Interactive Visual Analysis of Complex Climate Data

Johannes Kehrer[1,2]

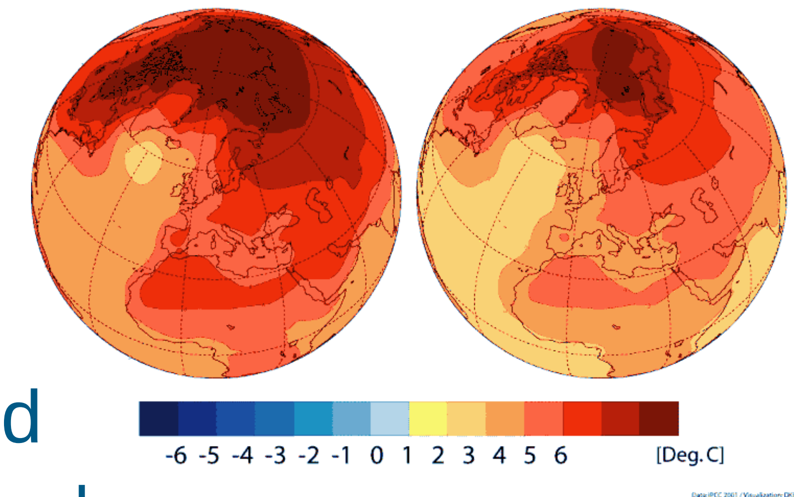
[1] Institute of Computer Graphics and Algorithms,
Vienna University of Technology, Austria
[2] Department of Informatics, University of Bergen, Norway

Motivation

- Insight into climate system

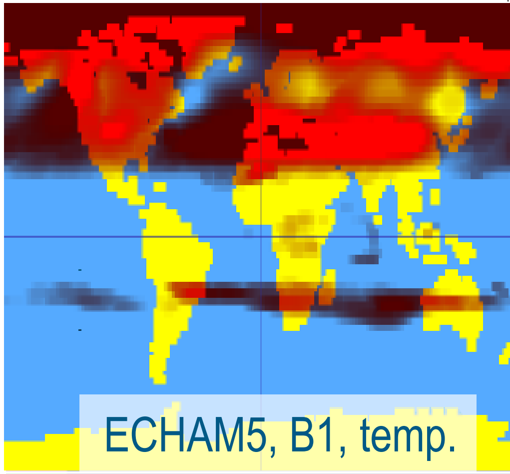
  → measurements & simulations (e.g., global climate models)

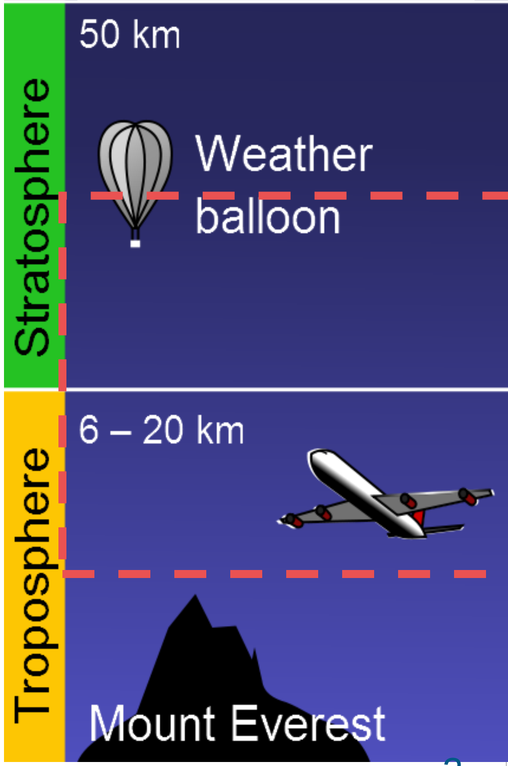
- Challenges
  - large, multi-variate data
  - time-dependent scenarios
  - deficiencies within data



- Difficult to analyze / understand
  - usually statistical methods used
  - require prior knowledge
  - challenging to perform a flexible analysis

- **Hypothesis Generation**
  [Kehrer et al. 2008, Ladstädter et al. 2009, 2010]
  - search for potential sensitive & robust **indicators** for climate change
  - characteristic climate signals that deviate from natural variability
  - useful to monitor atmospheric change
- Upper troposphere-lower stratosphere
  - known to be sensitive
  - investigate key climate parameters





Usual Workflow

- Set research focus
- Acquire data
- Iterate
  - explore / investigate data
  - formulate particular hypothesis
  - evaluate with statistics

Challenging to come up with new hypotheses

Goal: accelerate process (fast interactive visualization, more informed partner → more directed search)

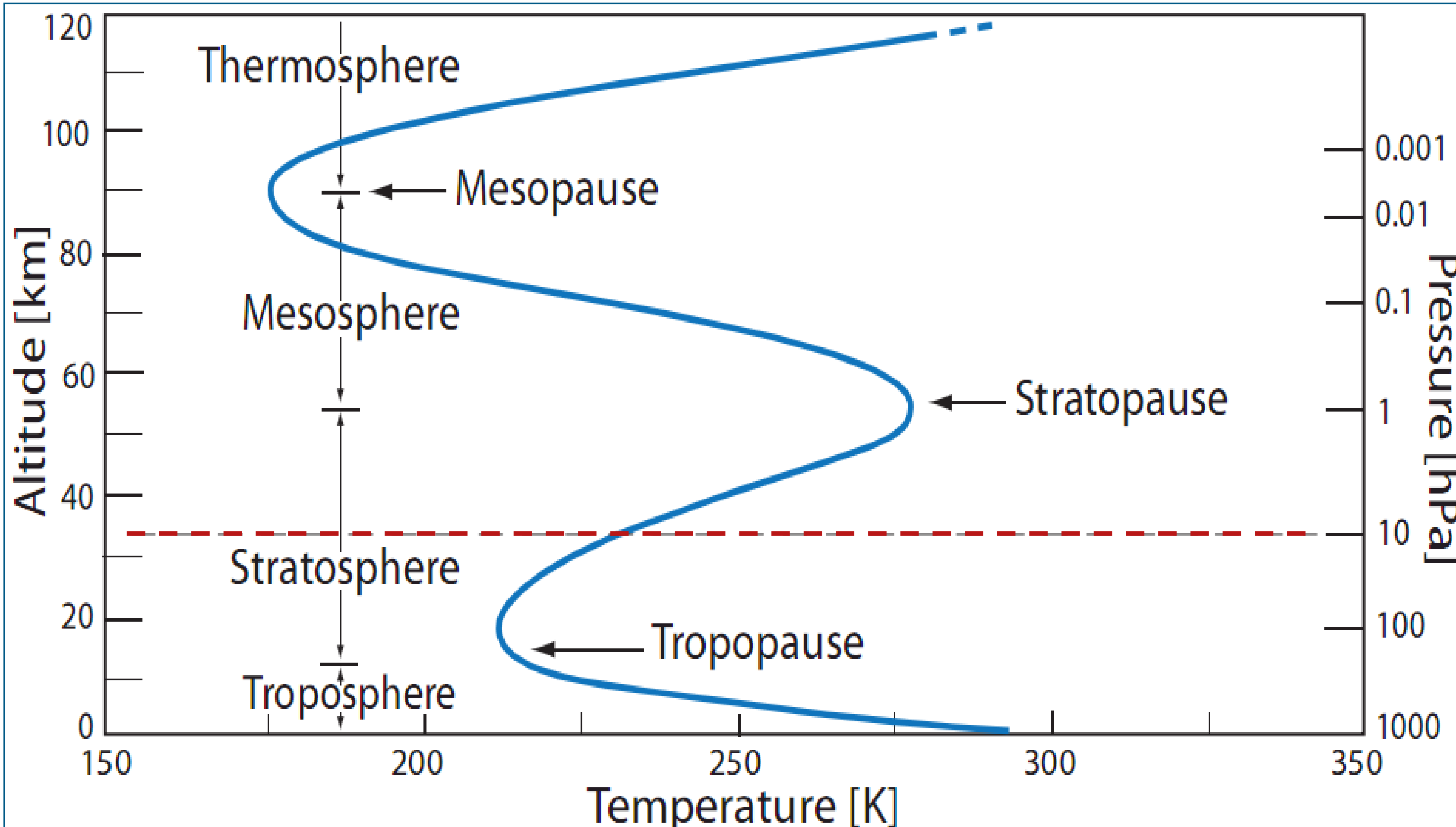
Data used in our Study

Climate Simulation Data

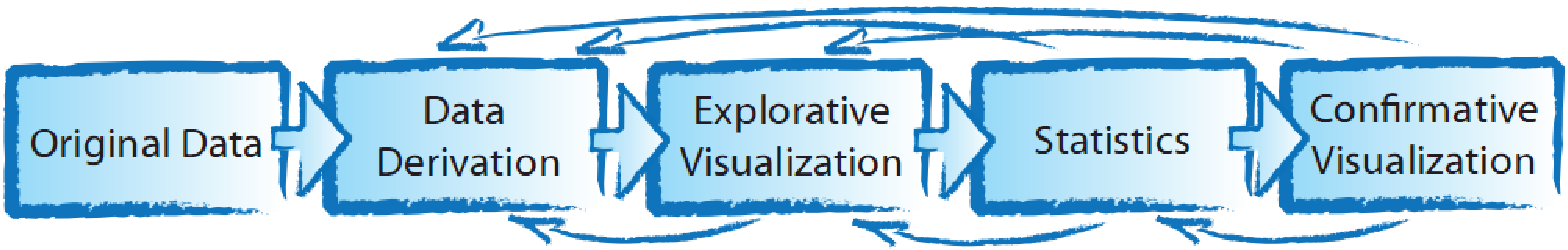
- ECHAM5 climate model, A2 scenario [MPI-M Hamburg] (IPCC 4th assessment report)
- temperature, years 1961–2061
- IPCC 20th century run before 2001
- 180.000 simulation cells
  → 2.5º x 2.5º, 18 pressure levels
- 108 time steps

