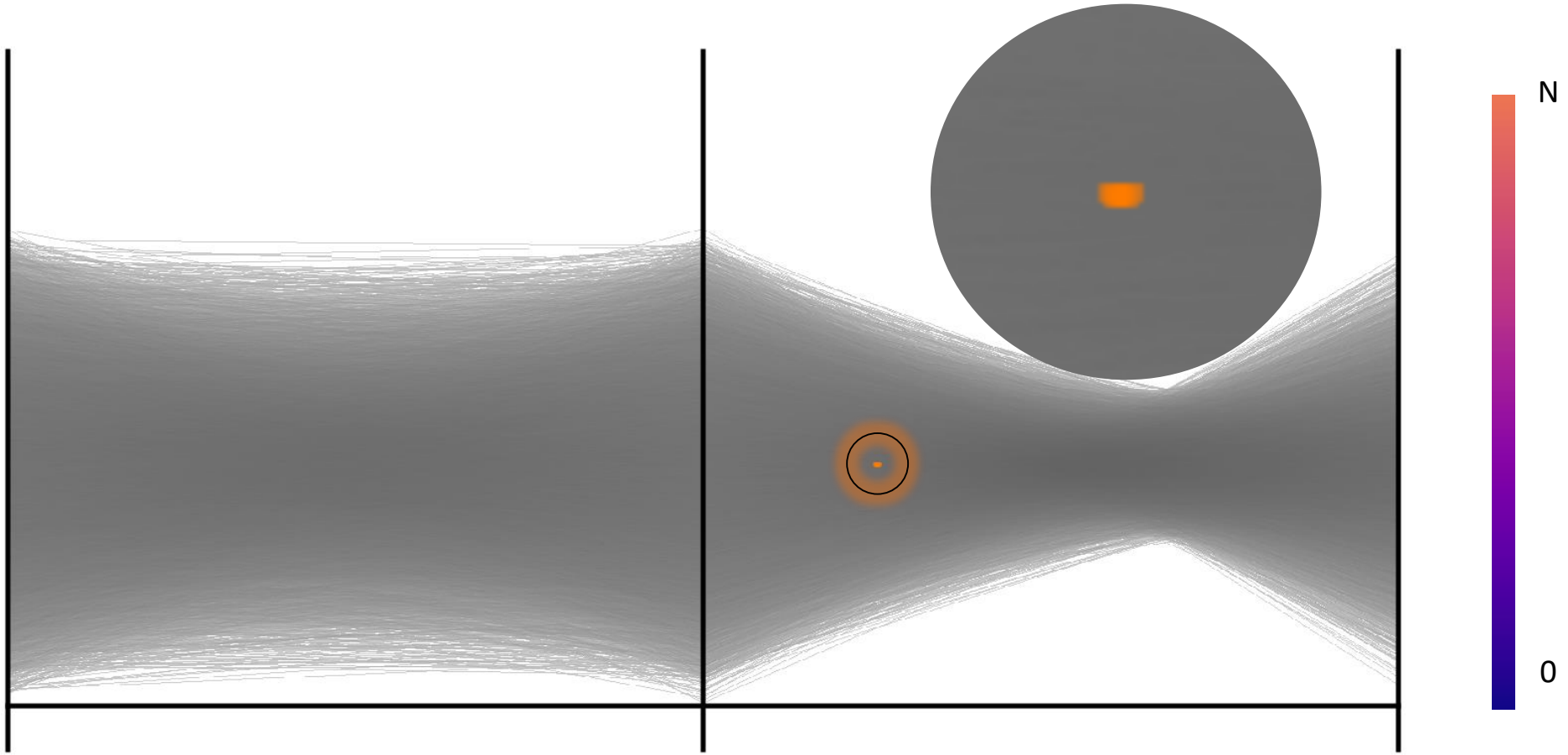
Patterns of Interest – 1-Flat Indexed Points





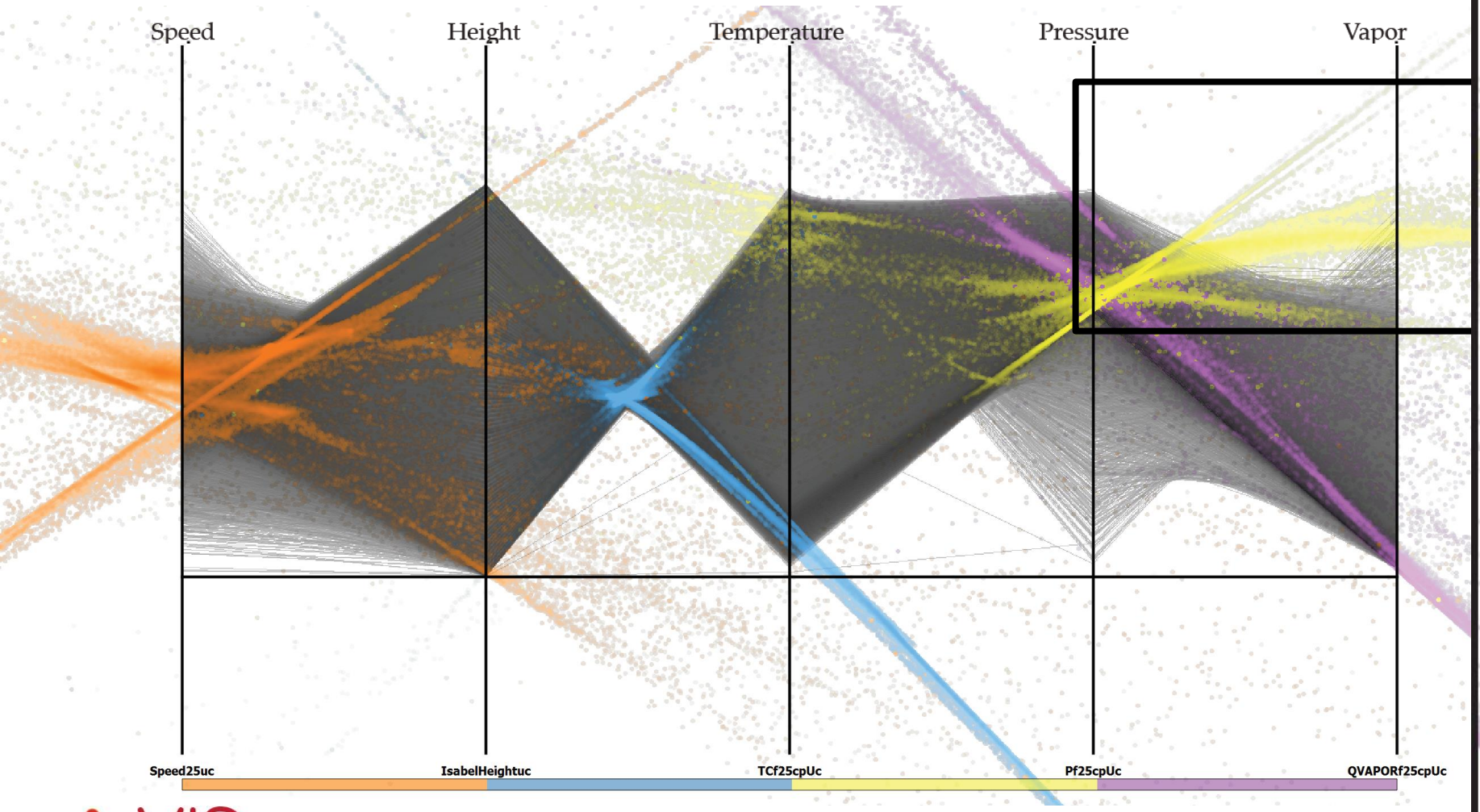
L. Zhou and D. Weiskopf, "Indexed-Points Parallel Coordinates Visualization of Multivariate Correlations," in *IEEE Transactions on Visualization and Computer Graphics*, doi: 10.1109/TVCG.2017.2698041.

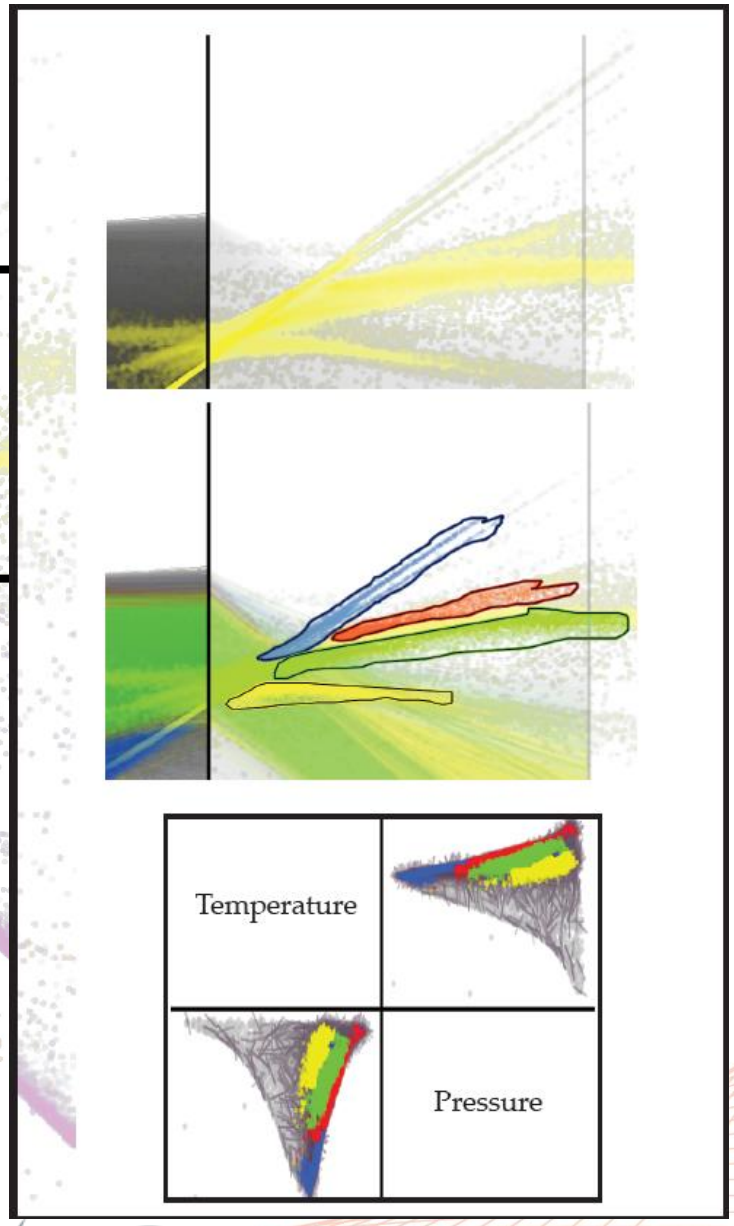
2-flat Indexed Points – A Single Plane



L. Zhou and D. Weiskopf, "Indexed-Points Parallel Coordinates Visualization of Multivariate Correlations," in *IEEE Transactions on Visualization and Computer Graphics*, doi: 10.1109/TVCG.2017.2698041.

