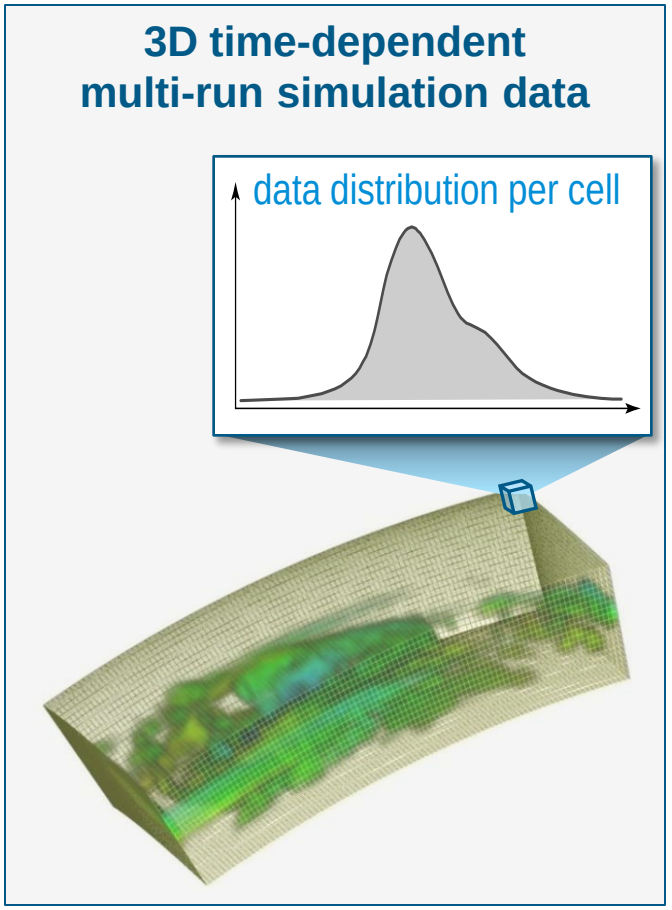
Brushing Moments in Interactive Visual Analysis

Johannes Kehrer[1], Peter Filzmoser [2], and Helwig Hauser[1]

[1] Department of Informatics, University of Bergen, Norway
[2] Department of Statistics and Probability Theory,
Vienna University of Technology, Austria

Higher-dimensional Scientific Data

- Considering “scientific” data **f**, i.e.,
  - some **data values f(p)** (e.g., temperature, pressure values)
  - measured/simulated wrt. a **domain p** (e.g., 2D/3D space, time, simulation input parameters)



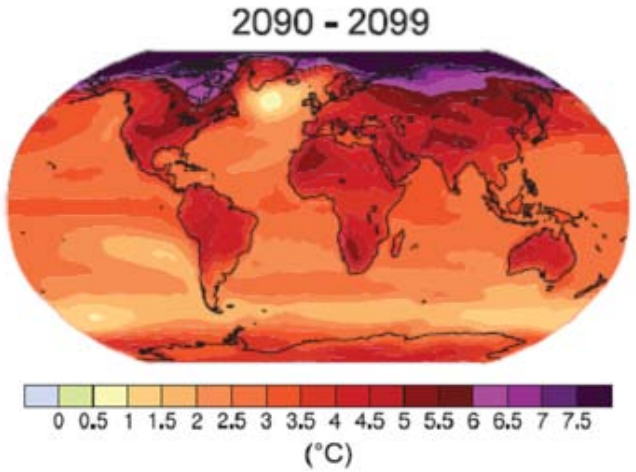
- If dimensionality of **p** > 3, then traditional visual analysis is hard
- Reducing the data dimensionality can help (e.g., computing statistical aggregates)

Reducing the Data Dimensionality

- **Statistics:** assess distributional characteristics along an independent data dimension (e.g., time, spatial axes)

- Integrate aggregated statistics into visual analysis through **attribute derivation**

[from IPCC AR #4, 2007]



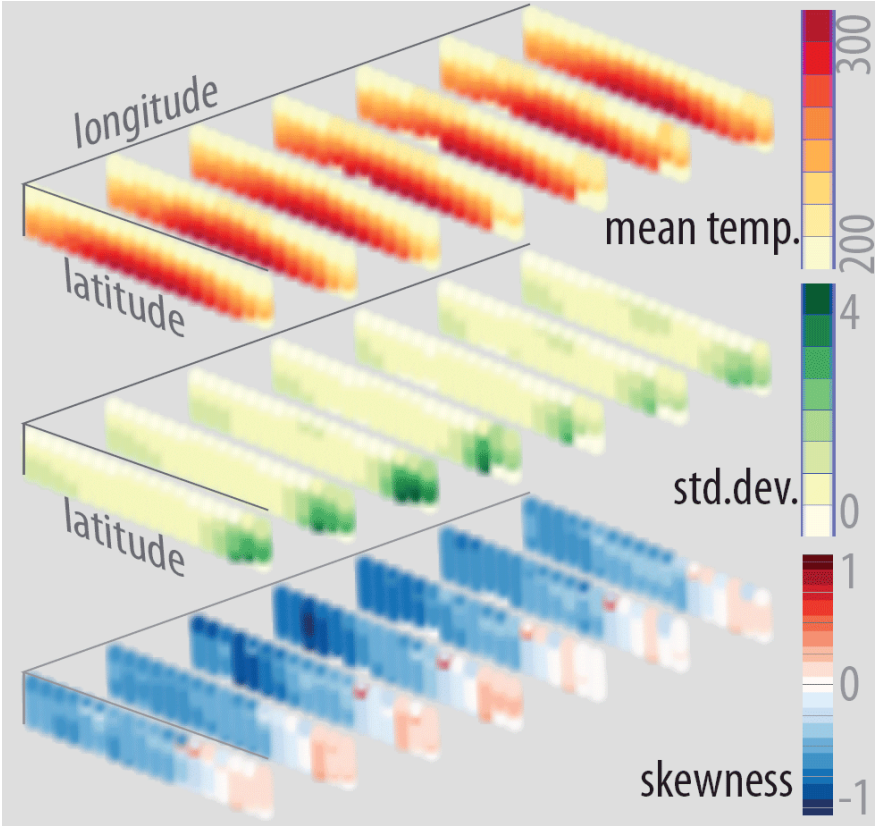
average temp. in ten years

Example: Multi-run Climate Simulation Data

CLIMBER-2 model: Meltwater outburst of Lake Agassiz

- 3D atmosphere
- 250 time steps
- 240 runs (7 model parameters)