Data used in our Study

Climate Simulation Data



Our Visual Exploration Process

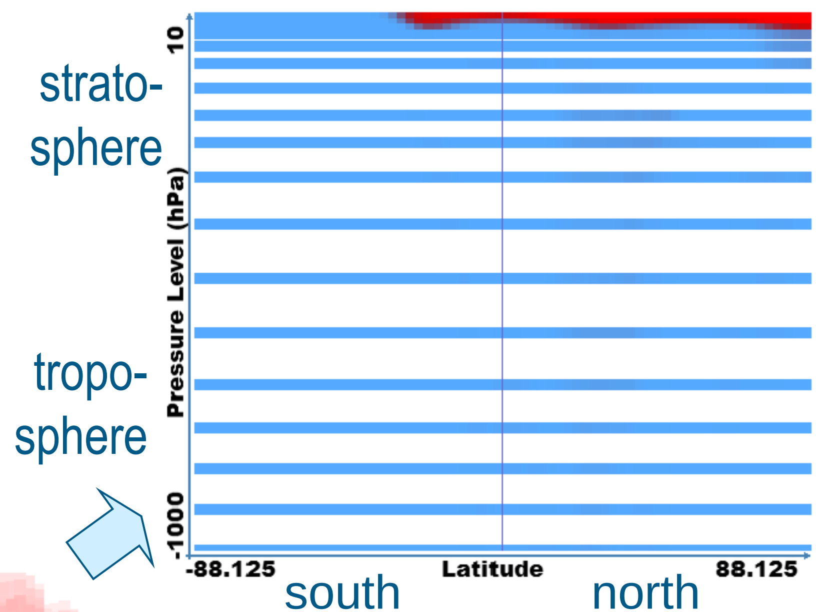


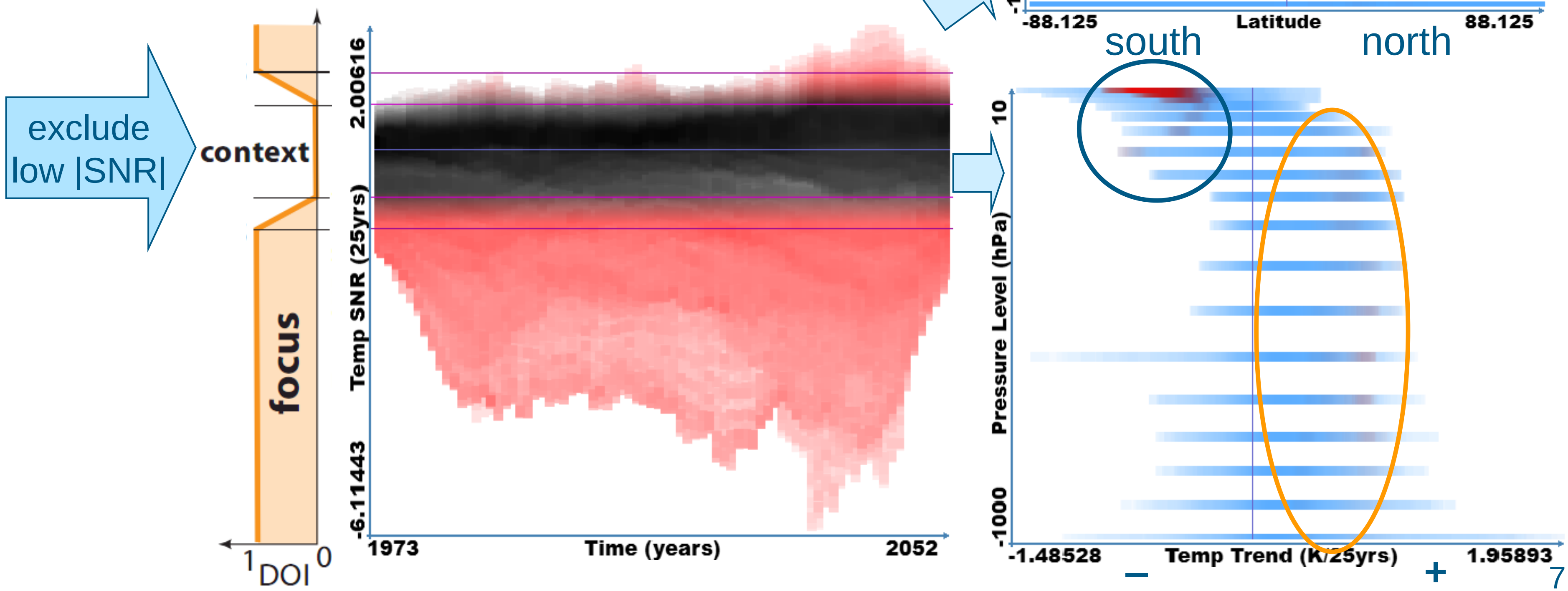
- Integrated **data derivation**
  → linear trends & signal to noise ratios (SNR)
- Interactive **visual exploration** for quick and flexible data investigation ("preview on statistics")
- Generated hypotheses evaluated using **statistics**
  → trend testing [Lackner et al. 08]
- Narrow down **parameters**

Focus on Expressive Data

Localize robust indicators

- areas with high significance
- smooth specification

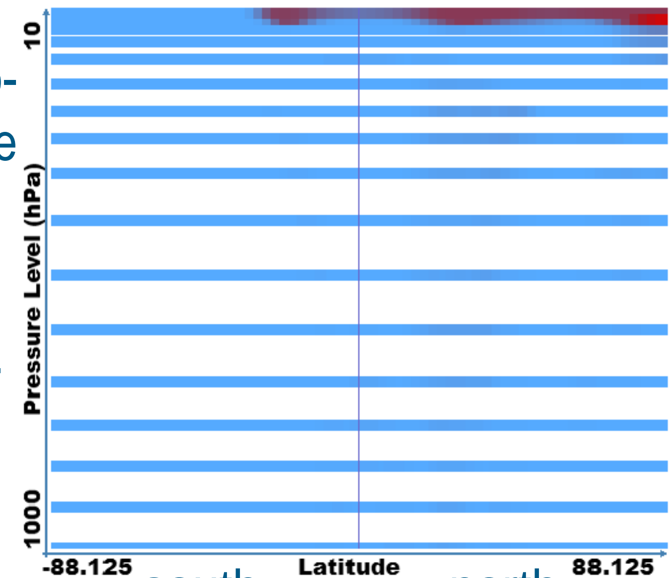


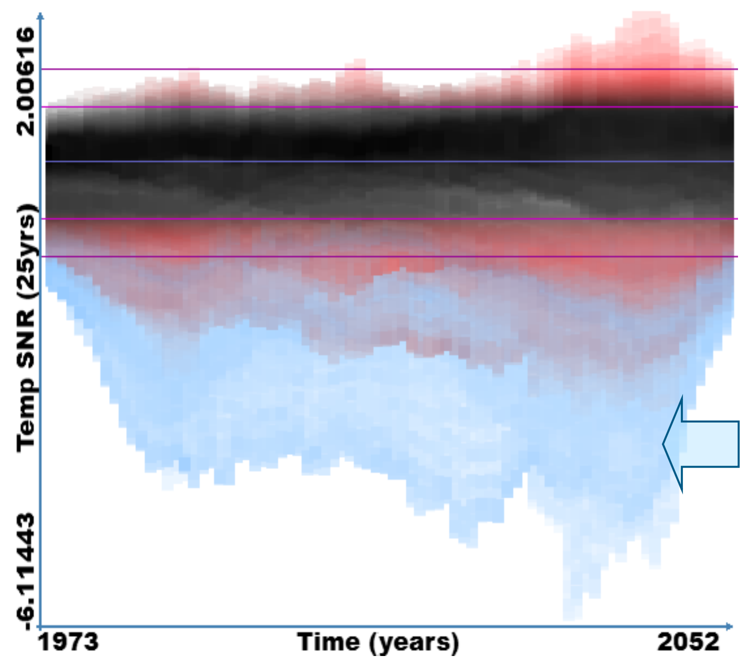


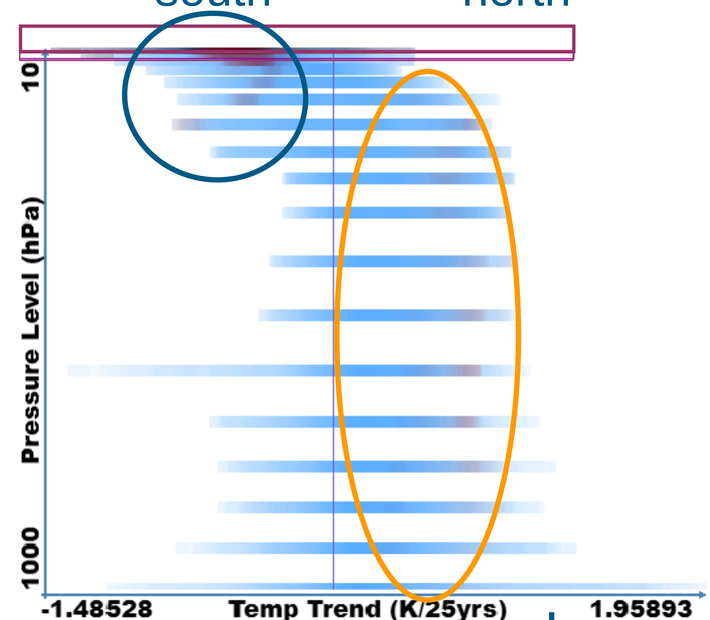
Further Refinement

Exclude upper pressure levels

→ known deficiencies
[Cordero & Forster ’06]



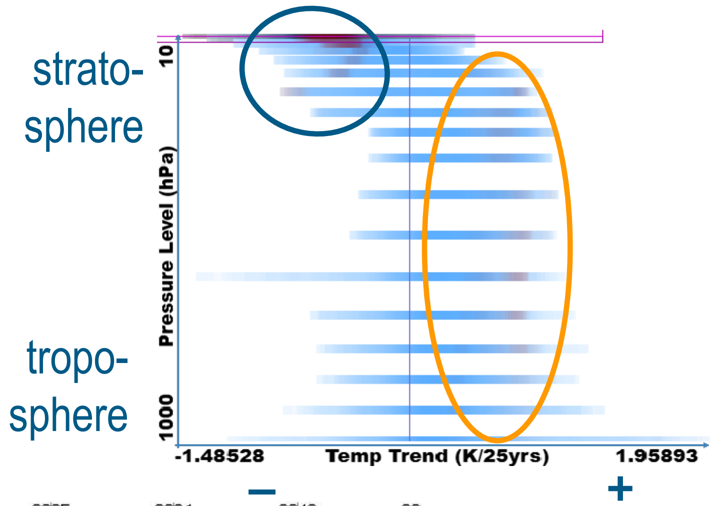


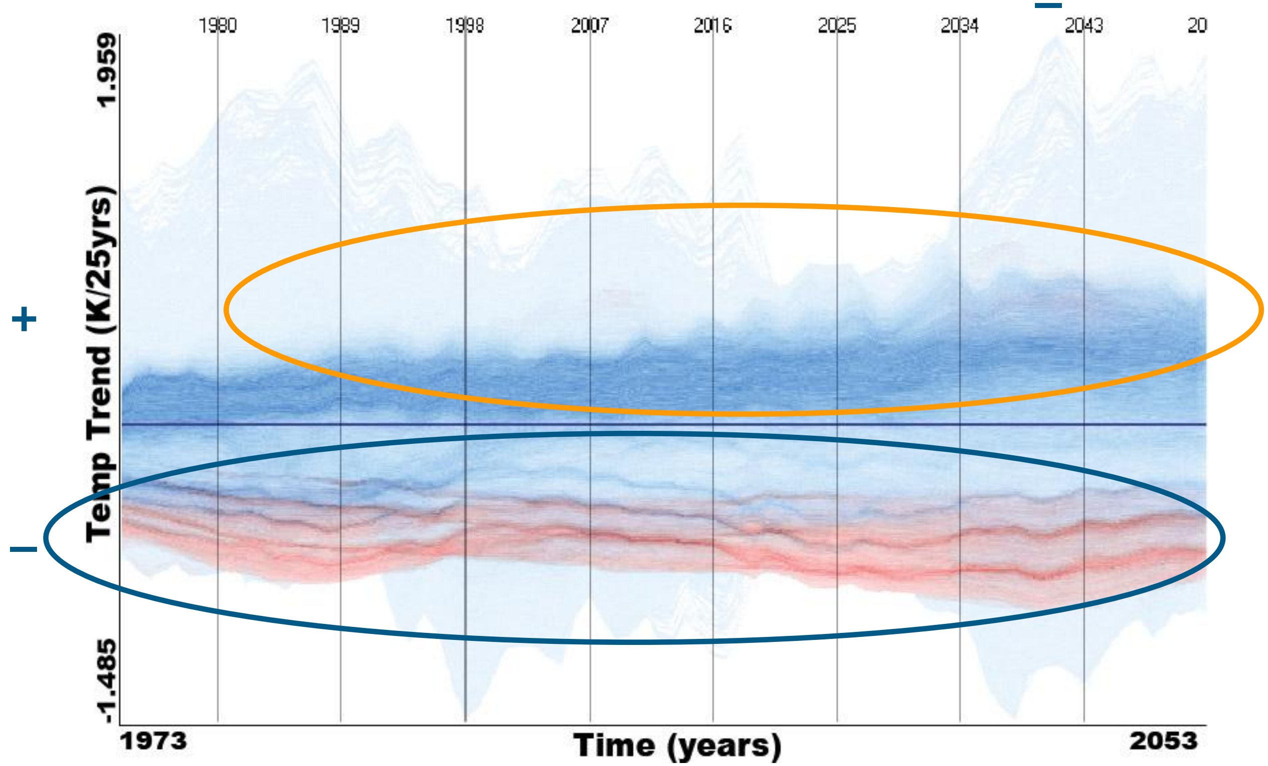


Kehrer et al.