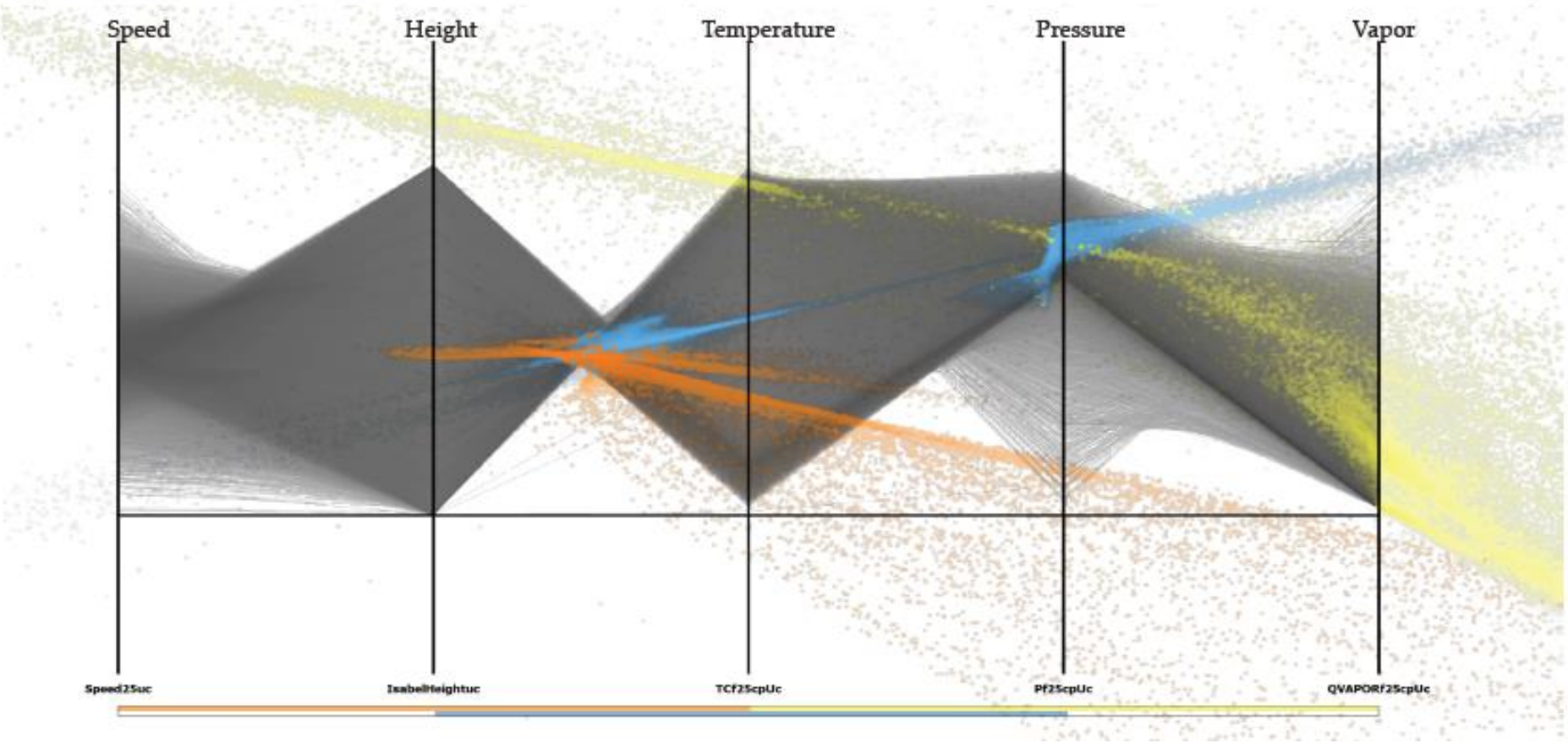
Examples – Hurricane Isabel 1-Flats





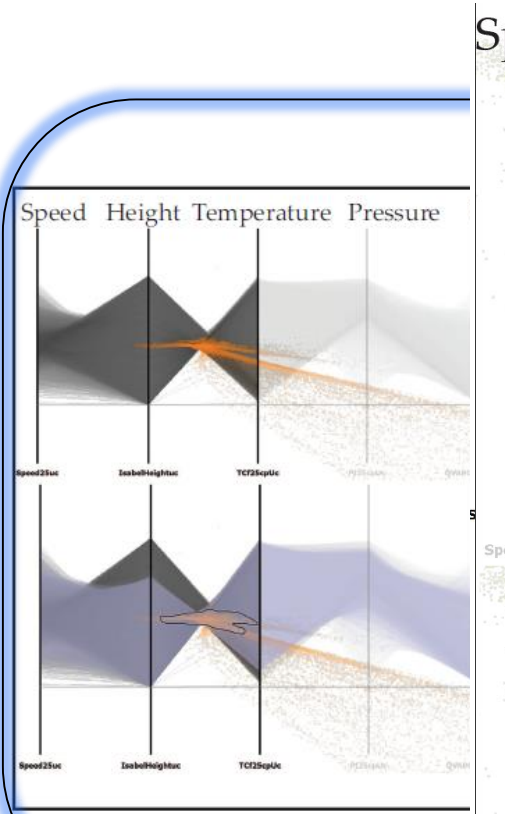
L. Zhou and D. Weiskopf, "Indexed-Points Parallel Coordinates Visualization of Multivariate Correlations," in *IEEE Transactions on Visualization and Computer Graphics*, doi: 10.1109/TVCG.2017.2698041.