→ Compute local statistics wrt. multiple runs

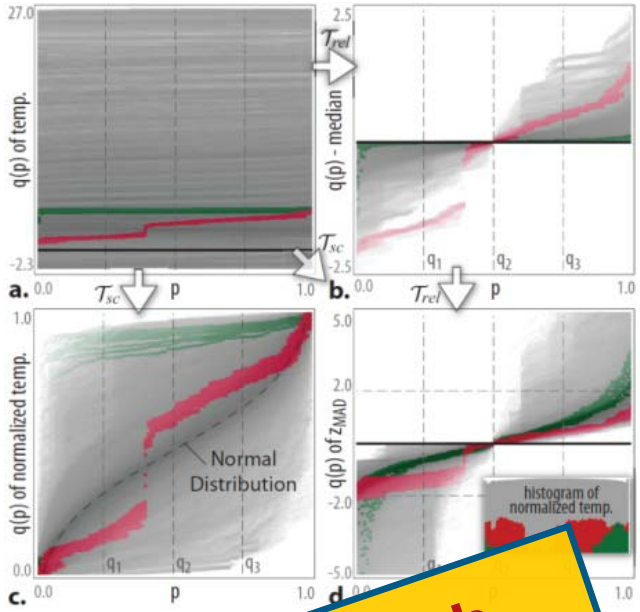


timestep 80

Moment-based Visual Analysis

- Get big picture (data trends & outliers)
- Multitude of choices, e.g,
  - 4 statistical moments (mean, std. deviation, skewness, kurtosis)
  - ×3 traditional and 2 robust estimates
  - ×2 compute relation (e.g., differences, ratio)
  - ×3 change scale (e.g., data normalization, log. scaling, measure of “outlyingness”)

= 72 possible configurations per axis

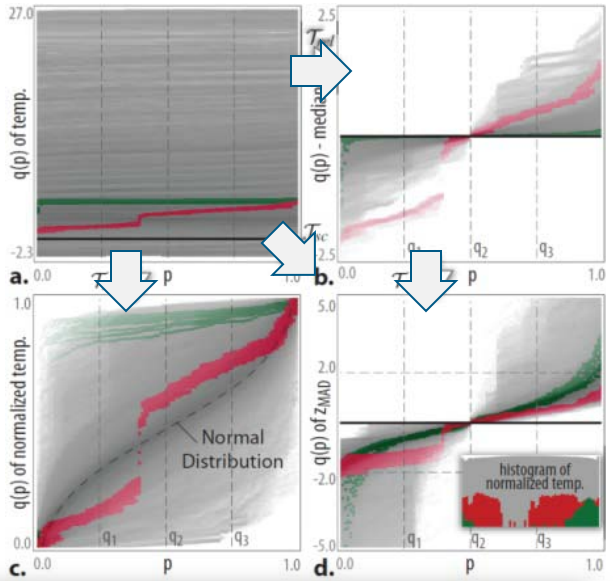


Structured approach to manage complexity

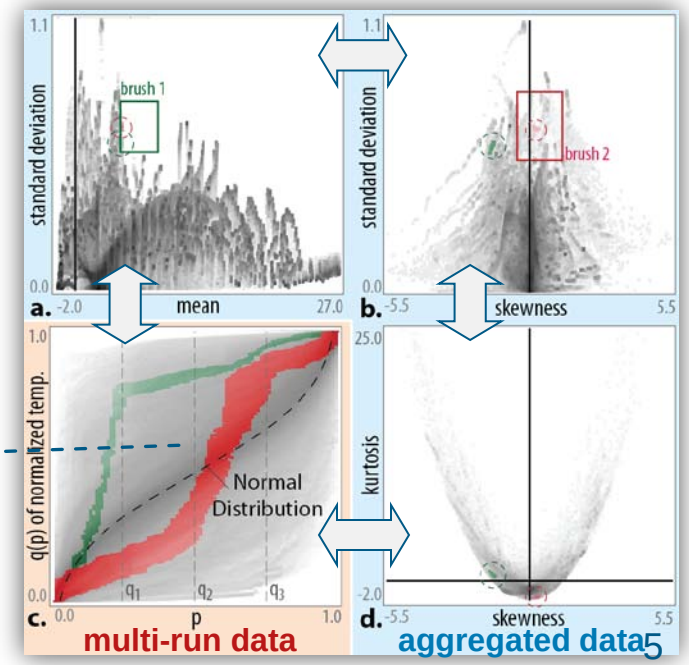
- How to deal with this “management challenge”?

Moment-based Visual Analysis

- Iterative view transformations
  - alter axis/attribute configuration
    (construct a multitude of informative views)
  - maintain mental model of views
  - classification of moment-based views



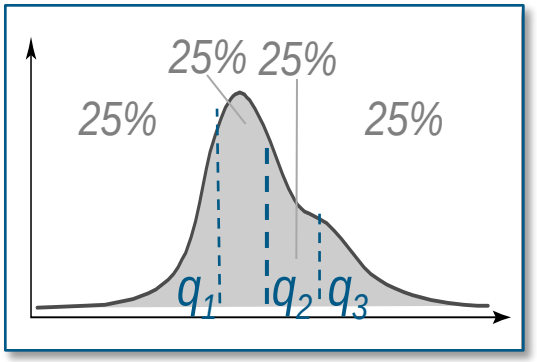
- Relate
  multi-run data ⇔ aggregated data