Explore Trend Variation over Time

Robust cooling trends

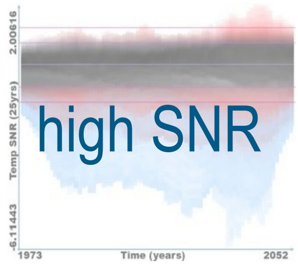


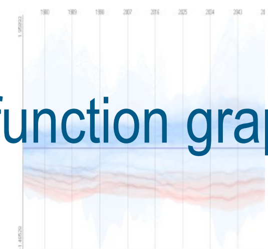


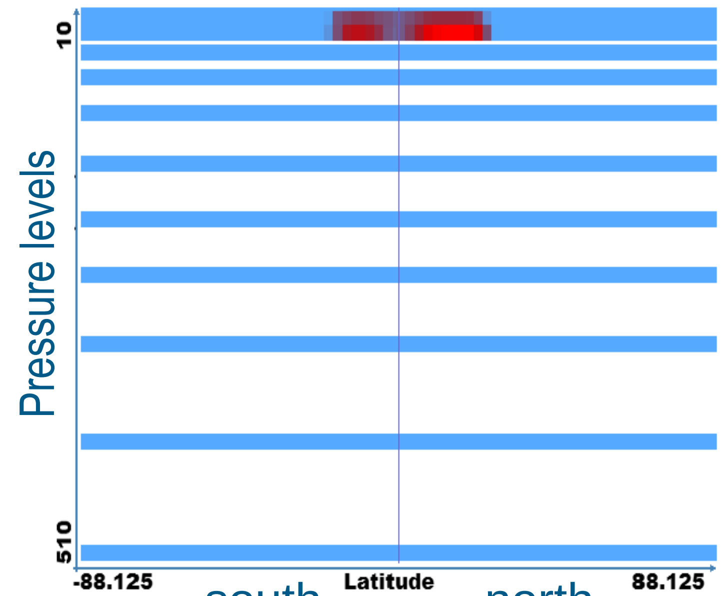
Relations between Dimensions

Up to now:

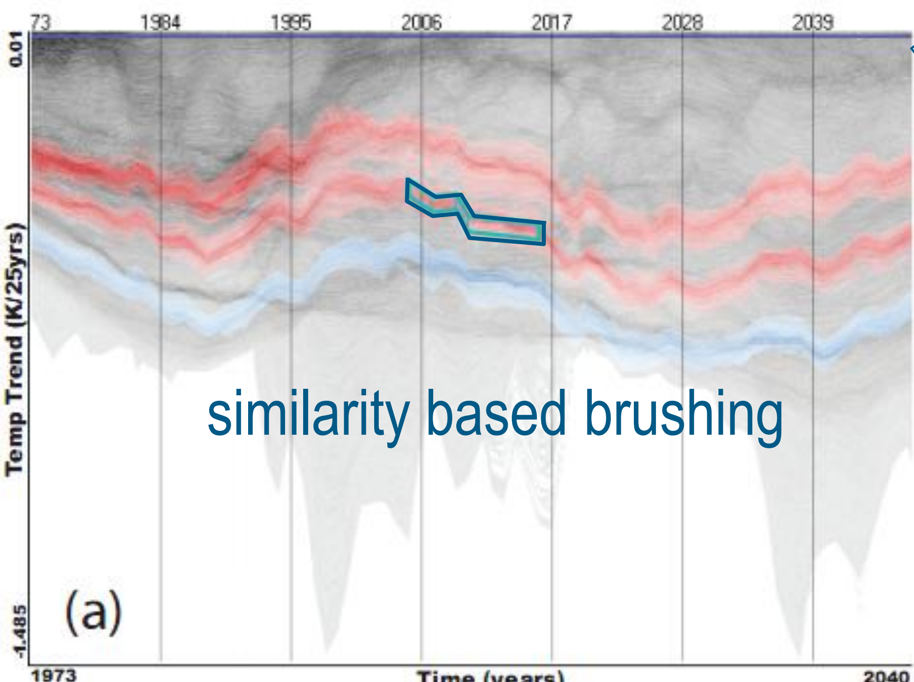
→ investigation in one direction

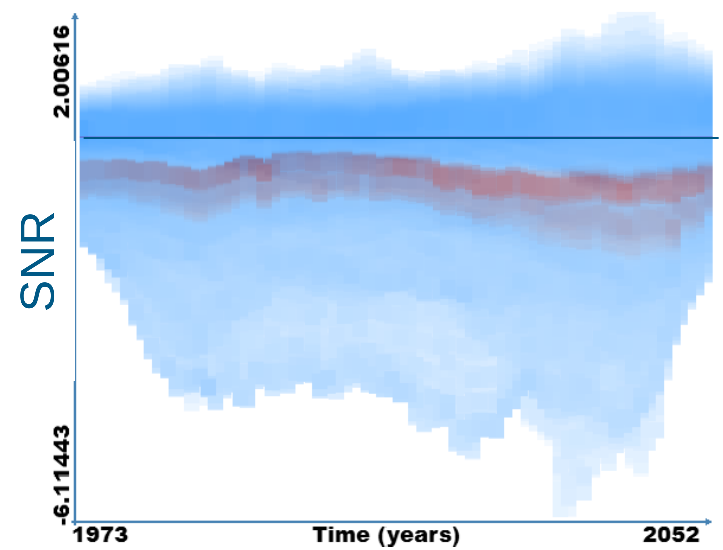






→ check relation in other direction

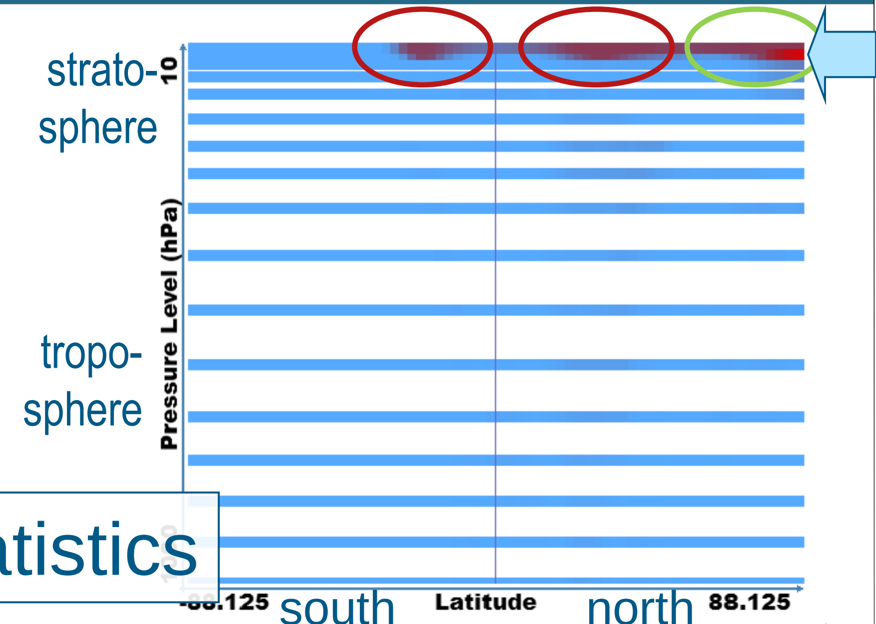


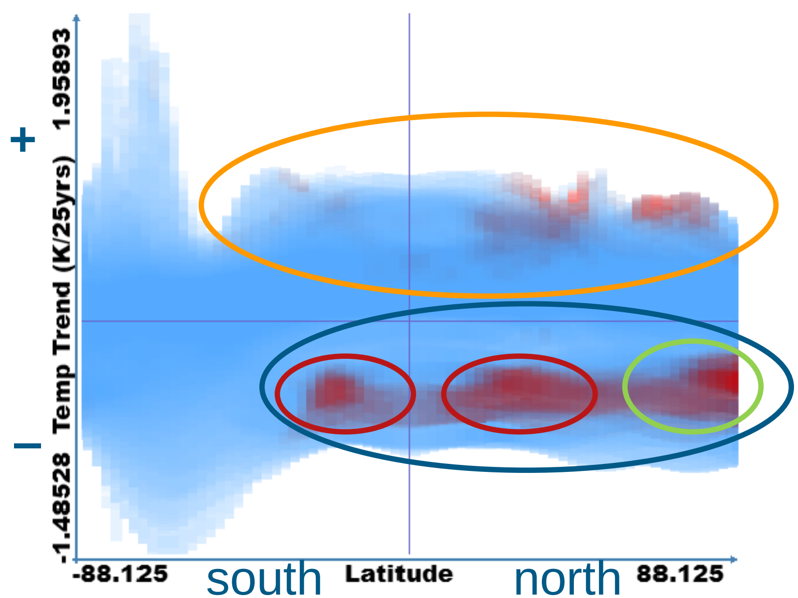


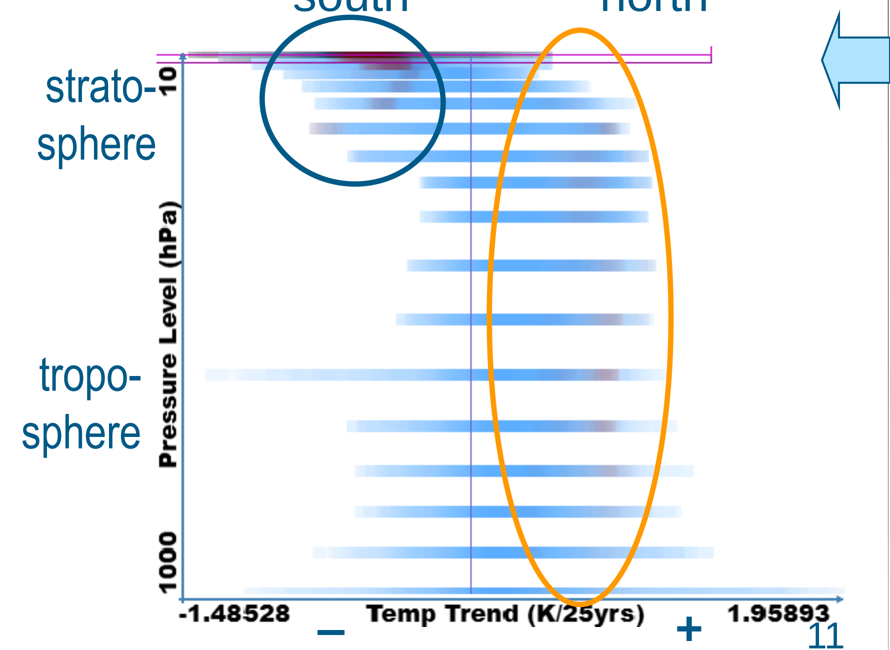
Generated Hypothesis/ECHAM5 temp.

Promising indicator region in **lower stratosphere** at **northern latitudes** & **tropics**. **Cooling trend** considered robust over investigated time span.

hypothesis handed over to statistics

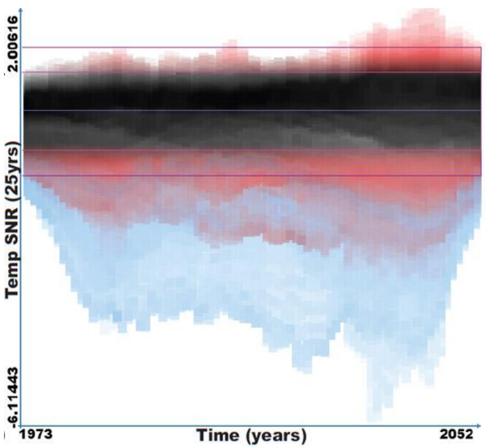


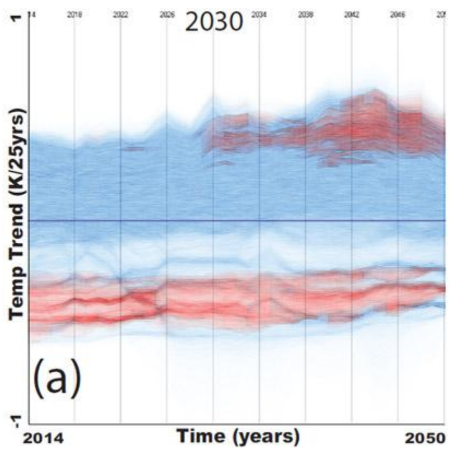


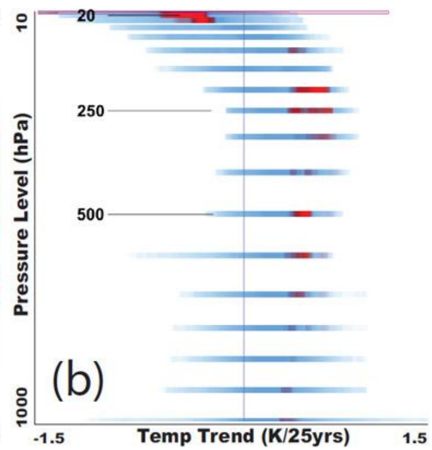


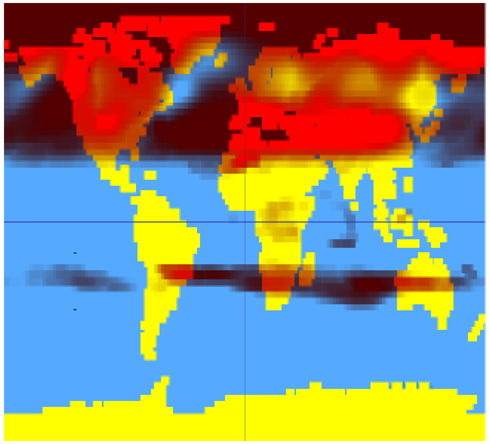
Hypothesis Generation with Visual Exploration