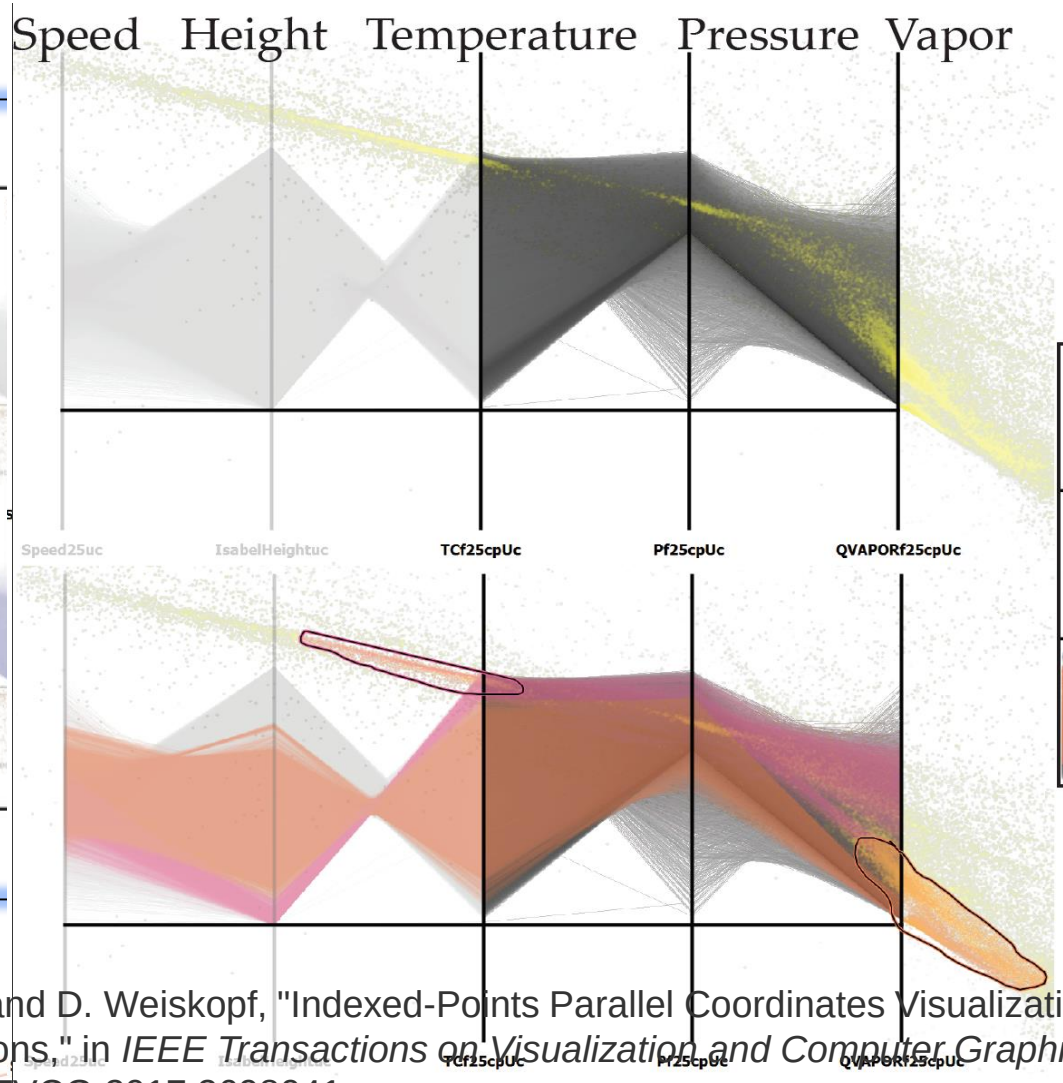
Examples – Hurricane Isabel 2-Flats

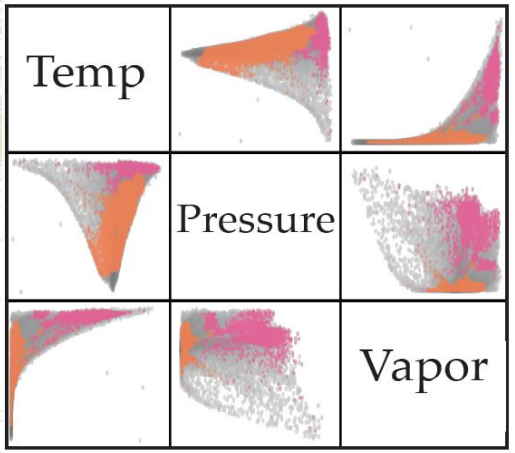


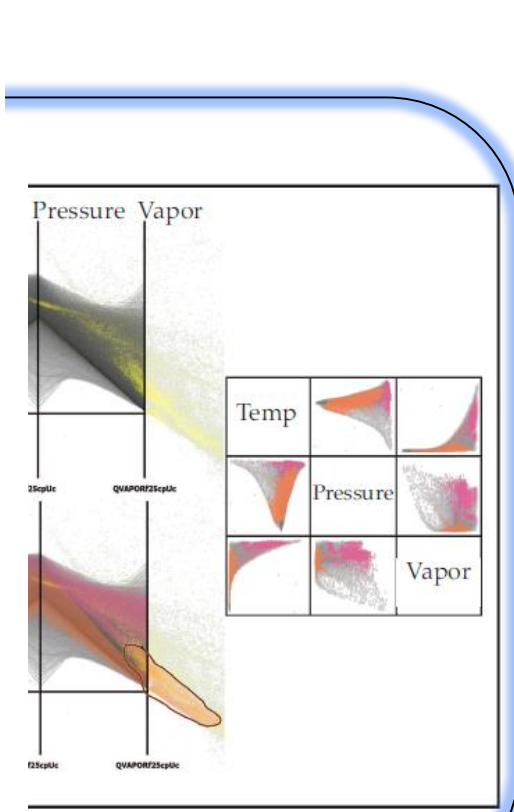
L. Zhou and D. Weiskopf, "Indexed-Points Parallel Coordinates Visualization of Multivariate Correlations," in *IEEE Transactions on Visualization and Computer Graphics*, doi: 10.1109/TVCG.2017.2698041.

Examples – Hurricane Isabel 2-Flats









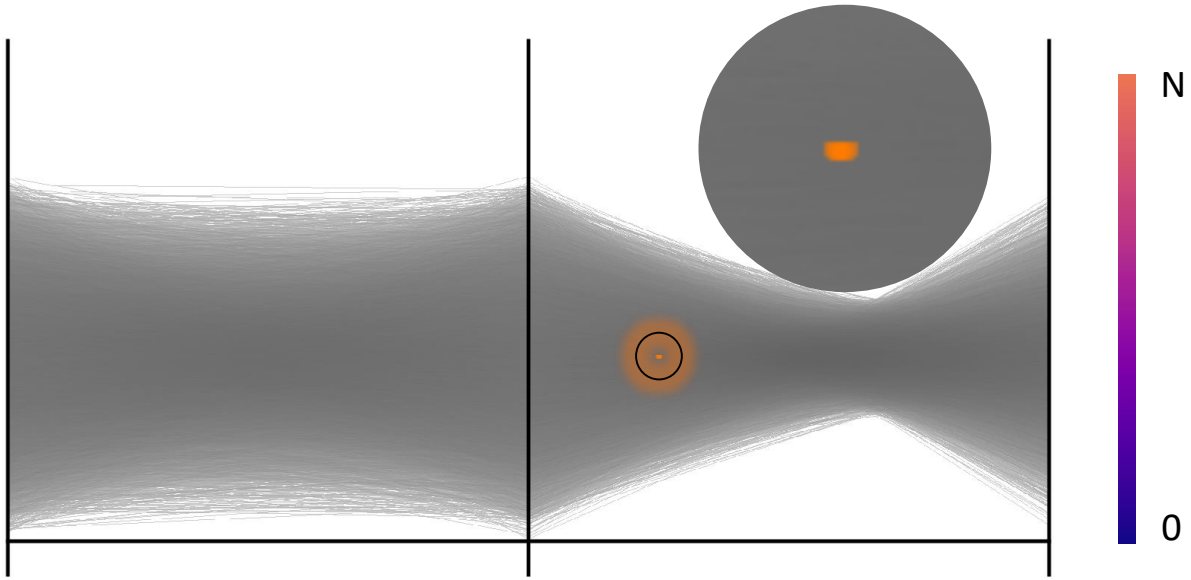
L. Zhou and D. Weiskopf, "Indexed-Points Parallel Coordinates Visualization of Multivariate Correlations," in *IEEE Transactions on Visualization and Computer Graphics*, doi: 10.1109/TVCG.2017.2698041.

新的问题：索引点显示

- 有时许多索引点会落在空间中同一位置（单一平面例子）→ 面积小，数值高的区域如何显示？
- 进一步，直方图、轨迹等数据中都会出现高动态范围的情况

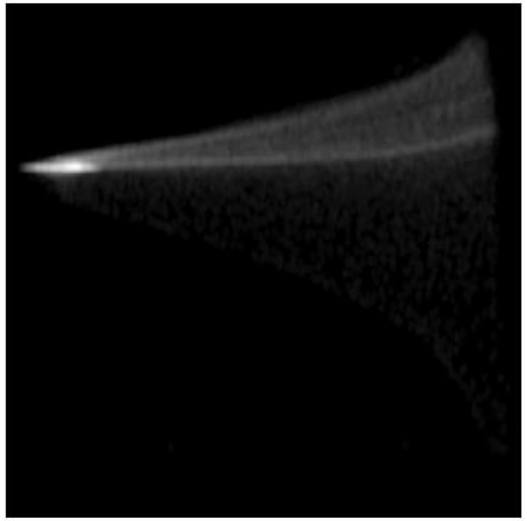
研究动机：

- 如何显示这些点同时显示整体分布？
- 方法符合视觉感知原理？

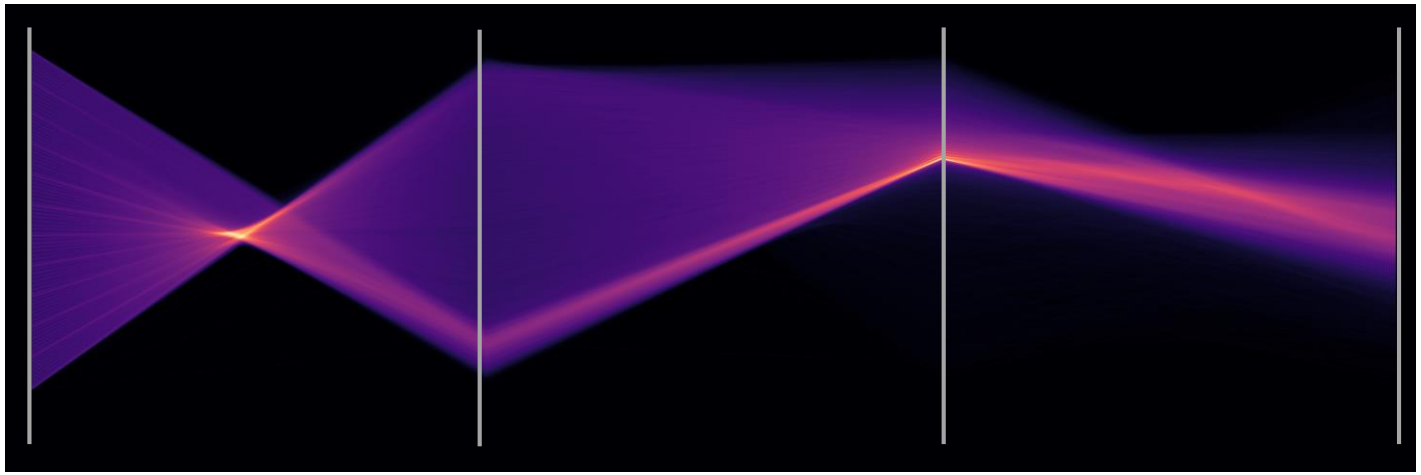


Visualizing High-Dynamic-Range (HDR) scalar data

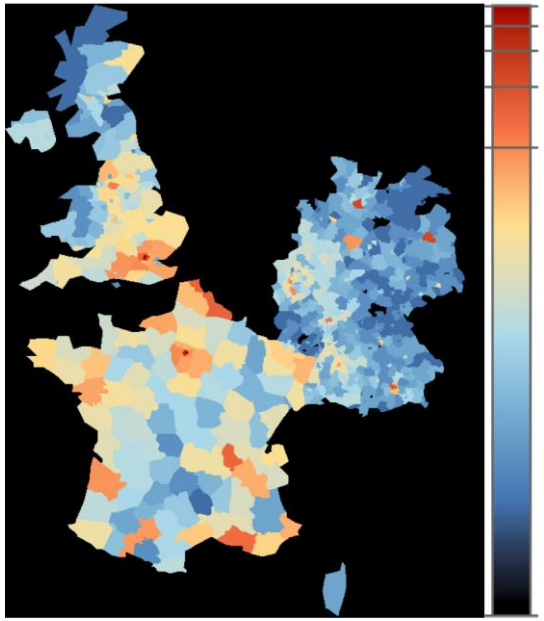
- HDR scalar data = scalar data values of very large (unbounded) ranges + 2D diagrams
- Scatterplots (matrices), parallel coordinates, trajectory plots, choropleth, node-link diagrams
- Visualizing with global transformation + color mapping