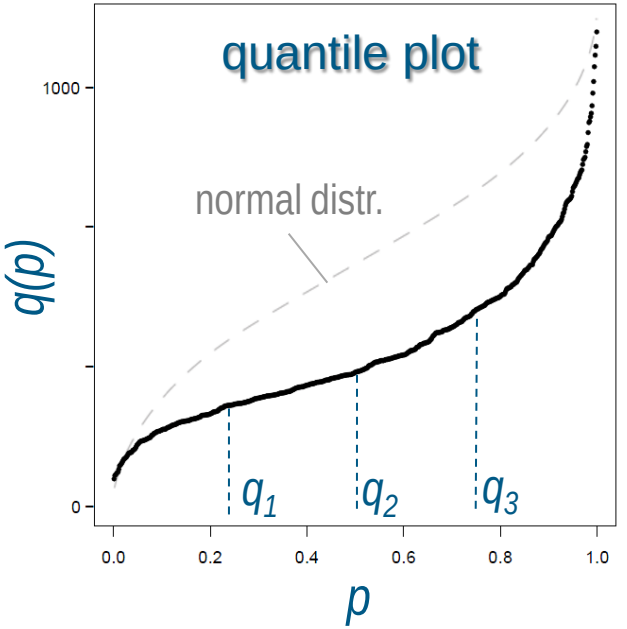
quantile plot
(focus+context)

Visualizing Data Distributions

- Sample quantile $q(p)$ of a distribution $\{x_1,\ldots,x_n\}$ ($p \in [0, 1]$)
  - at least $n \cdot p$ observations $\leq q(p)$
  - at least $n \cdot (1-p)$ observations $\geq q(p)$
- Examlpes: median $q_2 = q(½)$, quartiles q_1, q_3

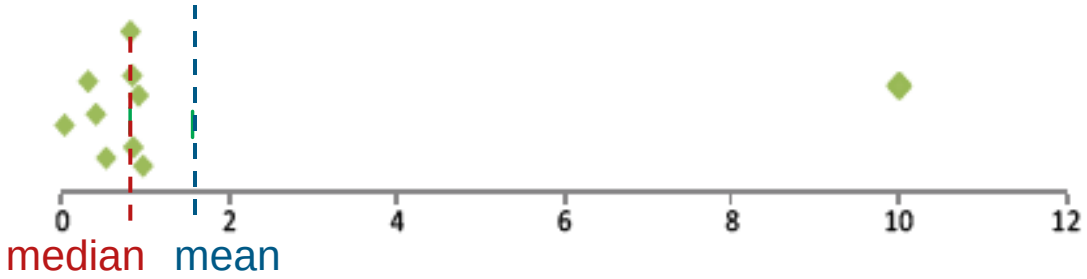


- Quantile plot
  - shows all data items of a distribution
  - assess data characteristics
    (normal dist., symmetrical, skewness, possible outliers, etc.)

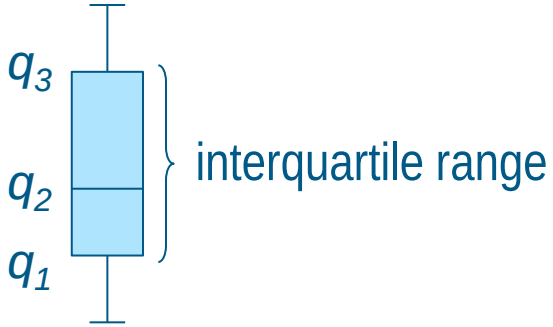


Robust Statistics

Outlier influence traditional estimates



- Robust estimates of std. deviation
  - 0.741· interquartile range (IQR)



  - median absolute deviation
    $MAD(x_1,\ldots,x_n) = 1.483 \cdot \text{med}_{1 \le i \le n}(|\, x_i - \text{median}\,|)$

Robust Statistics

- Estimates of skewness
  - traditional

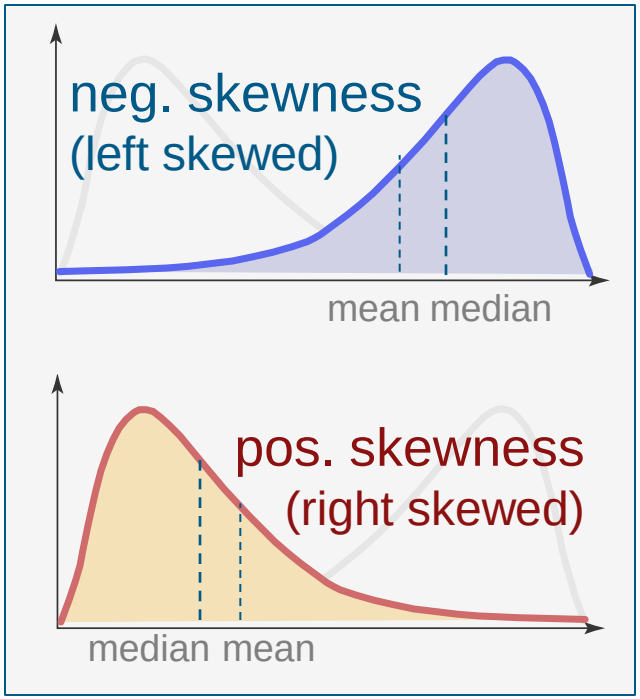
$$skewness = \frac{1}{n}\sum_{i=1}^{n}\frac{(x_i - \bar{x})^3}{s^3}$$

  - median/MAD-based

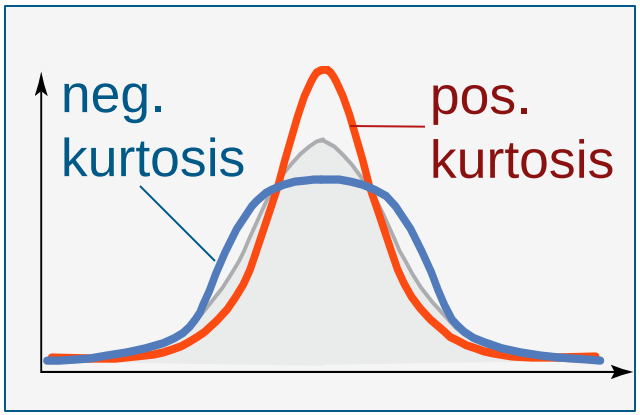
$$skew_{MAD} = \frac{1}{n}\sum_{i=1}^{n}\frac{(x_i - median)^3}{MAD(x_1,...,x_n)^3}$$