- Kehrer et al. **Hypothesis generation in climate research with interactive visual data exploration.** *IEEE TVCG, 14(6):1579–1586, 2008.*
- Ladstädter et al. **SimVis: an interactive visual field exploration tool applied to climate research**. In *New Horizons in Occultation Research,* pages 235–245. Springer, 2009.
- Ladstädter et al. **Exploration of climate data using interactive visualization**. *Journal of Atmospheric and Oceanic Technology, 27(4):667–679, 2010.*





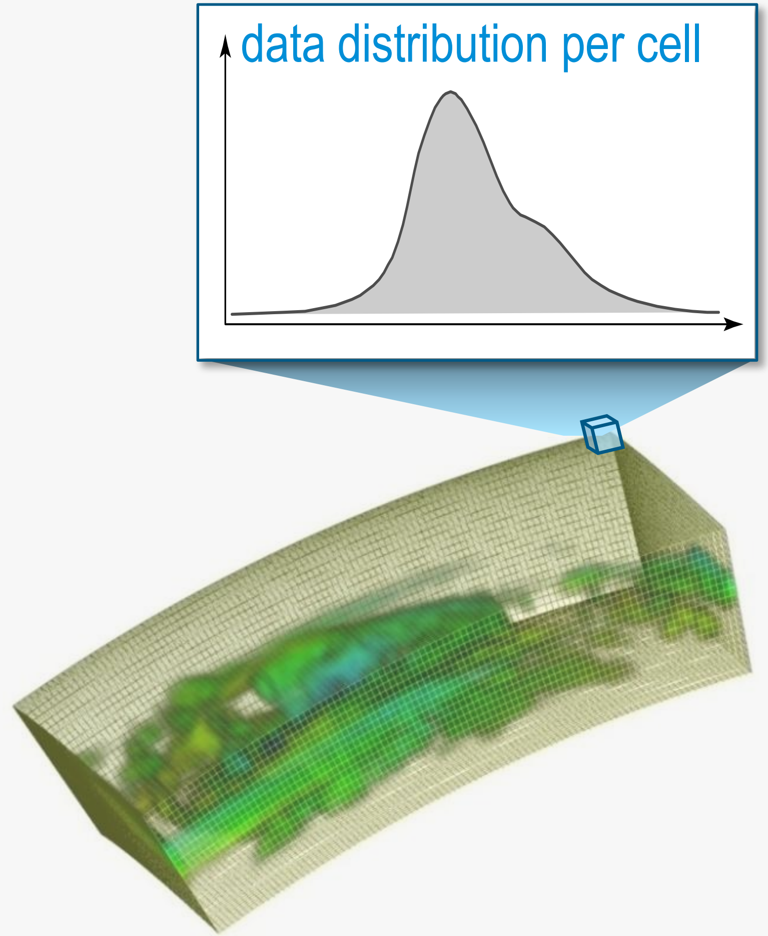




Higher-dimensional Scientific Data

- Data in scientific visualization
  - **data values d(x)**
    (e.g., temperature, pressure values)
  - measured/simulated wrt. a
    domain x (e.g., 2D/3D space,
    time, simulation input parameters)

- **Dimensionality reduction**
  (e.g., computing statistics wrt.
  time / spatial axes)

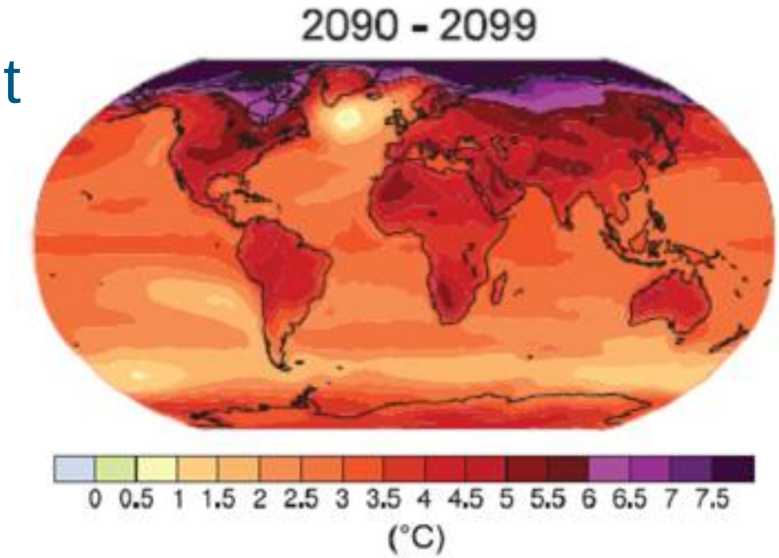


3D time-dependent
multi-run simulation data

Reducing the Data Dimensionality

[from IPCC AR #4, 2007]

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



average temp. in ten years

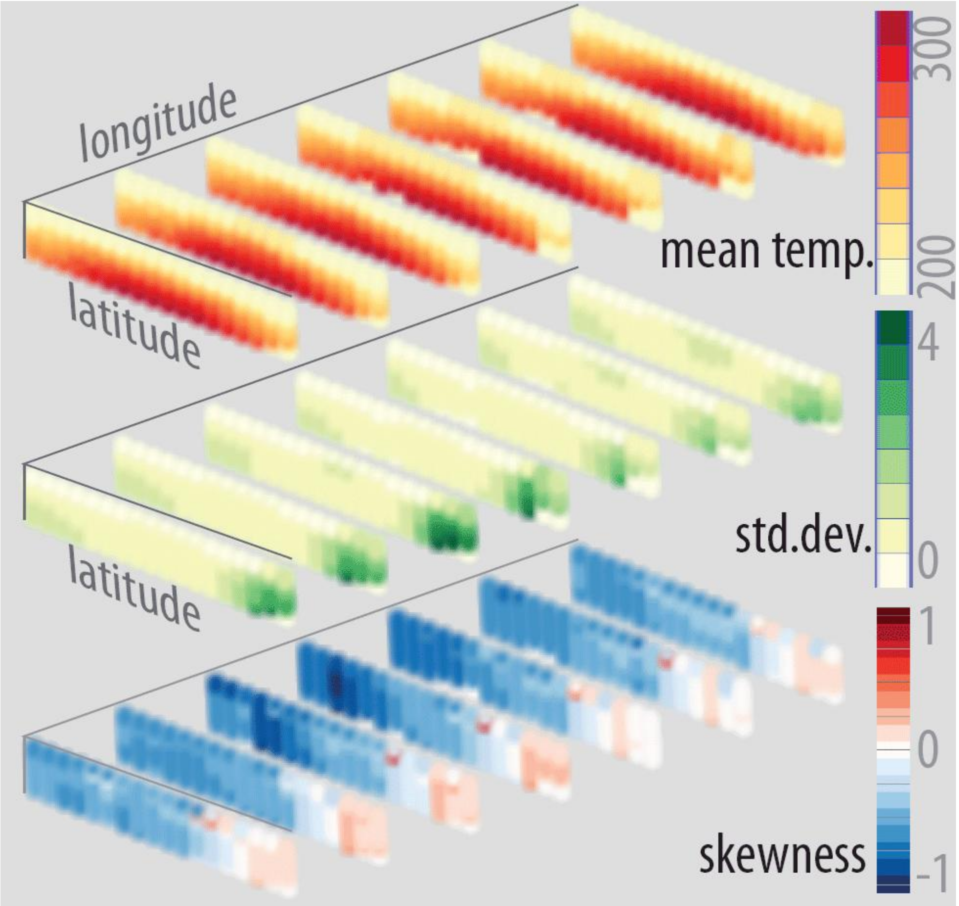
- Integrate into IVA through **attribute derivation**

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

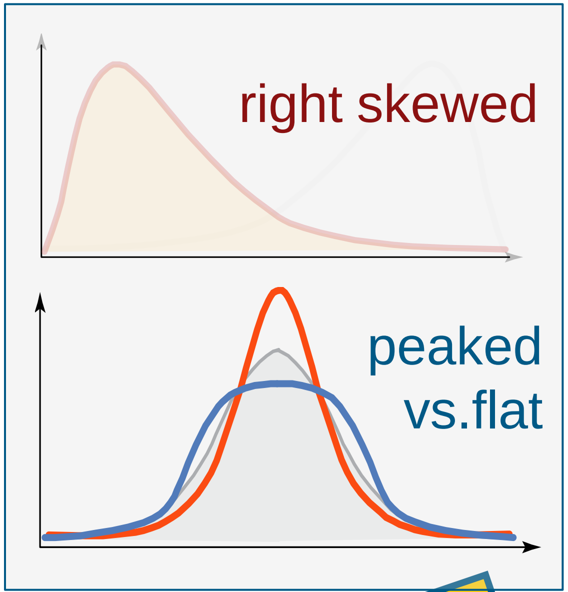


timestep 80

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

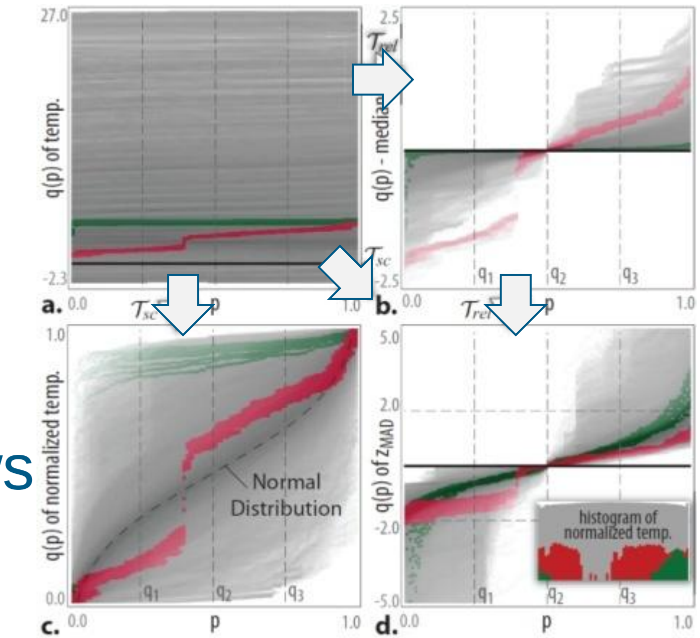


Structured approach to manage complexity

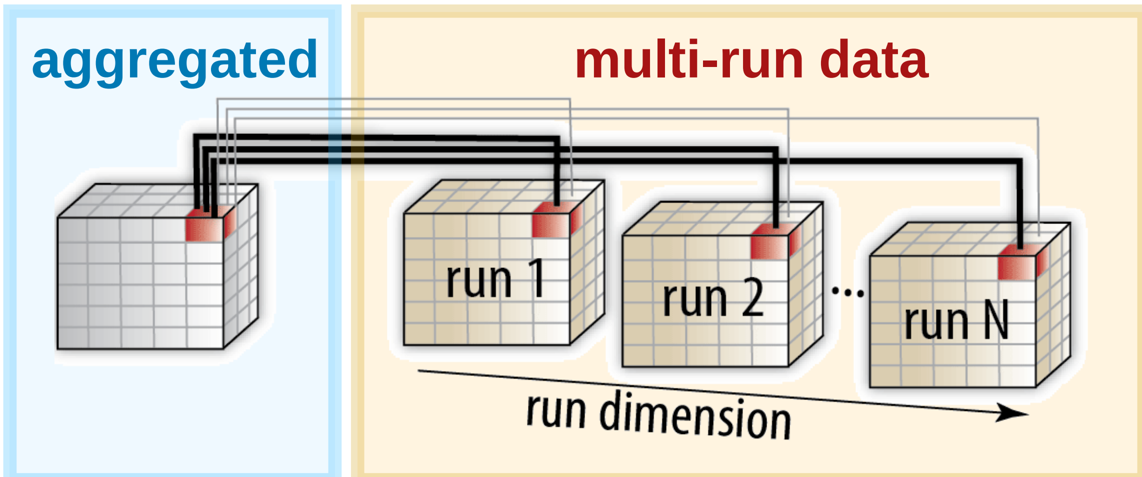
How to deal with this "management challenge"?