scatterplot



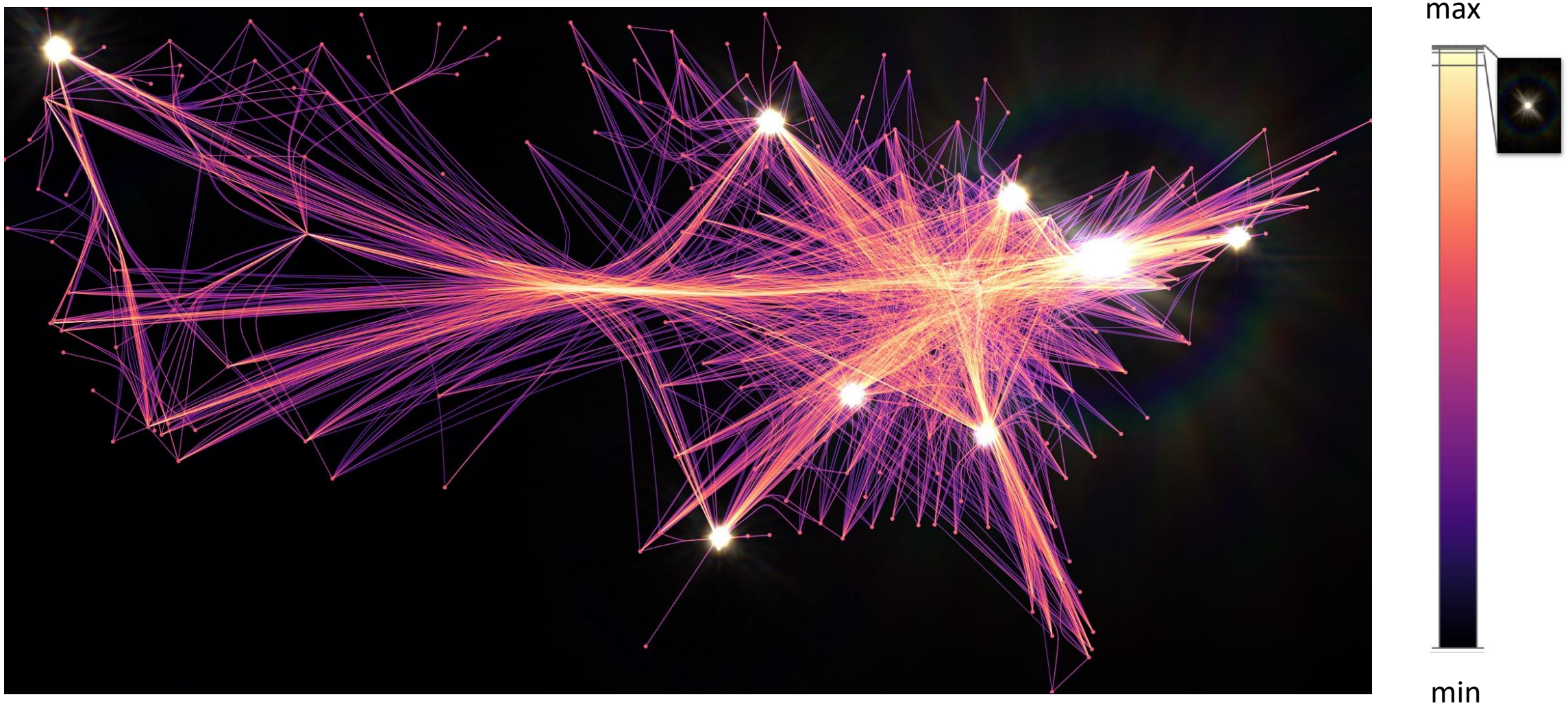
parallel coordinates



choropleth

L. Zhou, M. Rivinius, C. R. Johnson and D. Weiskopf, "Photographic High-Dynamic-Range Scalar Visualization," in IEEE Transactions on Visualization and Computer Graphics, doi: 10.1109/TVCG.2020.2970522.

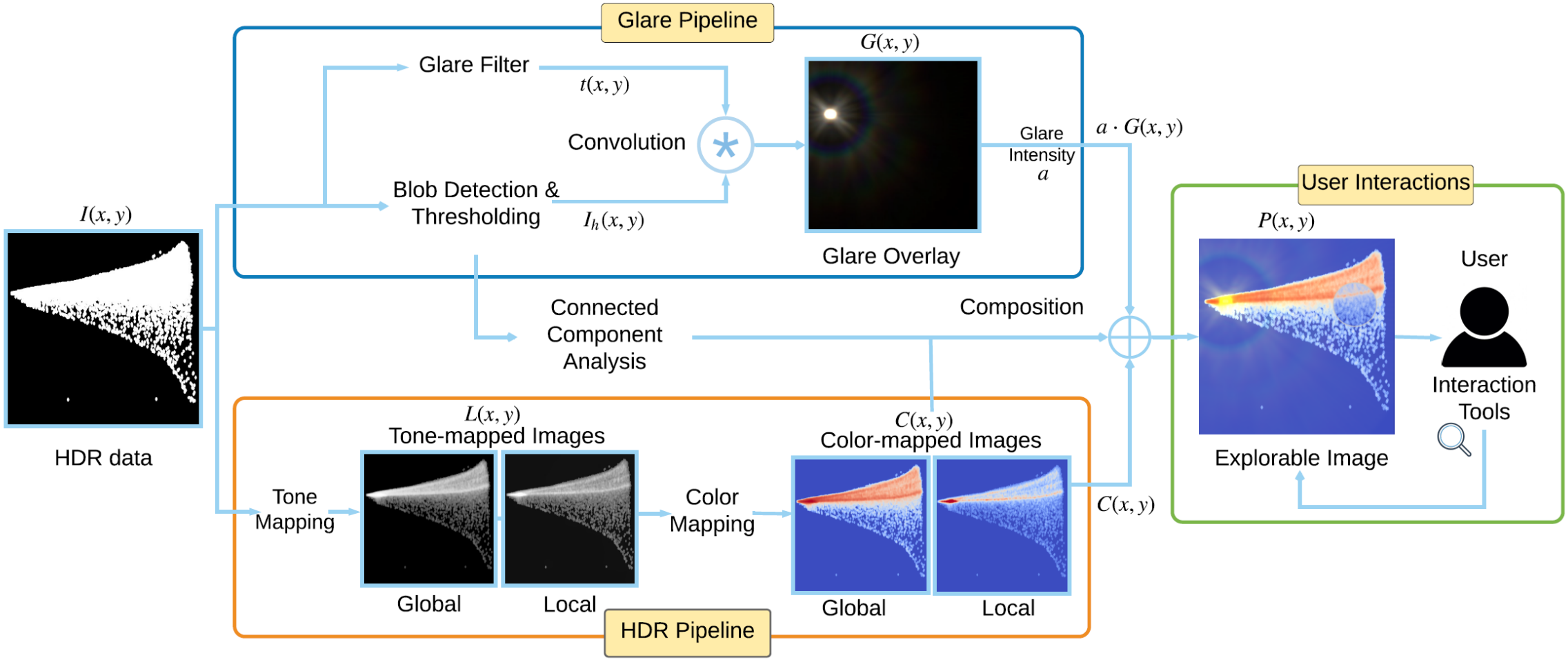
Visualizing High-Dynamic-Range (HDR) Scalar Data



Our new photographic HDR method

L. Zhou, M. Rivinius, C. R. Johnson and D. Weiskopf, "Photographic High-Dynamic-Range Scalar Visualization," in IEEE Transactions on Visualization and Computer Graphics, doi: 10.1109/TVCG.2020.2970522.