  - octile-based

$$skew_{oct} = \frac{e_7 + e_1 - 2e_4}{e_7 - e_1}$$



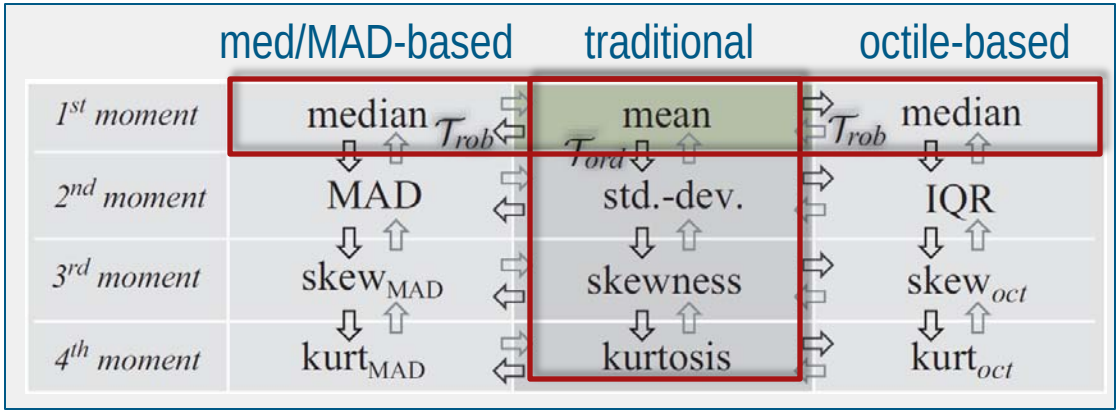
- Analogous estimates for kurtosis



Iterative View Transformations

Change axis/attribute configuration of view

- **change order** of moment
- **robustify moment**

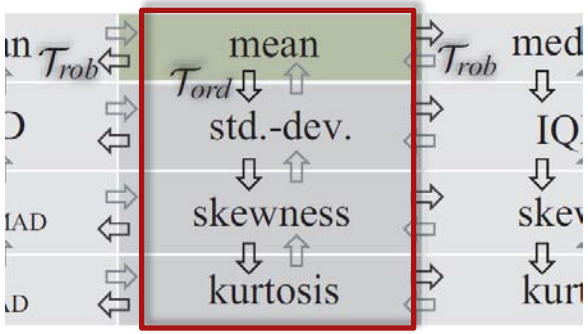


- **compute relation** (e.g., difference or ratio)
- **change scale** (e.g., normalize, z-standardization)

Closer related to data tranformations

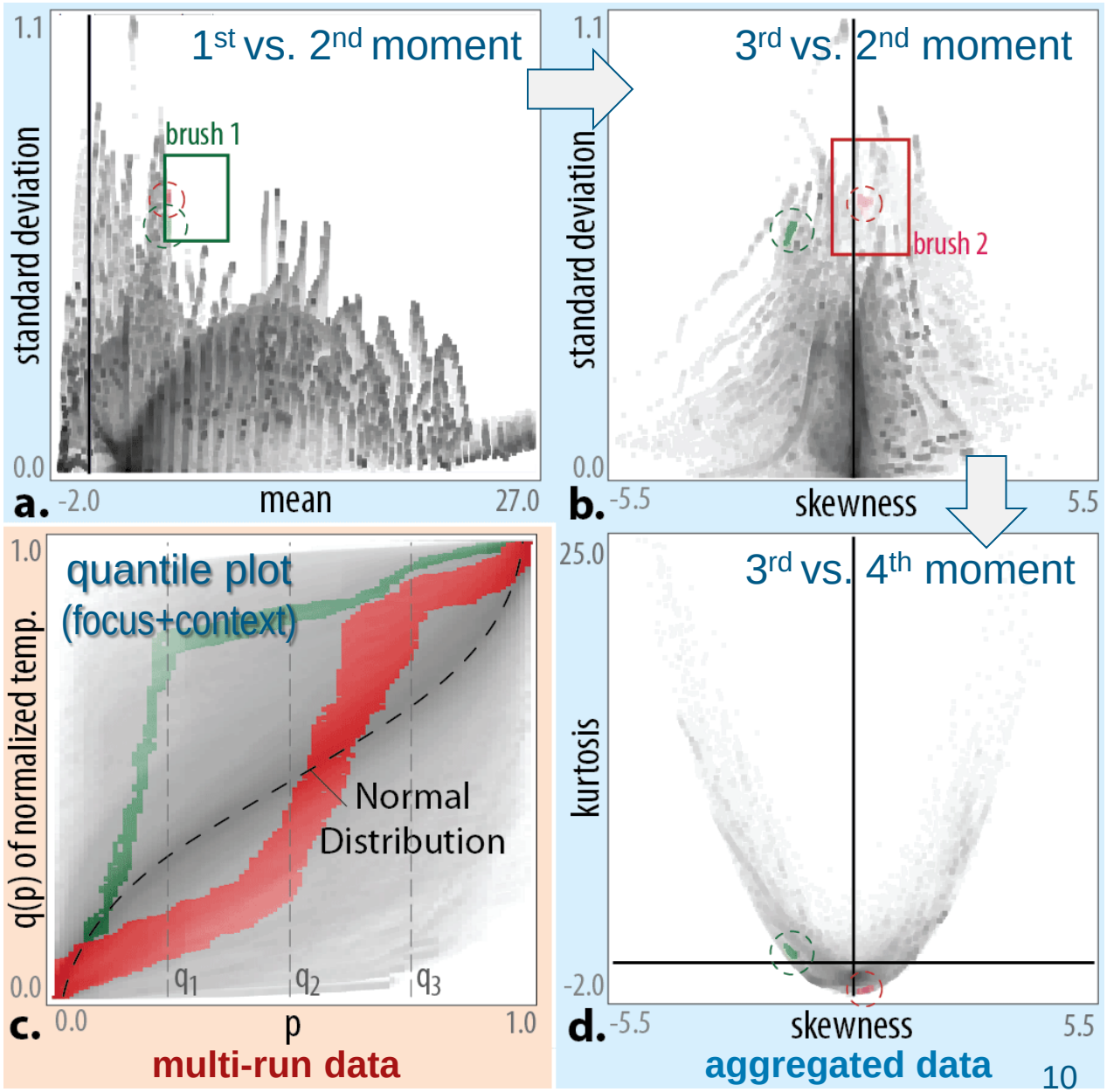
Basic View Setup: Opposing Different Moments