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

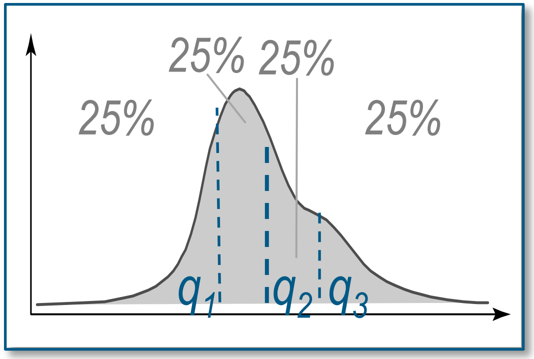


- One-to-many relation



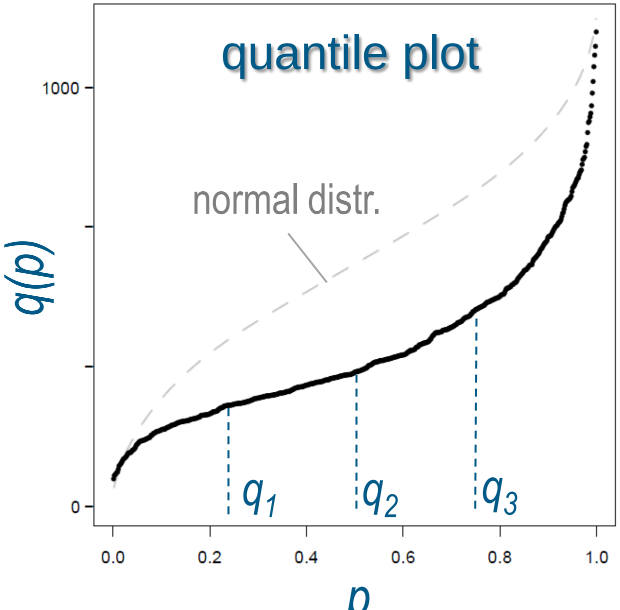
Visualizing Data Distributions

- 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 $\geq q(p)$



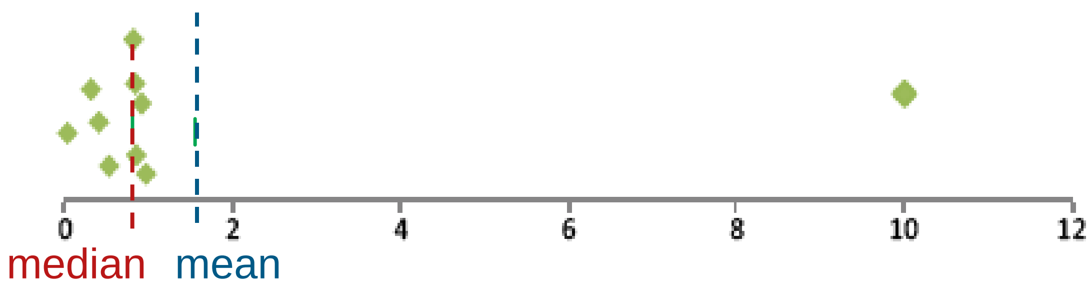
- Examlpes: median $q_2 = q(½)$, quartiles q_1, q_3

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

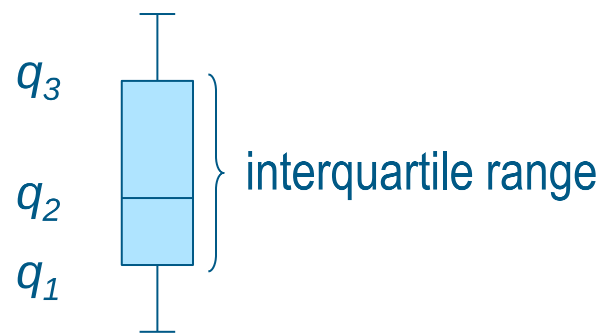


Robust Statistics

Outlier influence traditional estimates



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



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

- Estimates of skewness
  - traditional

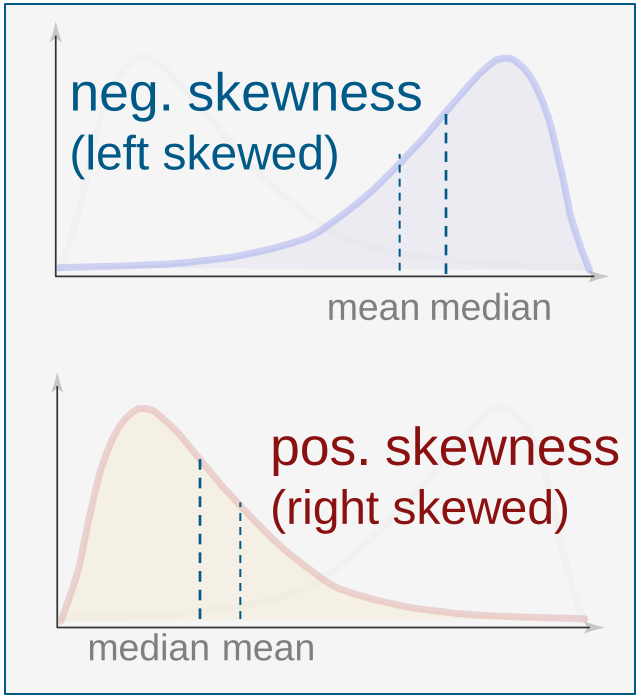
$$skewness = \frac{1}{n}\sum_{i=1}^{n}\frac{(x_i - \bar{x})^3}{s^3}$$

  - median/MAD-based

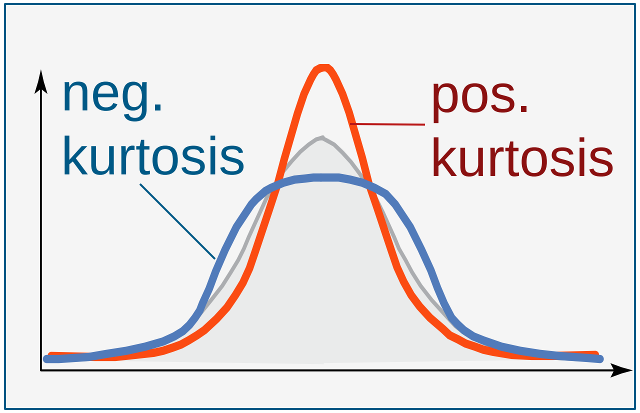
$$skew_{MAD} = \frac{1}{n}\sum_{i=1}^{n}\frac{(x_i - median)^3}{MAD(x_1,\ldots,x_n)^3}$$

  - 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**

	med/MAD-based	traditional	octile-based
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}

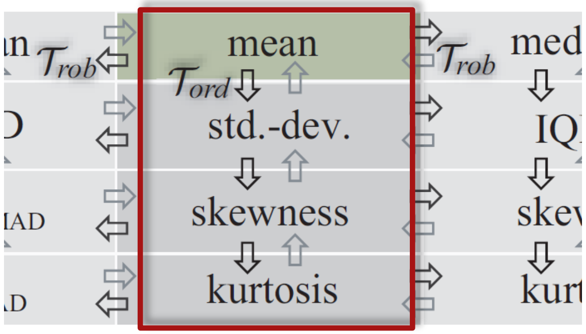
$\mathcal{T}_{rob}$ $\mathcal{T}_{ord}$ $\mathcal{T}_{rob}$

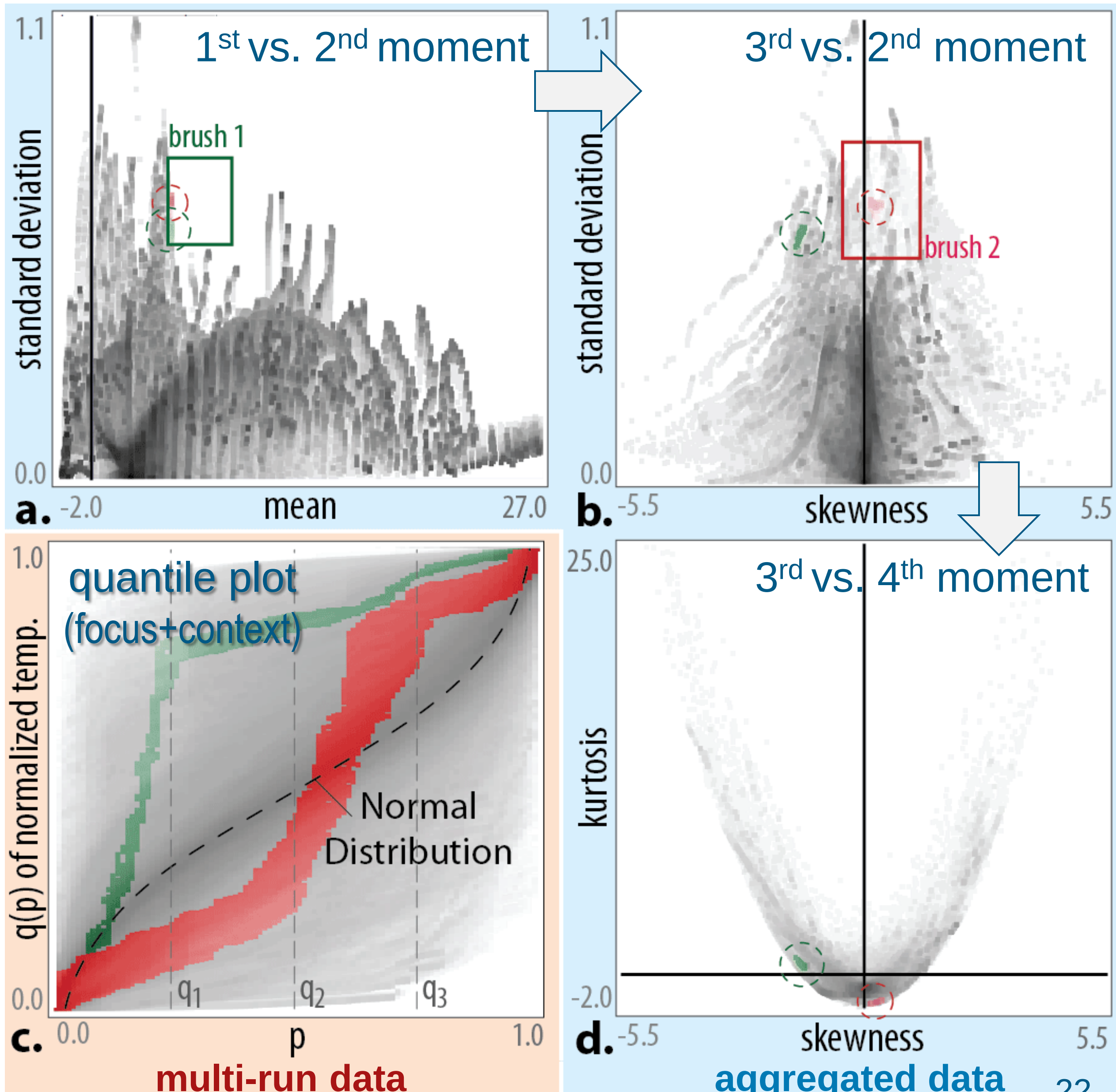
- **compute relation**
  (e.g., difference or ratio)
- **change scale**
  (e.g., normalize, z-standardization)