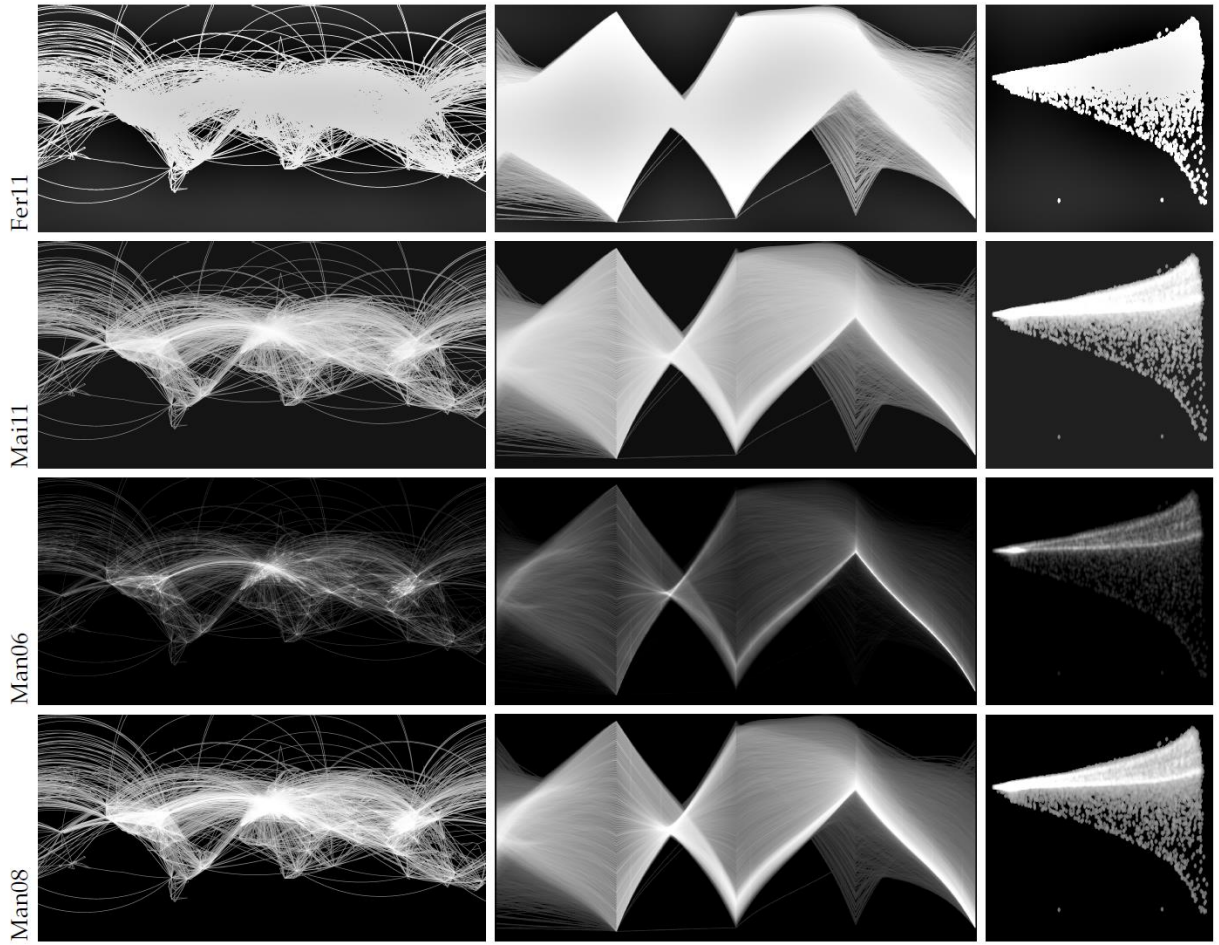
The workflow



L. Zhou, M. Rivinius, C. R. Johnson and D. Weiskopf, "Photographic High-Dynamic-Range Scalar Visualization," in IEEE Transactions on Visualization and Computer Graphics, doi: 10.1109/TVCG.2020.2970522.

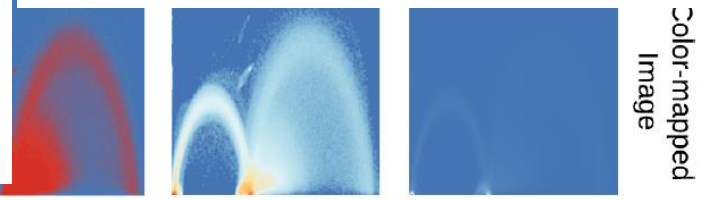
The HDR pipeline

- Tone-mapping operator → color mapping
- Systematic, quantitative evaluation ([Aydin et al. 08]) on TMOs for visualization
- Cu



Mean distortion metric of tone-mapping operators in our evaluation.

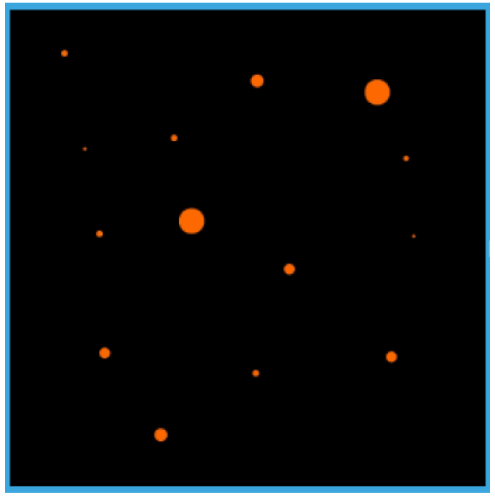
TMO	IsabelTvP	World Flights	IsabelPC	Averaged
Man08 [4]	0.483	0.372	0.698	0.517
Man06 [3]	0.541	0.796	0.558	0.632
Mai11 [5]	0.532	0.610	1.076	0.739
Ash02 [1]	0.749	1.653	0.731	1.044
Rei05 [11]	0.650	0.685	1.850	1.062
log	0.756	0.954	1.604	1.105
Dur02 [9]	0.803	1.211	1.632	1.216
Fat02 [7]	0.802	1.350	1.645	1.266
Fer11 [6]	1.267	1.666	1.681	1.538
Dra03 [8]	1.280	1.838	1.590	1.570
Pat00 [2]	1.339	1.684	1.696	1.573
Rei02 [10]	1.259	1.792	1.781	1.610
gamma	1.564	1.801	1.962	1.775
linear	2.580	2.663	2.626	2.623