change order of moment



→ study relations btw. moments

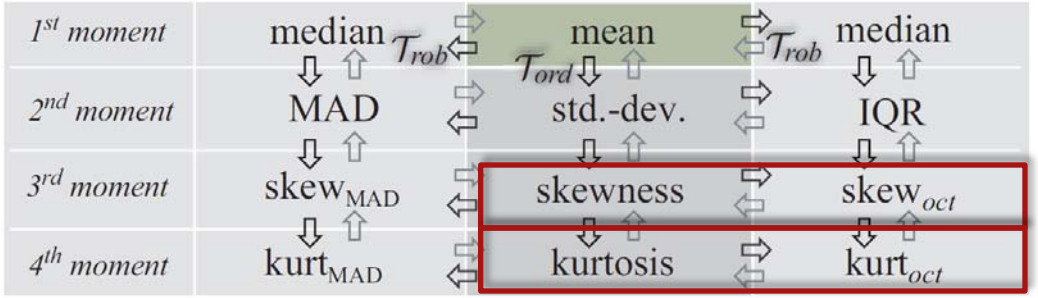
→ investigate basic characteristics of distributions

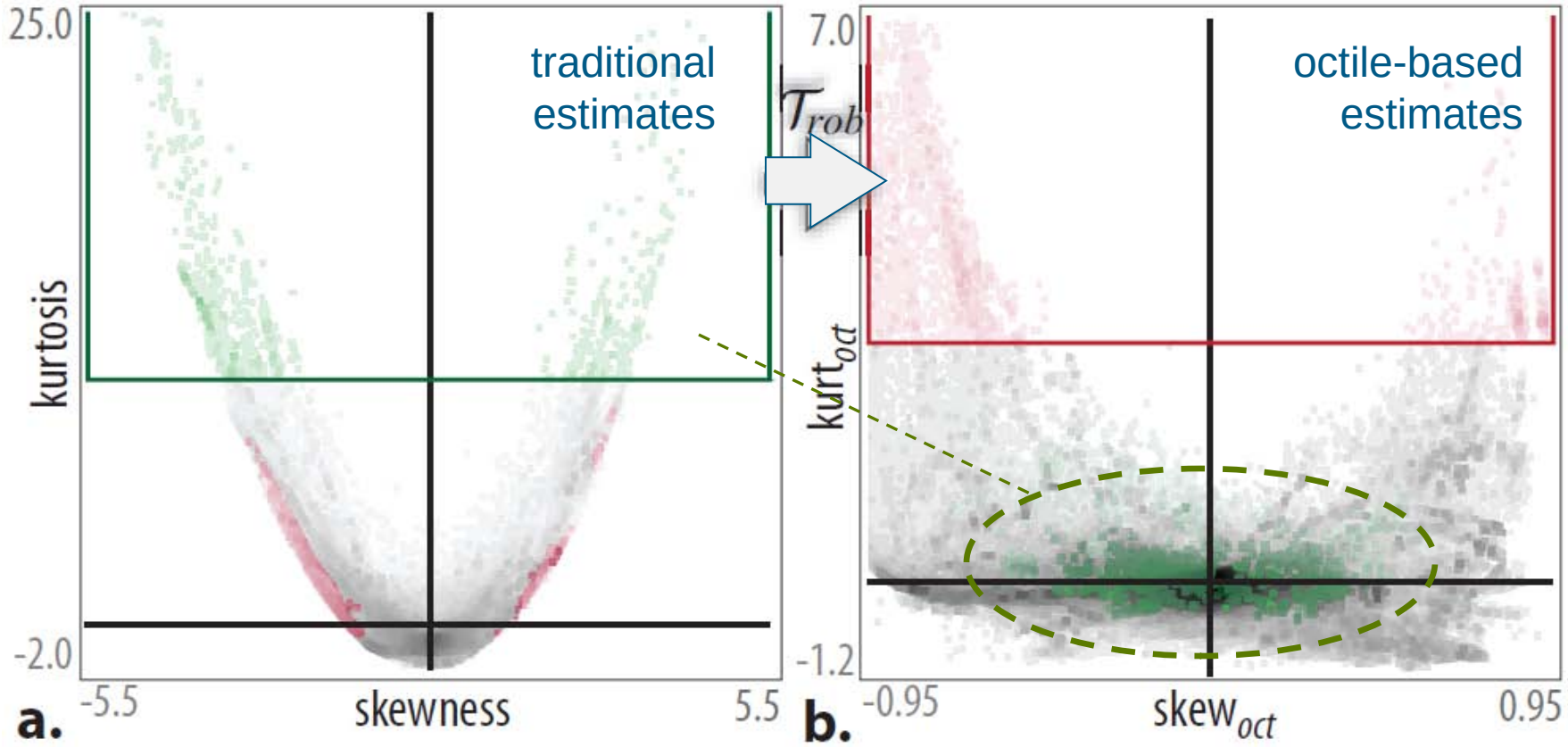


Views: Opposing Different Moments

robustify moment

→ assess influence of outliers