Closer related to
data tranformations

Basic View Setup: Opposing Moments

change order of moment





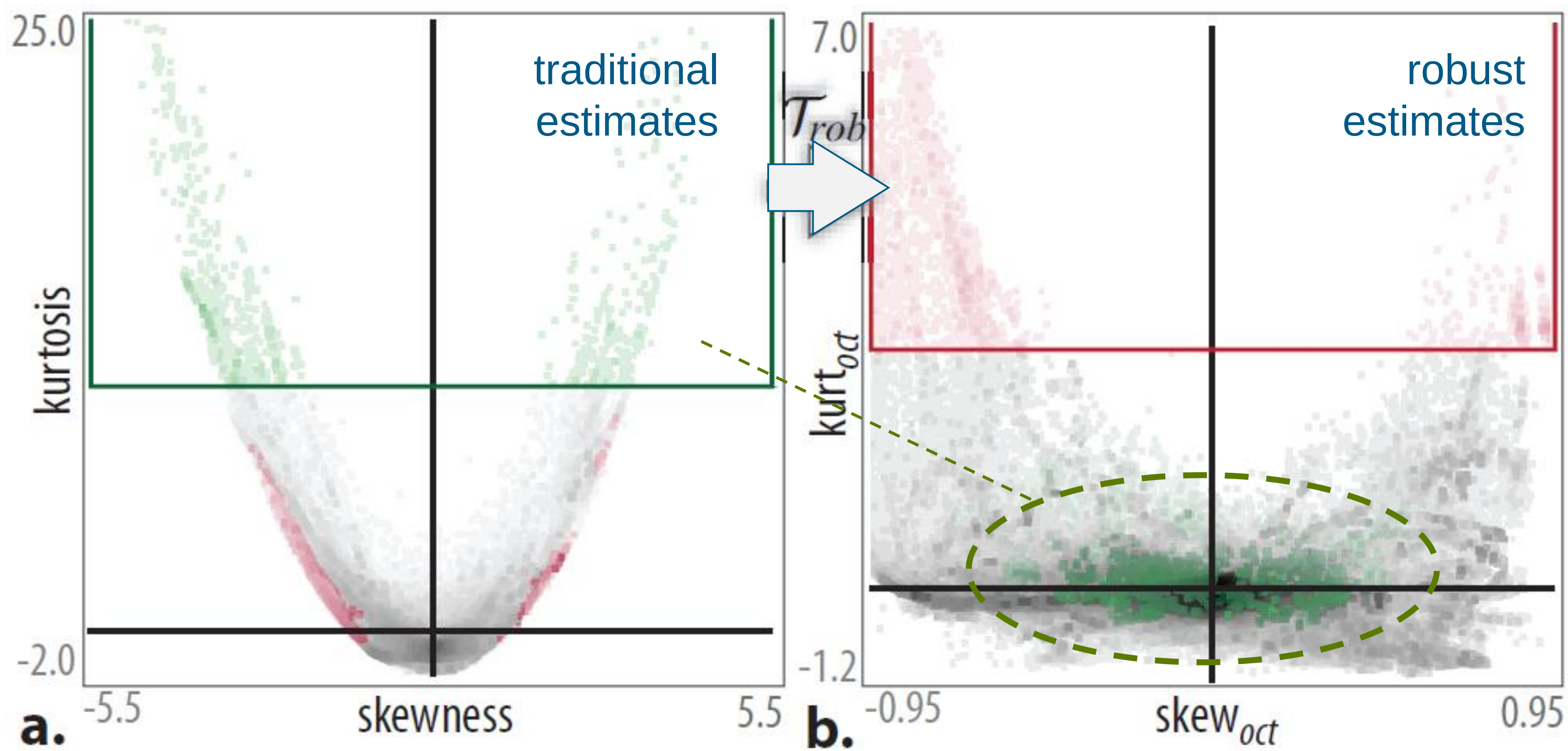
→ study relations betw. moments

→ investigate basic characteristics of distributions

robustify moment

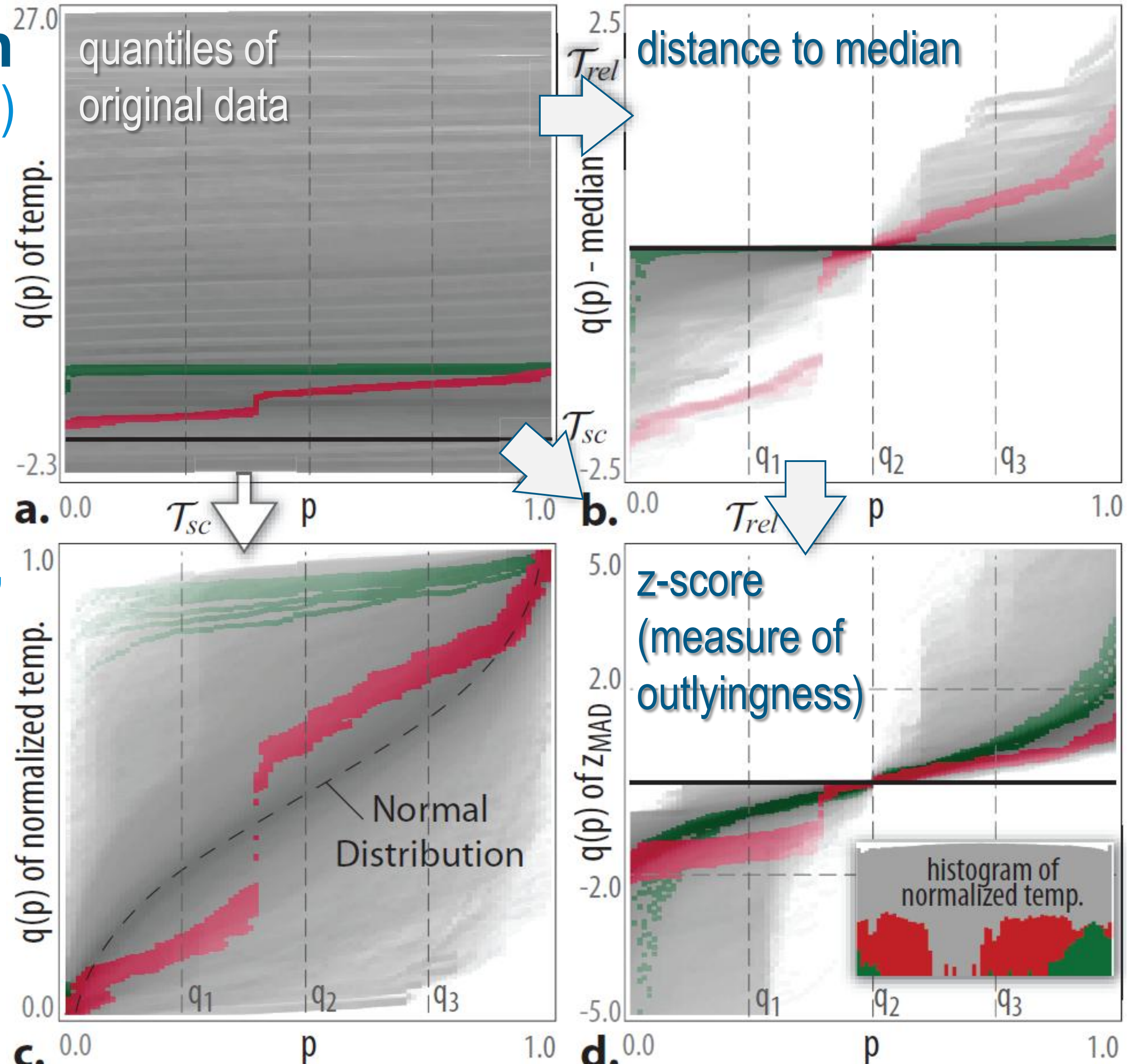
→ assess influence of outliers

1^{st} moment	median	$\mathcal{T}_{rob}$	mean $\mathcal{T}_{ord}$	$\mathcal{T}_{rob}$	median
2^{nd} moment	MAD		std.-dev.		IQR
3^{rd} moment	skew_{MAD}		skewness		skew_{oct}
4^{th} moment	kurt_{MAD}		kurtosis		kurt_{oct}

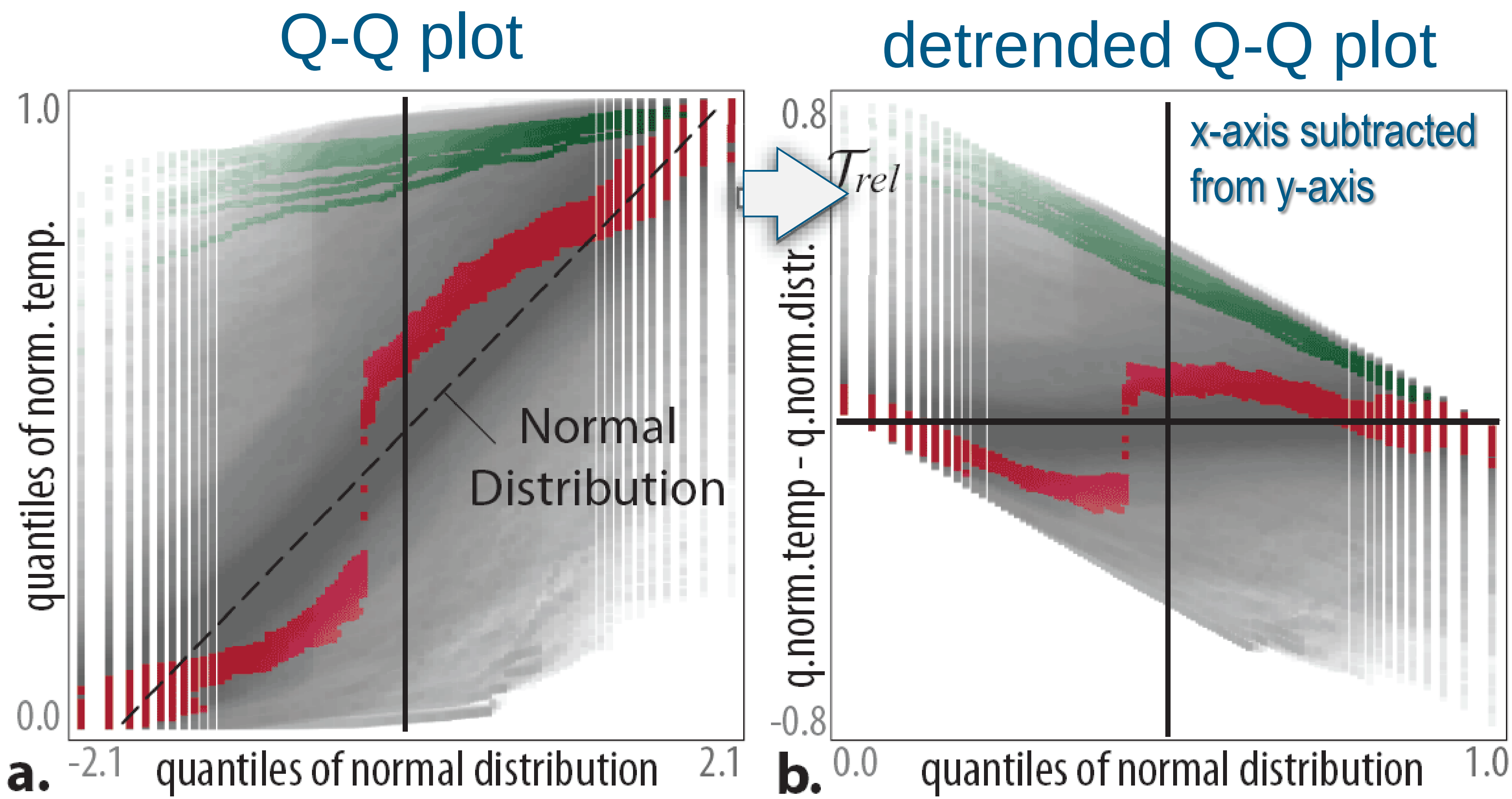


Other View Transformations

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



Other View Transformations

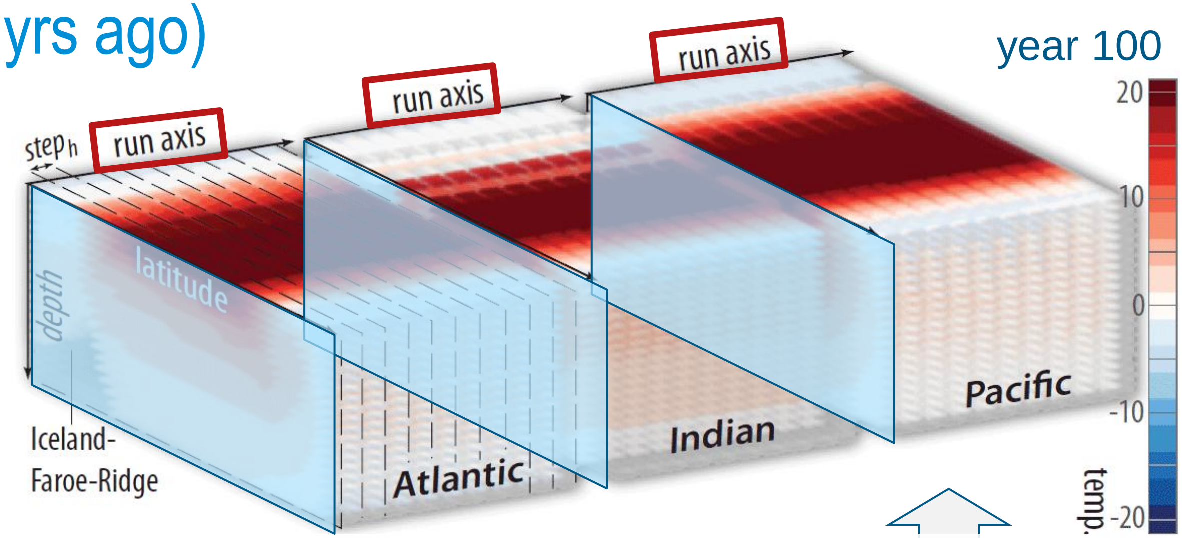


change scale
→ compare with
theoretical distribution

compute relation
(e.g., difference)