The glare pipeline

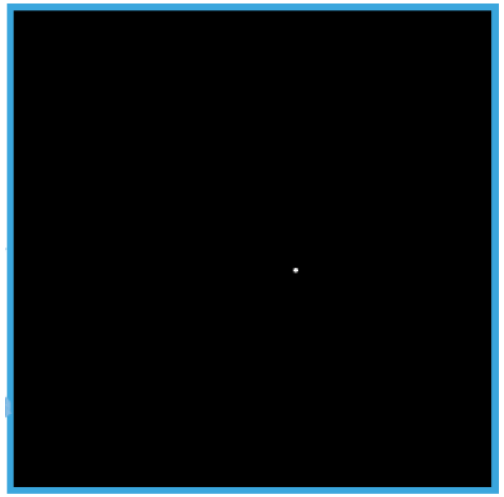
Input data

Blob detection

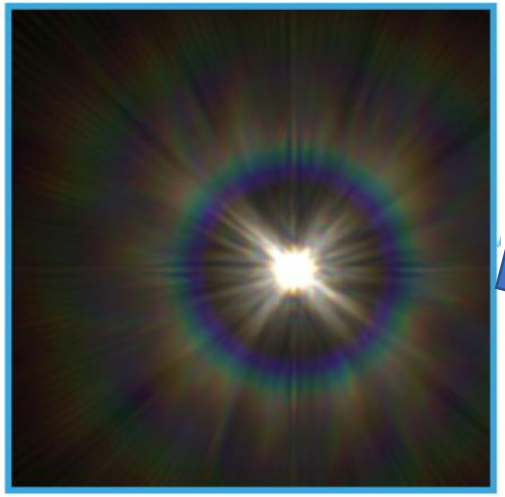


Blob mask

Thresholding

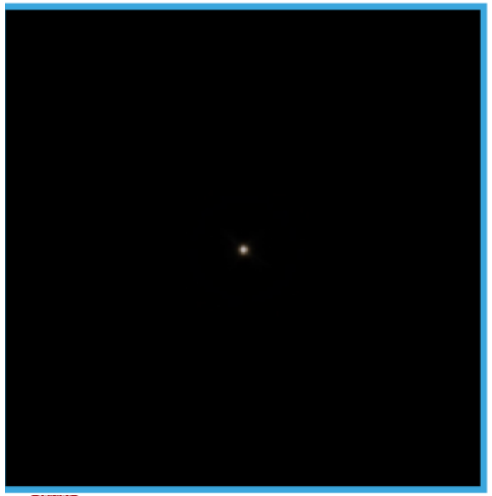


High-value pixels



Glare overlay

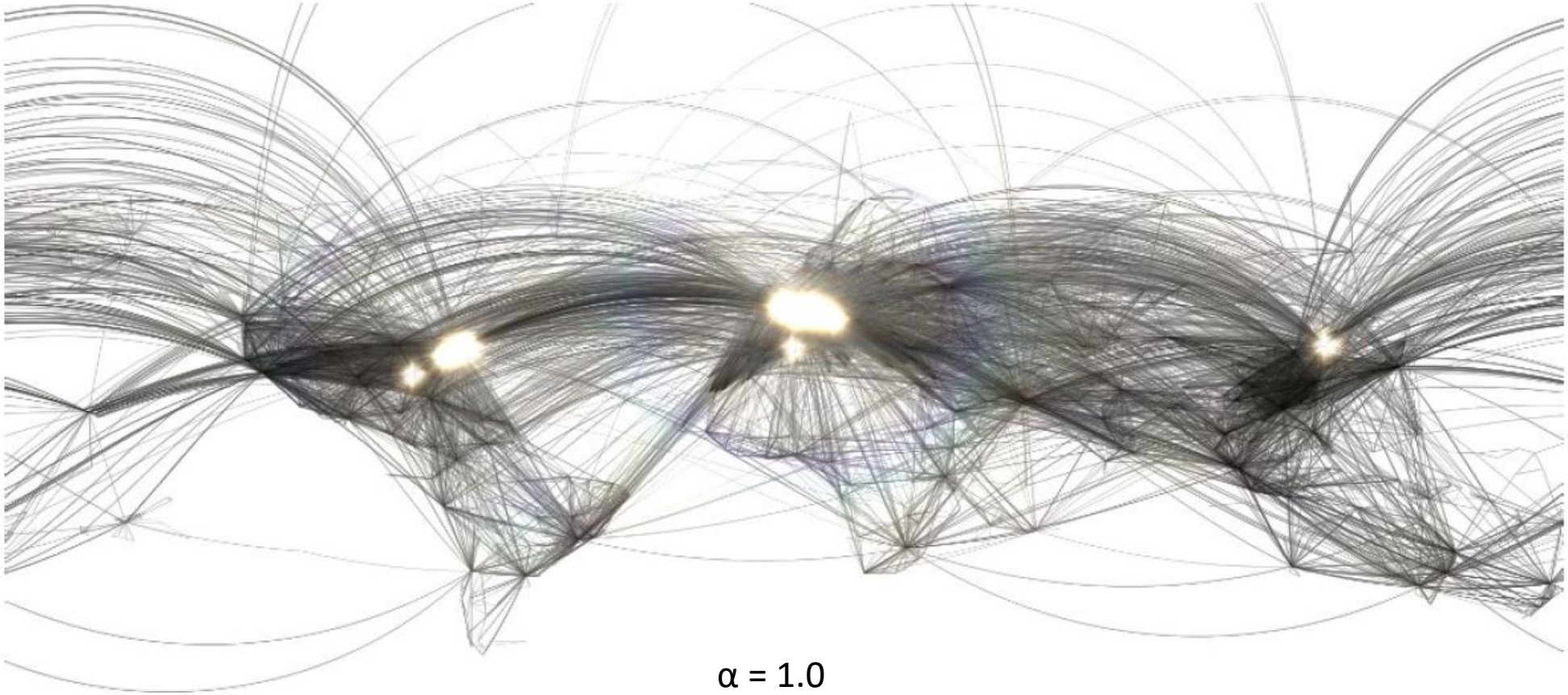
Convolution



Glare filter

Effect of glare intensity

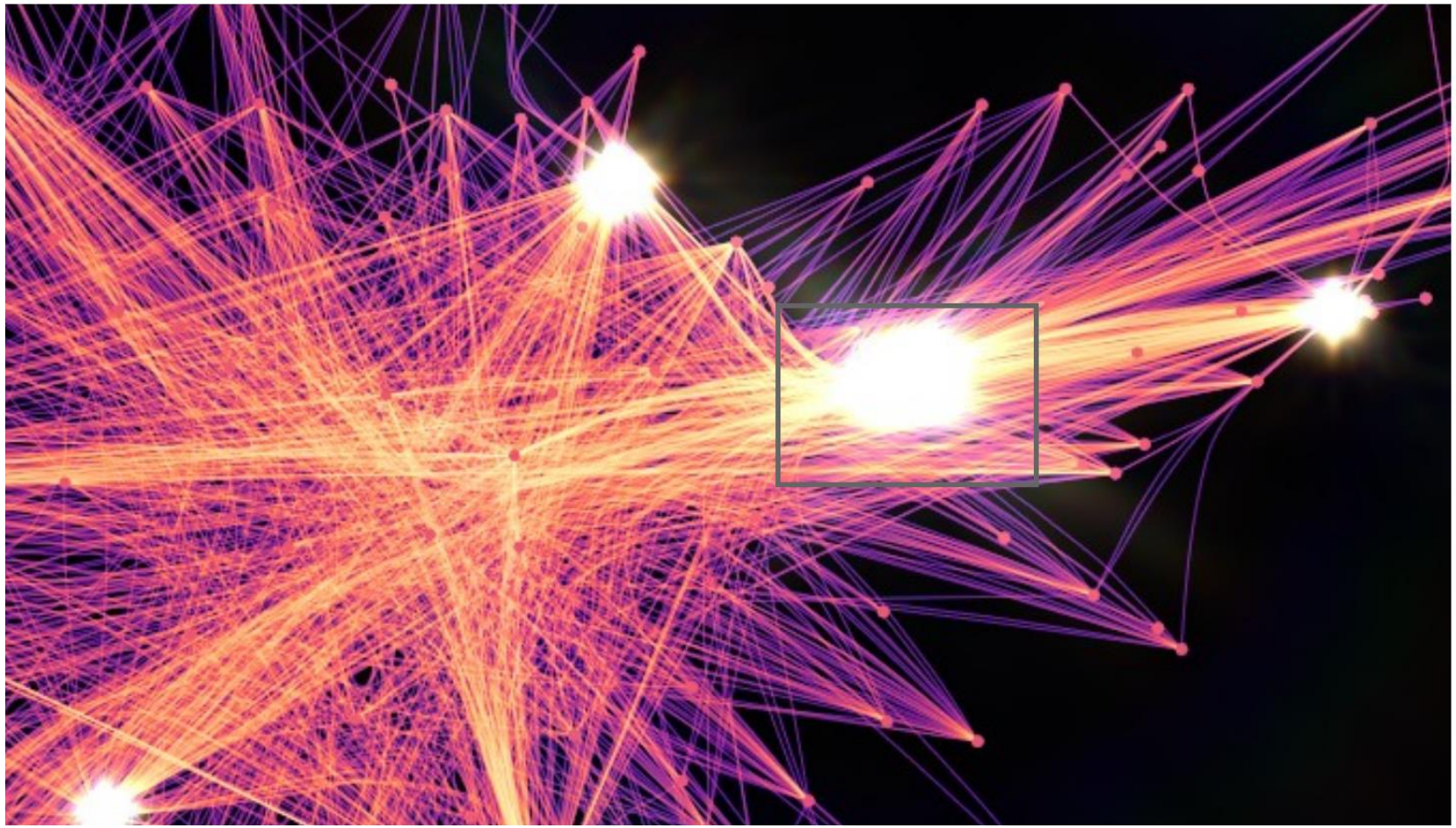
- Glare intensity α

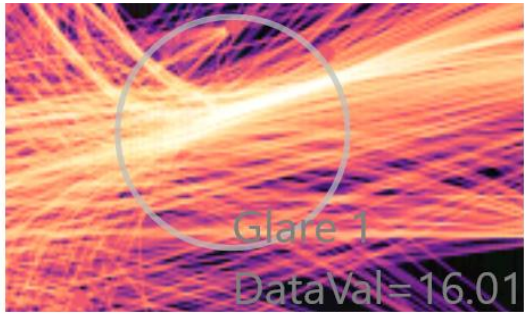


α = 1.0

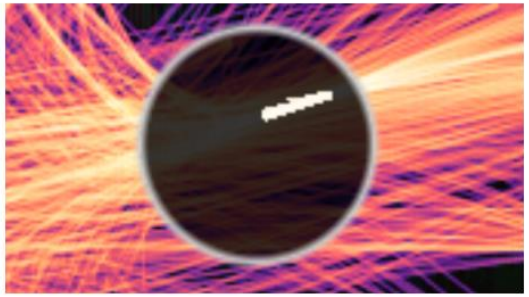
User interaction

- Magic lens and glare switch to “see through” glares
- Three modes of the magic lens