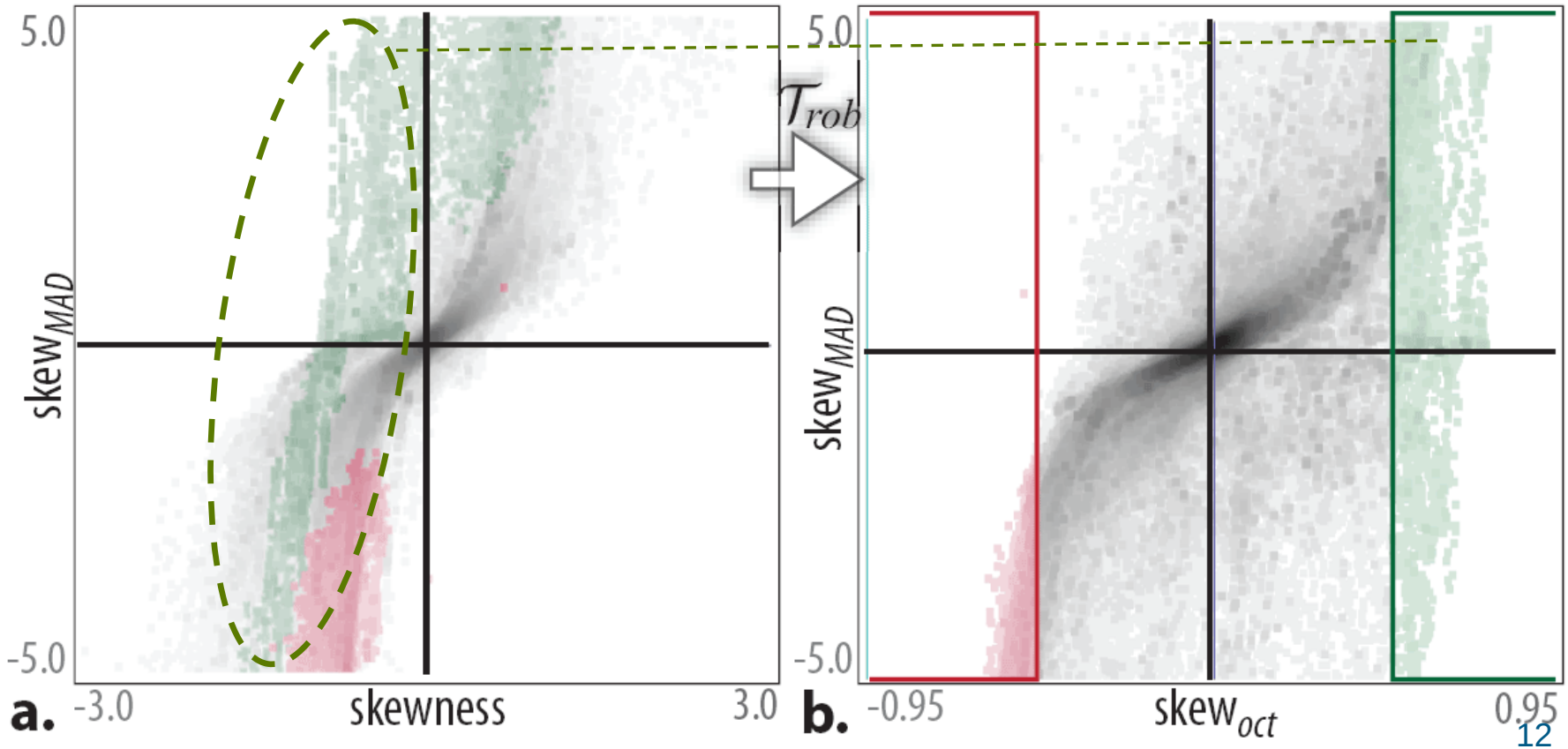


Views: Traditional vs. Robust Estimates

robustify moment

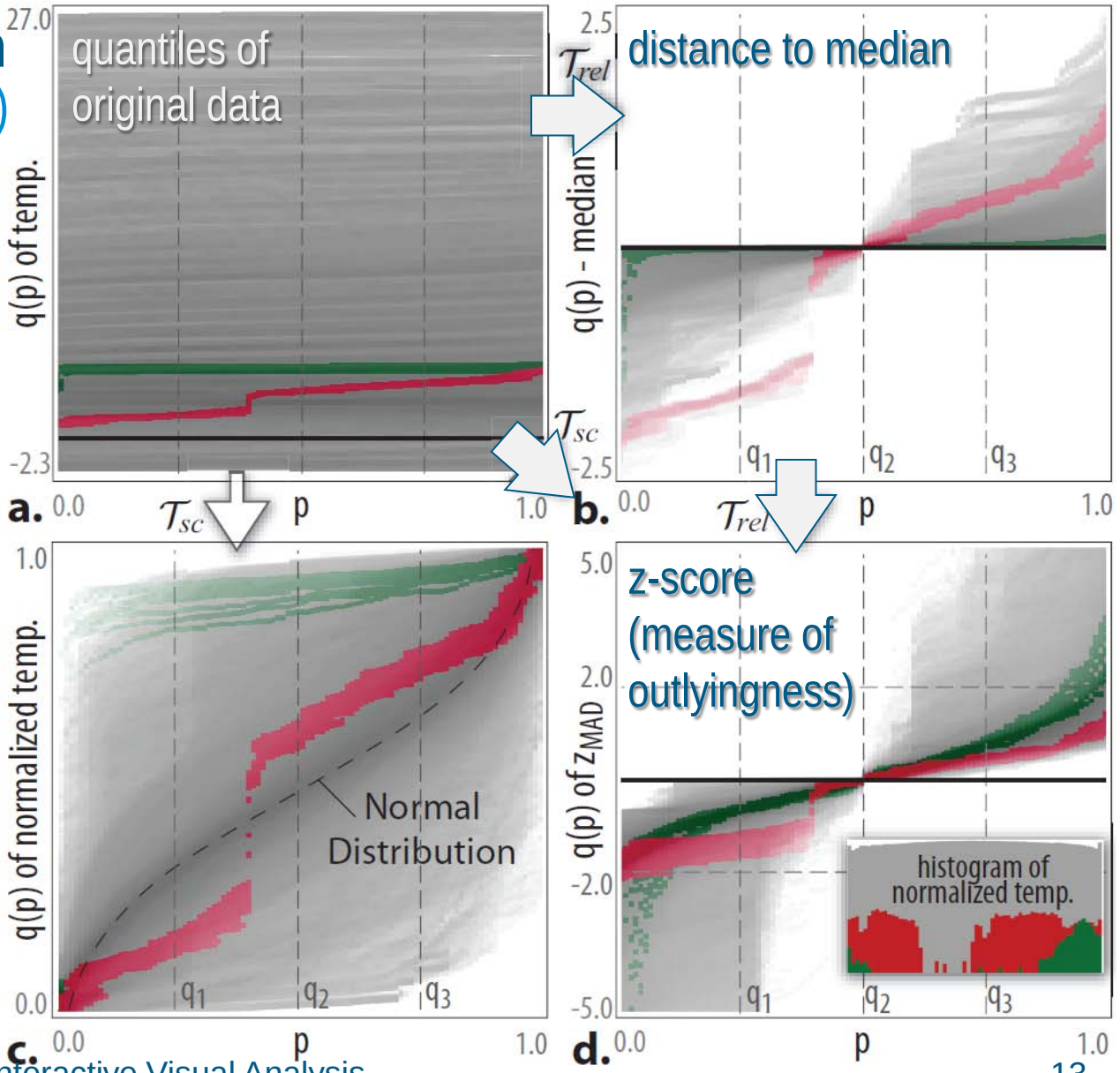
→ assess influence of outliers

1st moment	median	mean	median
2nd moment	MAD	std.-dev.	IQR