Showing Aggregated Statistics

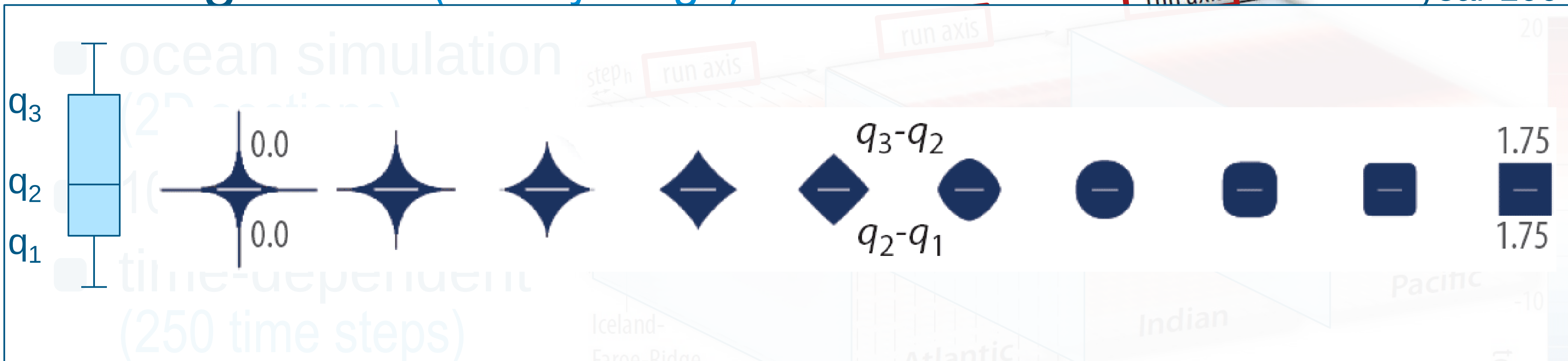
- Cooling event (8200 yrs ago)
  - ocean simulation (2D sections)
  - 10 x 10 = 100 runs
  - time-dependent (250 time steps)



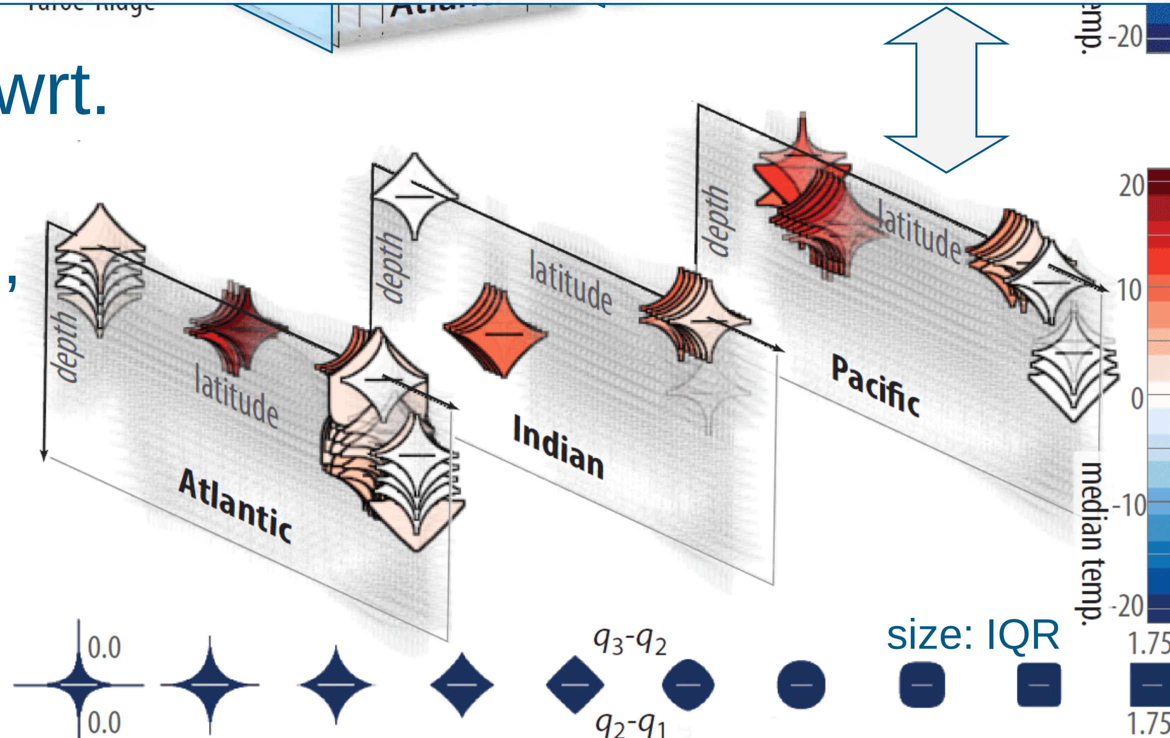
- Compute statistics wrt. the multiple runs
  - median, quartiles, etc.
  - billboard glyphs [Lie et al. 2009]

Showing Aggregated Statistics

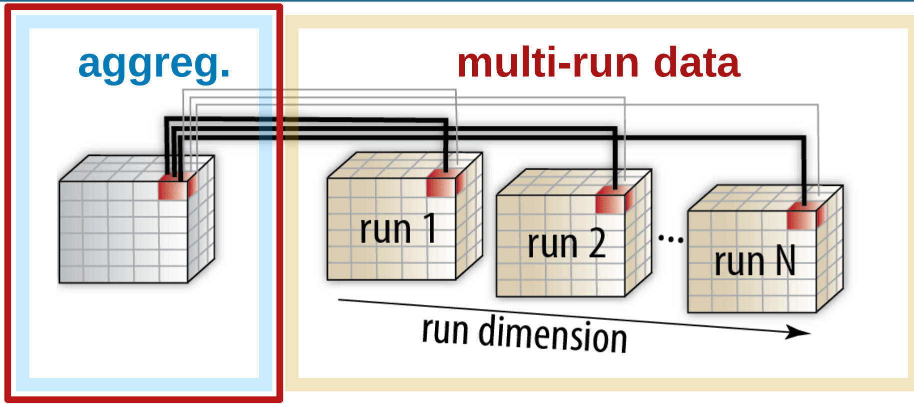
- Cooling event (8200 yrs ago)

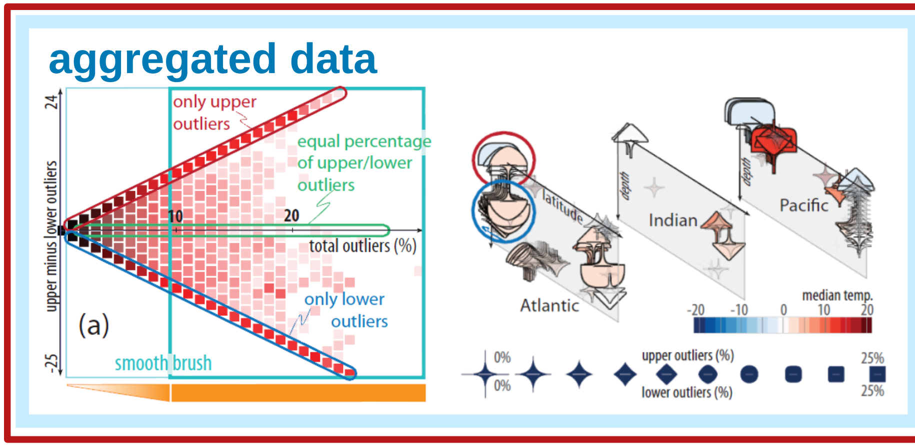


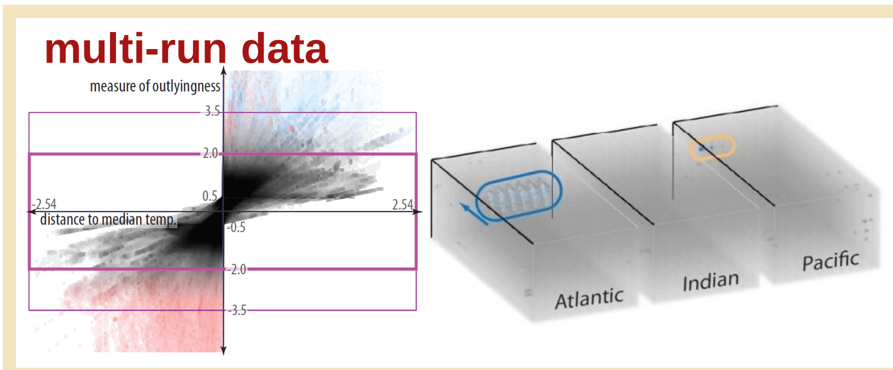
- Compute statistics wrt. the multiple runs
  - median, quartiles, etc.
  - billboard glyphs [Lie et al. 2009]