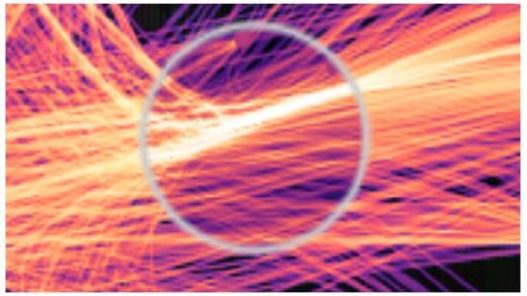




Reveal mode



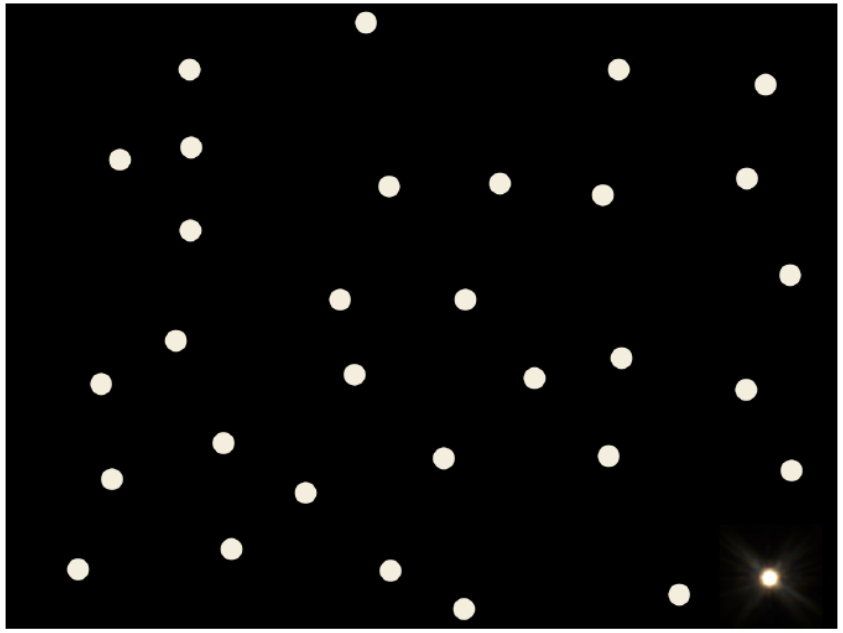
Bright pixel mode



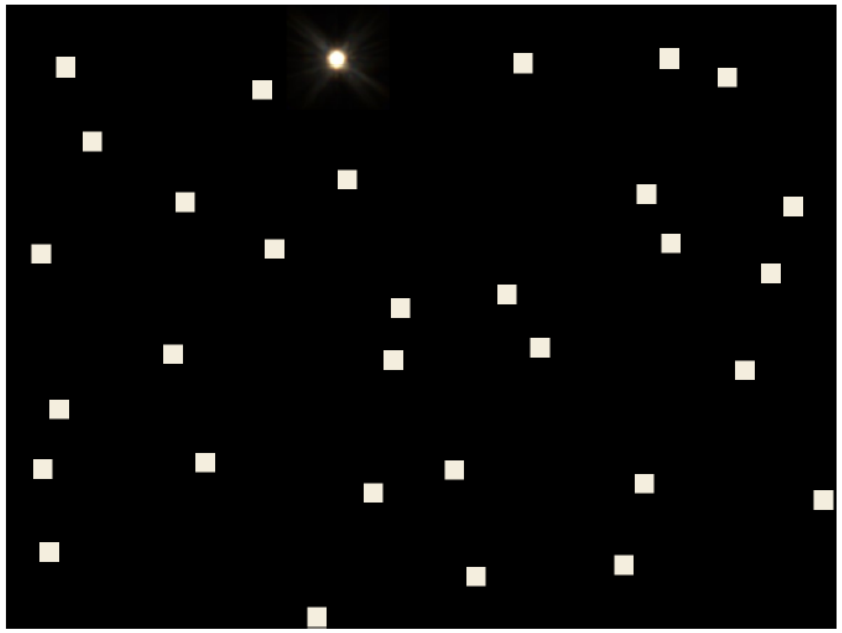
Contrast enhanced mode

为选用眩光提供证据——感知研究

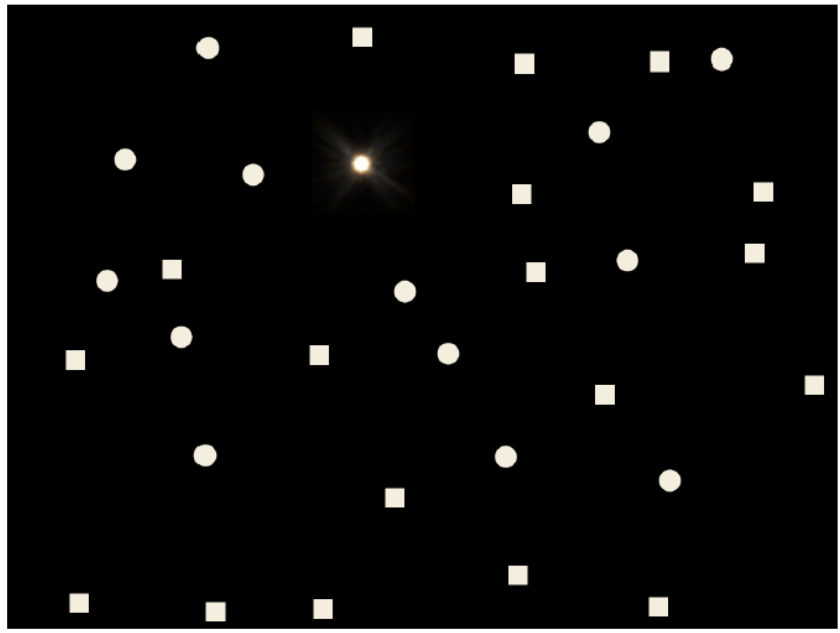
- Hypothesis: **glare is a preattentive visual cue**
- Typical state-of-the-art preattentive study design ([Krekhov and Krüger 19; Krekhov et al. 20])
  - Main factor: number of objects; second factor: type of distractors
  - Each condition has 48 tests → 48 * 2 * 3 = 288 tests per participant
  - The horizontal size of the test image is about 15°
- Task: each stimuli is shown for 250ms → participants indicate whether a glare was present



Disk



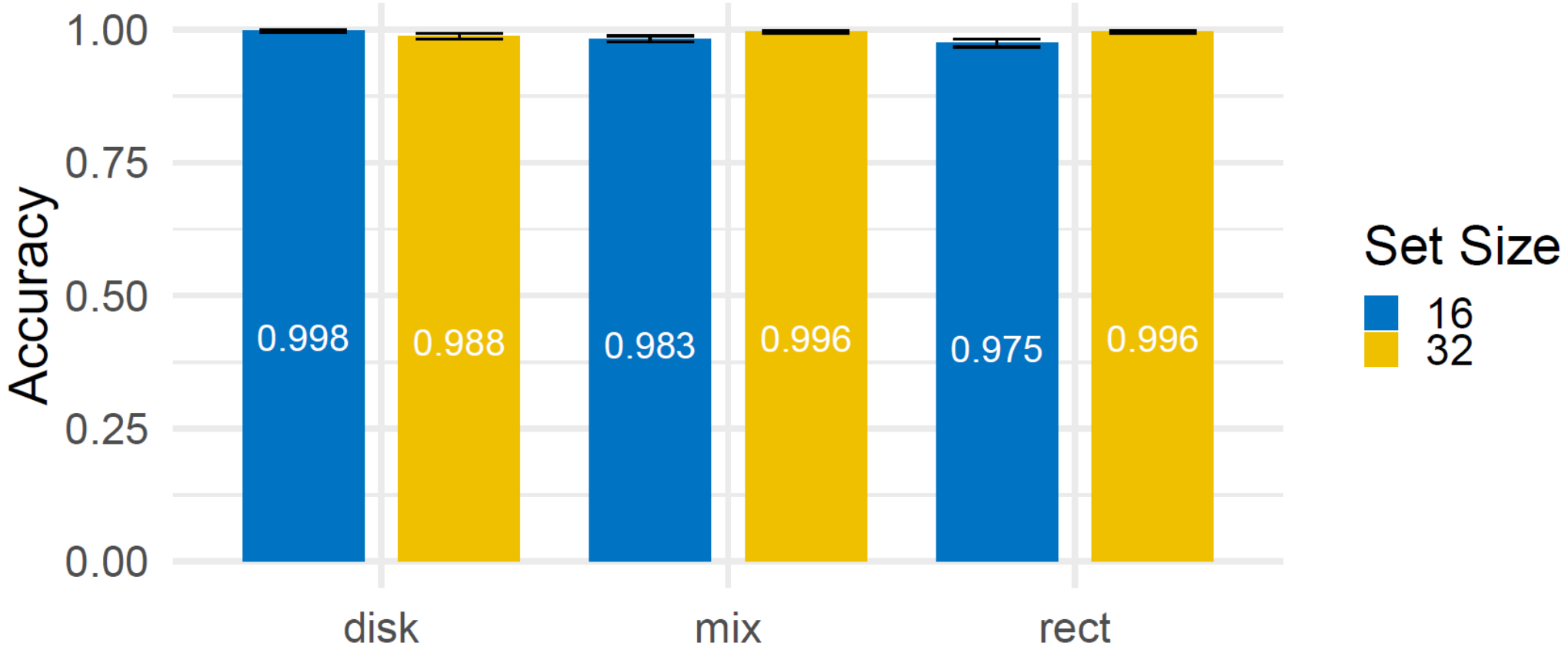
Rectangle



Mixed

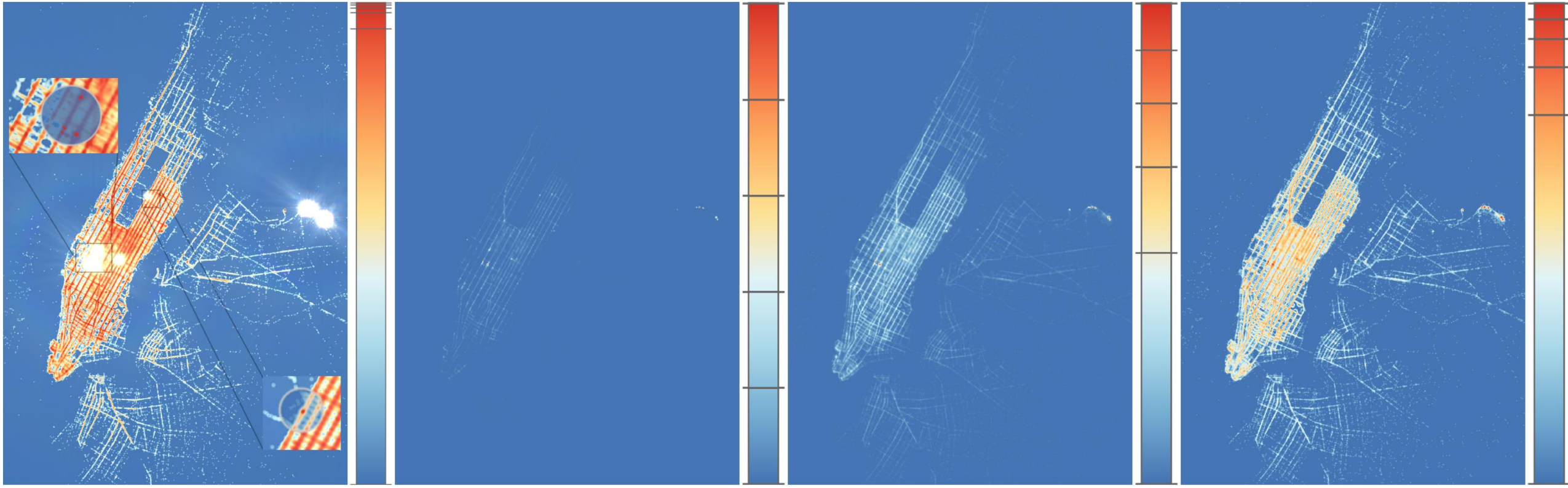
Perception study results

- Ten subjects (3 females, 7 males)
- High accuracies across all conditions→ similar to other preattentive visual cues
- Two-way ANOVA: no significant difference in accuracy between set size ($F(1; 9) = 1.678$; $p = 0.227$) nor type of distractors ($F(2; 18) = 1.316$; $p = 0.293$)
- **Glare is confirmed to be a preattentive visual cue**