3rd moment	$skew_{MAD}$	skewness	$skew_{oct}$
4th moment	$kurt_{MAD}$	kurtosis	$kurt_{oct}$



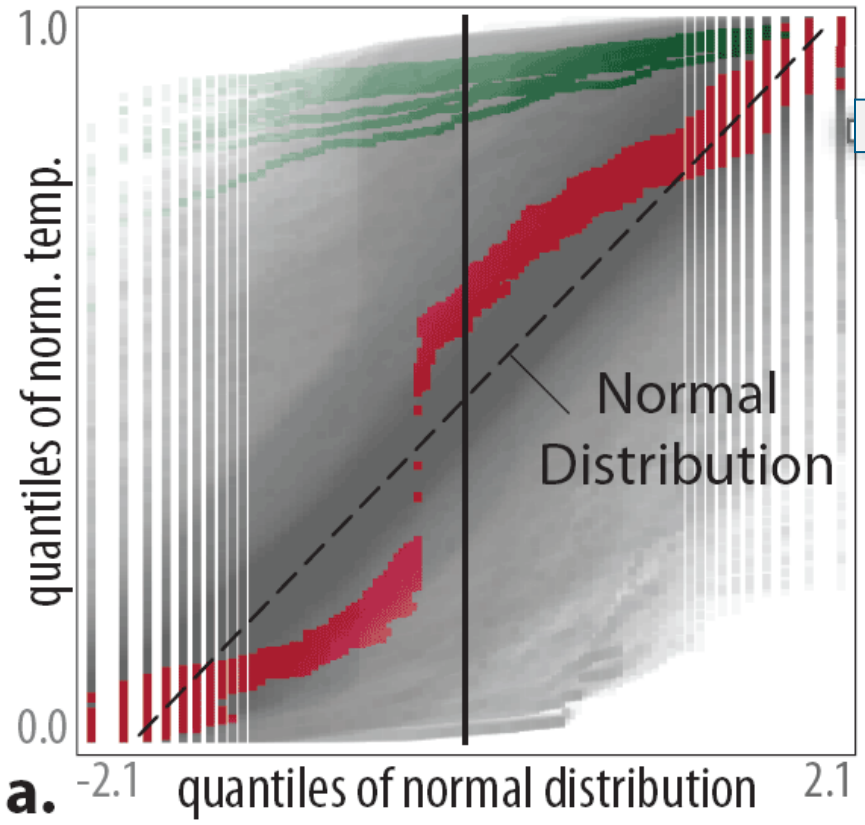
Other View Transformations

- **compute relation**
  (e.g., difference or ratio)
- **change scale**
  (e.g., z-standardization, normalize to [0,1])