Examples: dot-based geospatial data

- NYC taxi pick-up data



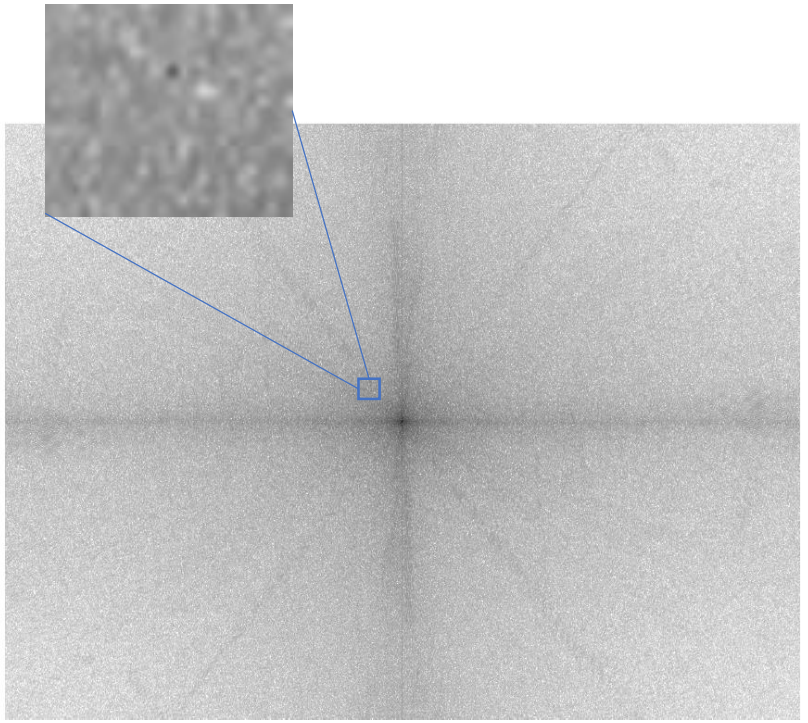
ours | linear mapping | gamma mapping | logarithmic mapping

L. Zhou, M. Rivinius, C. R. Johnson and D. Weiskopf, "Photographic High-Dynamic-Range Scalar Visualization," in IEEE Transactions on Visualization and Computer Graphics, doi: 10.1109/TVCG.2020.2970522.

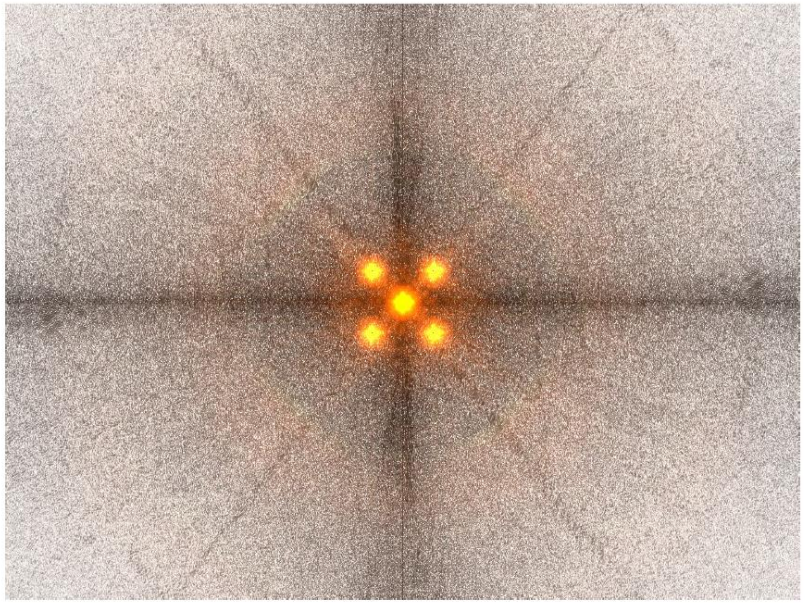
Examples: Image processing

- Power spectrum of a noisy image

Image corrupted by sine functions



logarithmic mapping

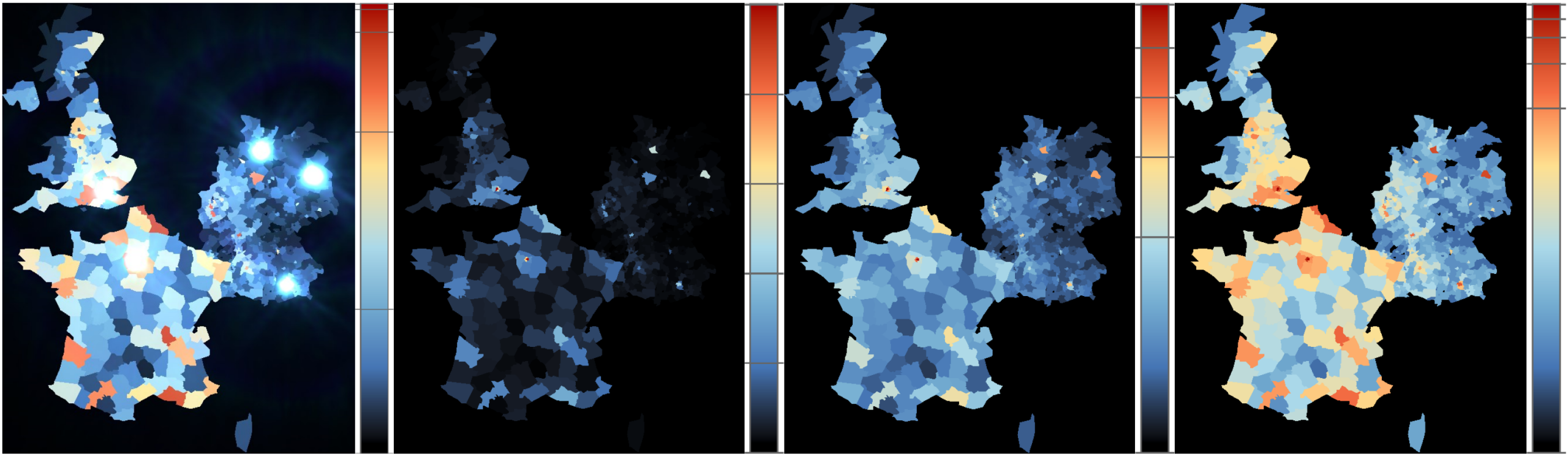


ours

L. Zhou, M. Rivinius, C. R. Johnson and D. Weiskopf, "Photographic High-Dynamic-Range Scalar Visualization," in IEEE Transactions on Visualization and Computer Graphics, doi: 10.1109/TVCG.2020.2970522.

Examples: choropleth

- GDP per-person of Germany, France, and Great Britain of year 2008

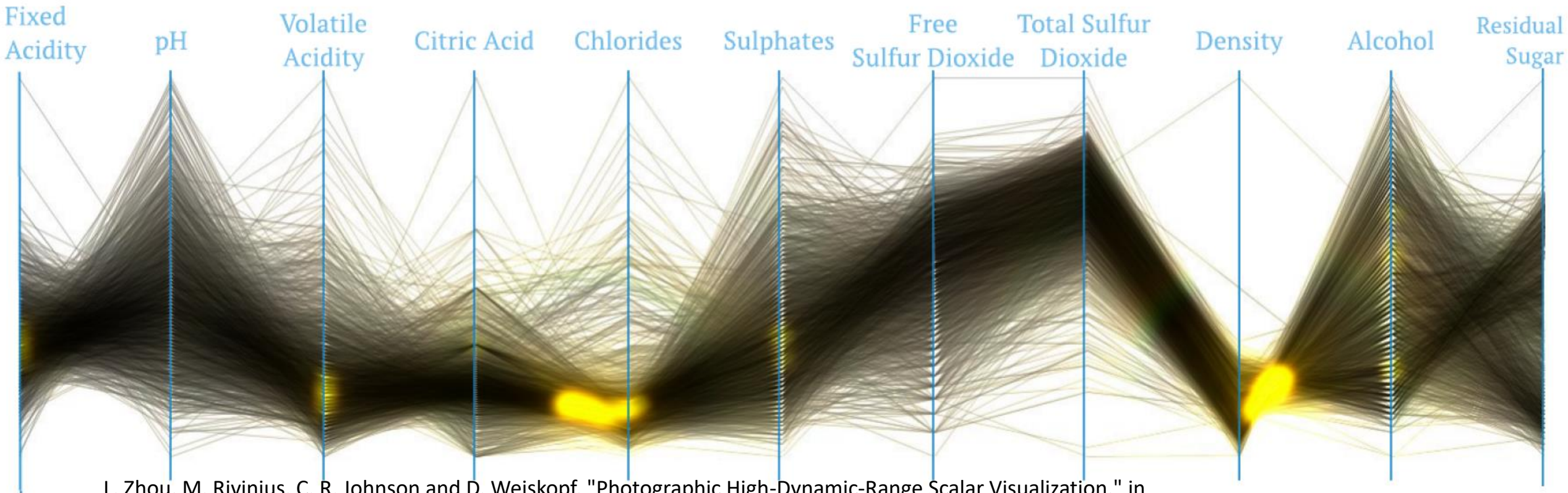


ours | linear mapping | gamma mapping | logarithmic mapping

L. Zhou, M. Rivinius, C. R. Johnson and D. Weiskopf, "Photographic High-Dynamic-Range Scalar Visualization," in IEEE Transactions on Visualization and Computer Graphics, doi: 10.1109/TVCG.2020.2970522.

Examples: parallel coordinates

- Wine quality dataset



L. Zhou, M. Rivinius, C. R. Johnson and D. Weiskopf, "Photographic High-Dynamic-Range Scalar Visualization," in IEEE Transactions on Visualization and Computer Graphics, doi: 10.1109/TVCG.2020.2970522.

研究经验总结

- 研究可视化基础问题
- 发现现有方法中的未解决的部分 （滚雪球？）
- 了解相关问题在其他领域的工作
- 通过（正确设计的）用户实验提供证据能很好提高文章的录用概率
- 导师指导非常重要!