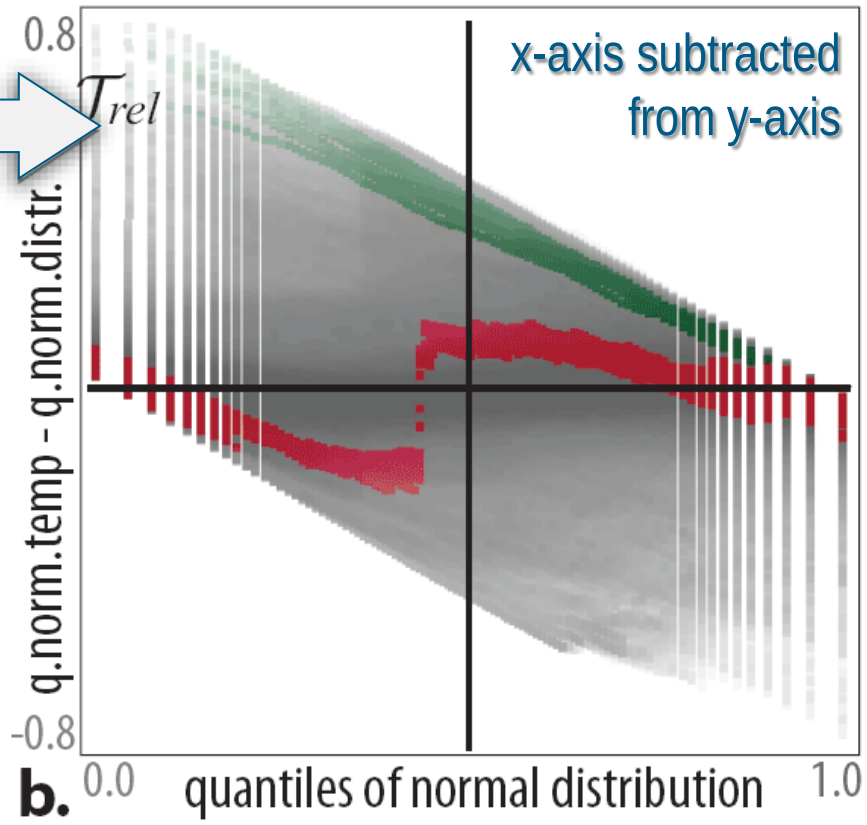


Other View Transformations

Q-Q plot



detrended Q-Q plot



change scale
→ compare with
theoretical distribution

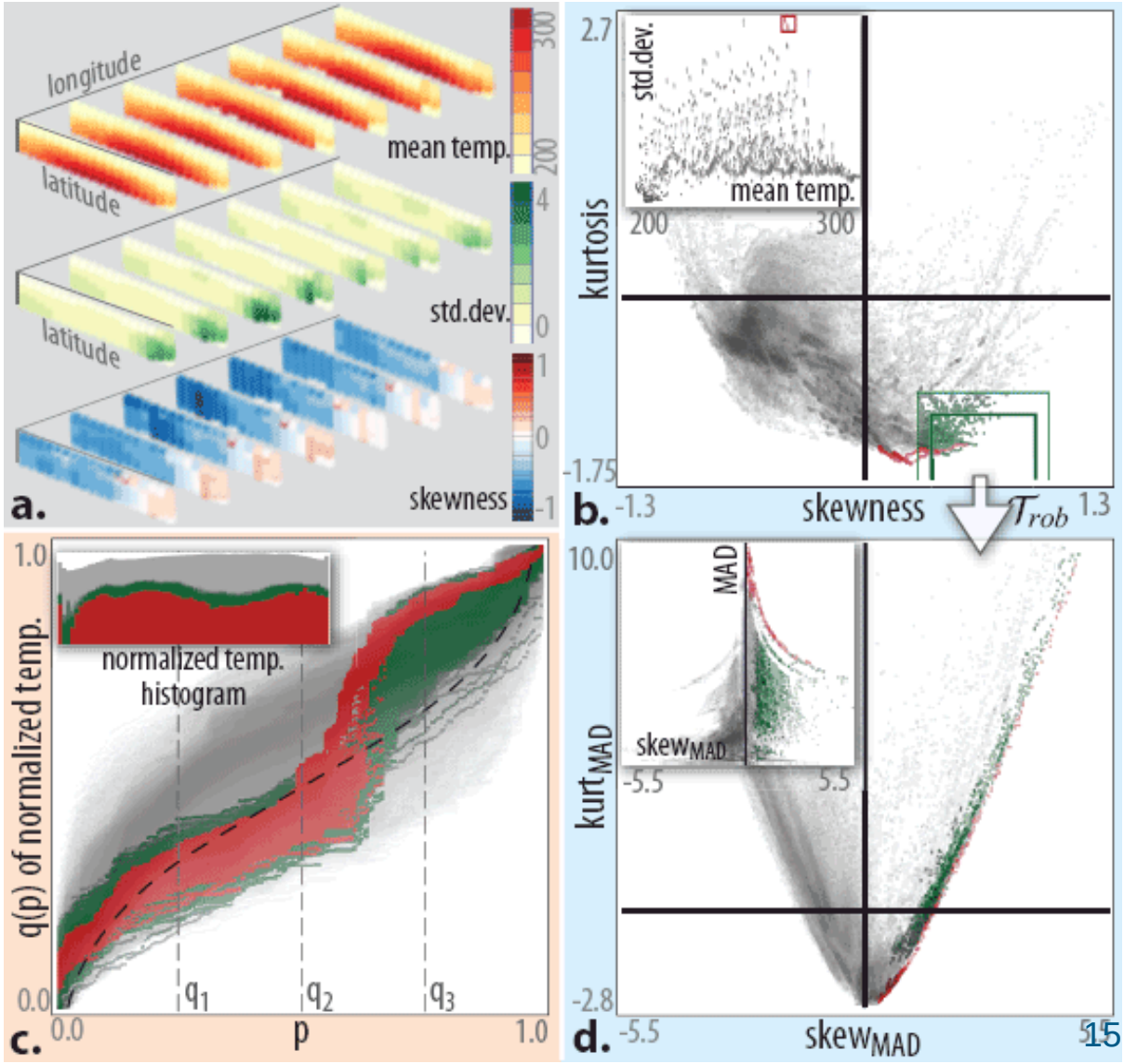
compute relation
(e.g., difference)

Analysis of Multi-run Climate Data

CLIMBER-2: Meltwater outburst of Lake Agassiz

(7 model parameters)

- 3D atmosphere
- 250 time steps
- 240 runs



Conclusion

- Data aggregation enables complex analysis
  - relating higher-dimensional data ⇔ aggregated statistics
  - analyst can work with both (original data, stat. properties)
- Traditional and robust estimates of moments
  (many opportunities also create management challenge)
- Iterative view transformations
  - helps analyst to maintain a mental model of views
  - matches the iterative nature of visual analysis
- Classification of informative views
  - opposing different moments
  - traditional vs. robust estimates of same moment

Acknowledgements

- Helmut Doleisch, Philipp Muigg
  www.SimVis.at
- Thomas Nocke, Michael Flechsig (PIK)

  Datasets are courtesy of

POTSDAM INSTITUTE FOR CLIMATE IMPACT RESEARCH

- VisGroup in Bergen,
  especially Armin Pobitzer, Stian Eikeland