Case Study: Outlier Analysis



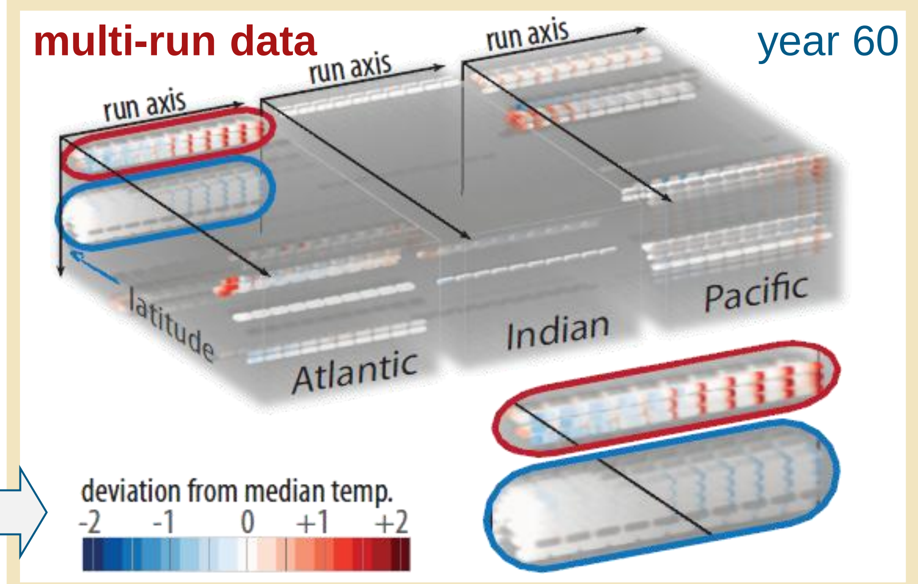


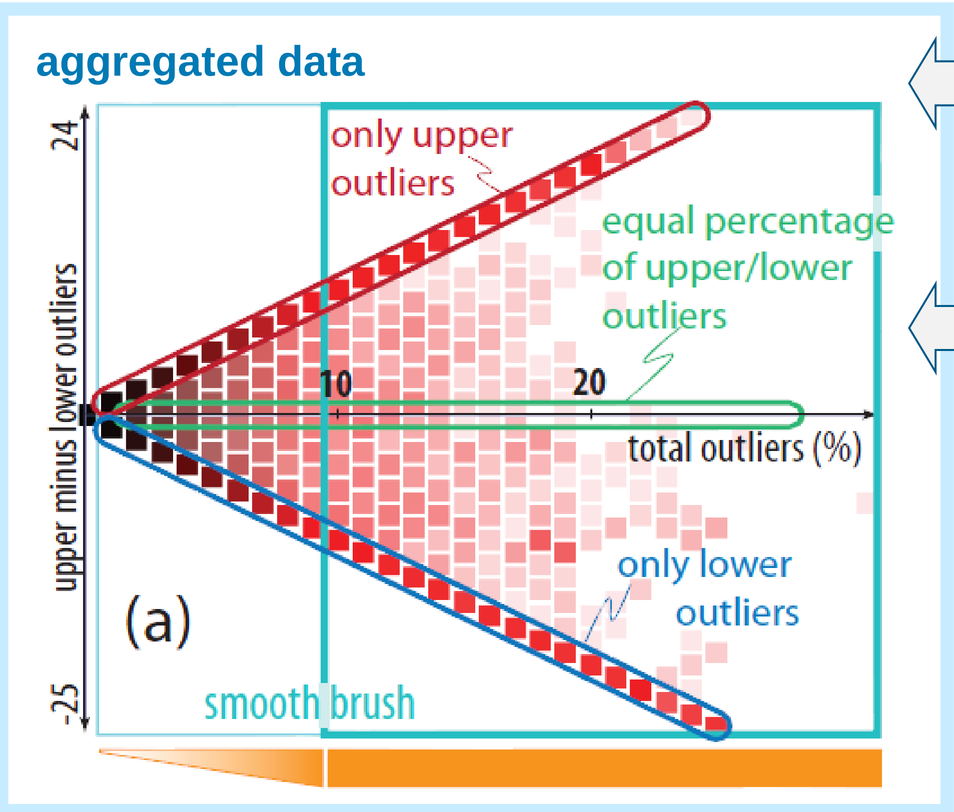


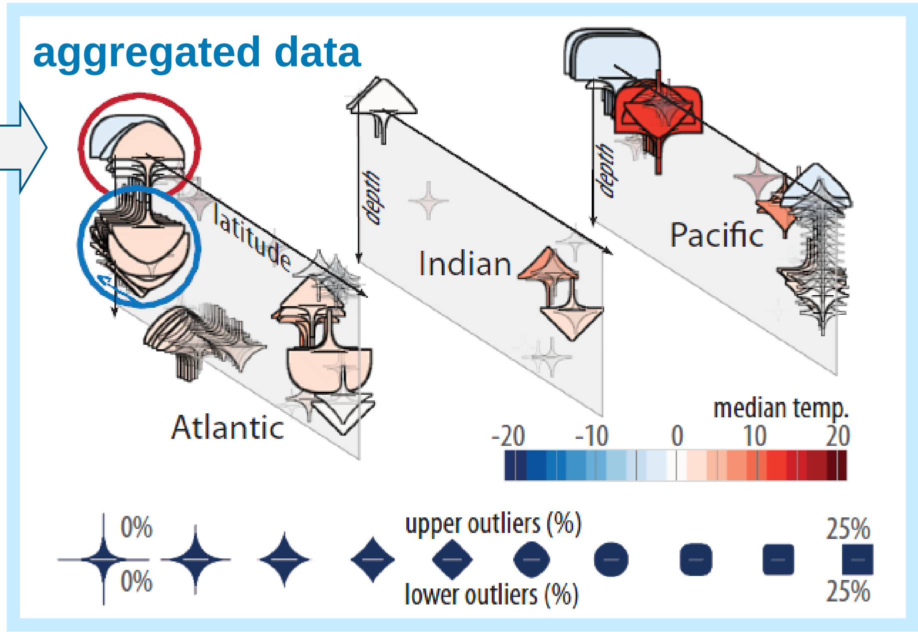
- Starts at aggregated level (summary statistics)
  - specify features via brushing
  - derive new attributes
- Refine features at multi-run level (details)
- Investigate further

Outlier Analysis: Aggregated Data

- Derive total number of outlier
- Where are outlier located?
  (% outlier > 3rd quartile + 1.5 IQR) –
  (% outlier < 1st quartile – 1.5 IQR)



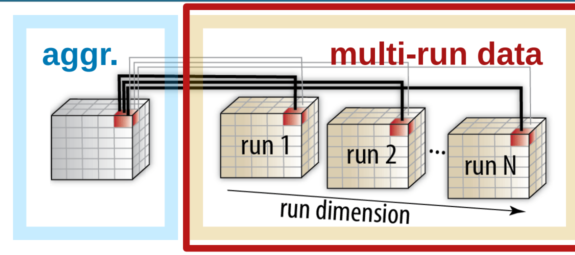


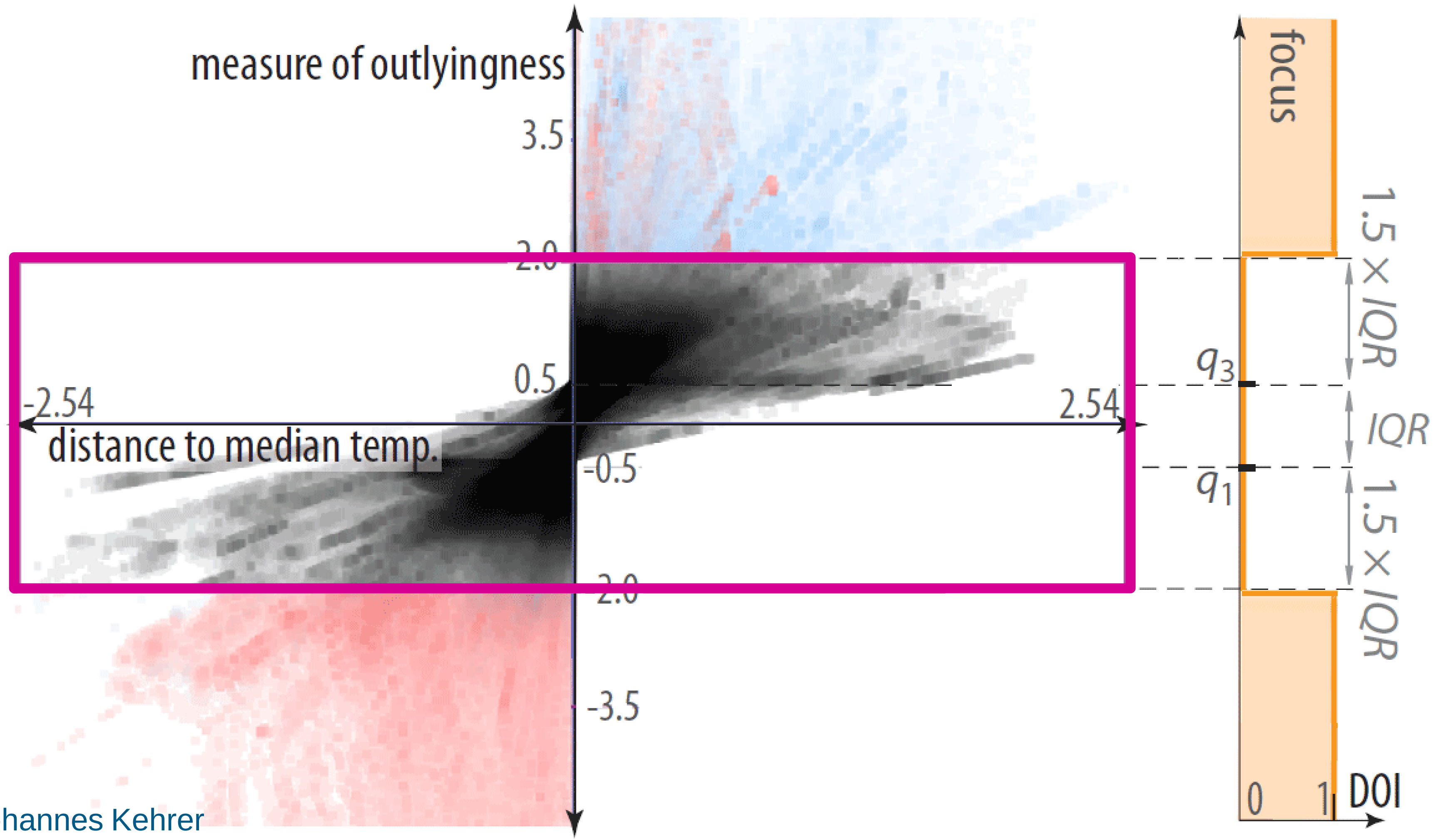


Feature Refinement: Multi-run Data

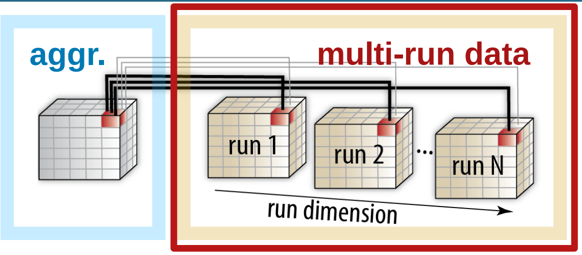
Derive measure of *outlyingness*

[Kehrer et al. 2010]

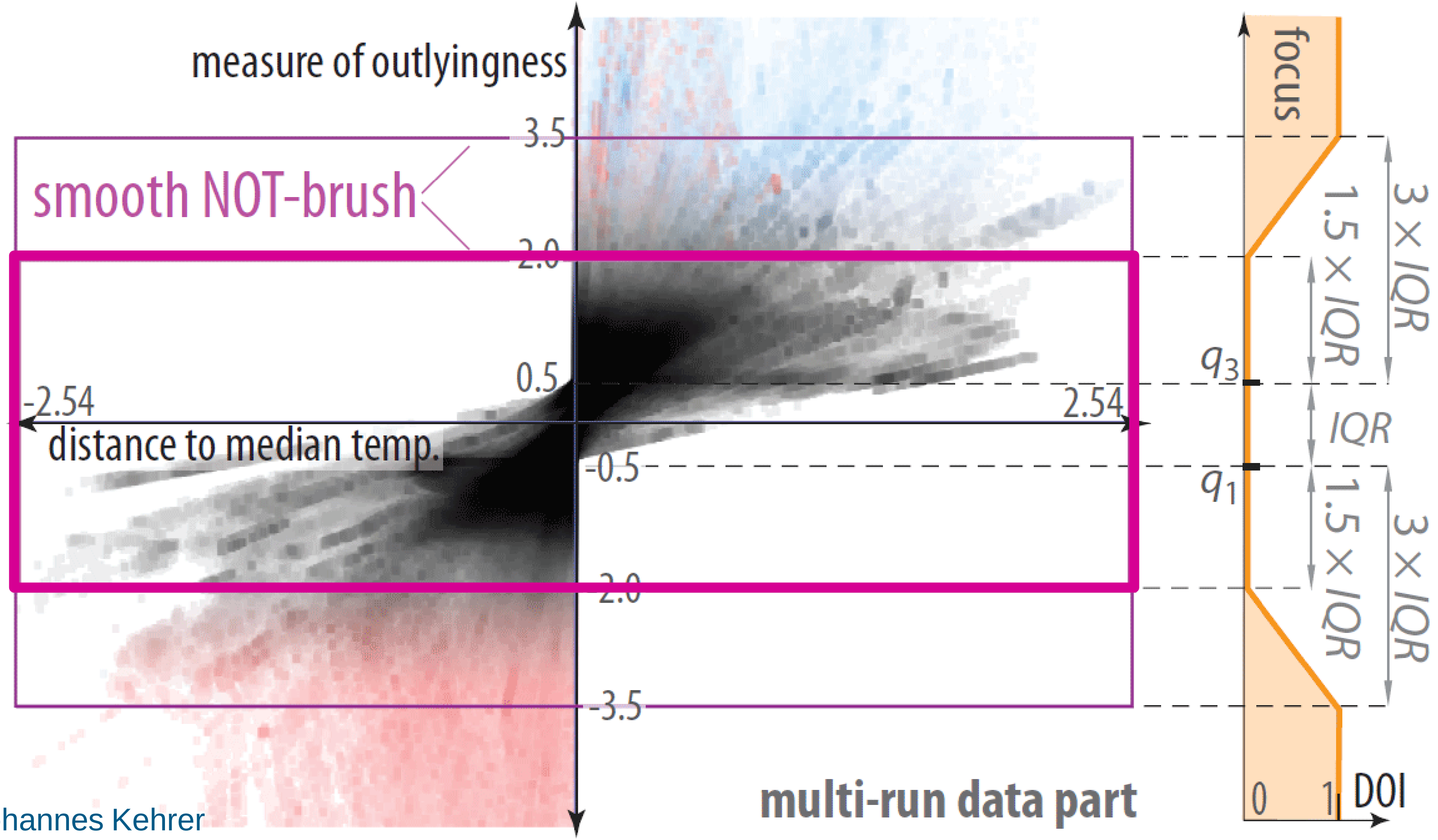




Feature Refinement: Multi-run Data

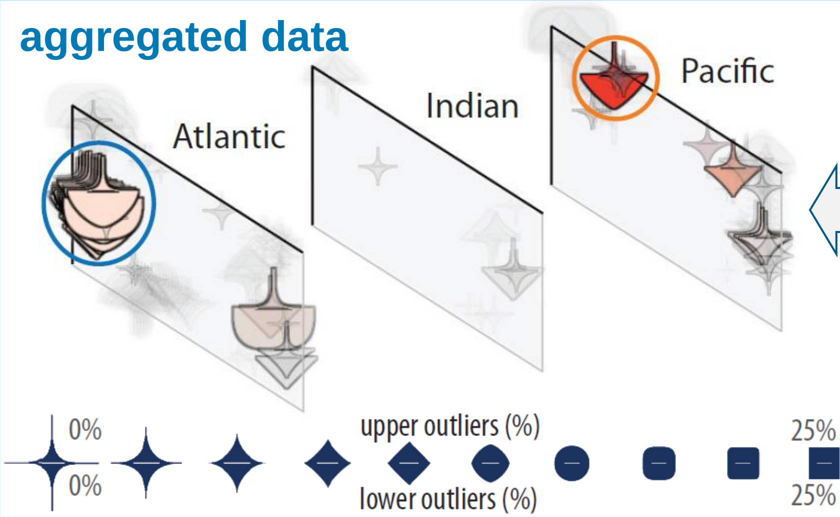


Derive measure of *outlyingness*
[Kehrer et al. 2010]

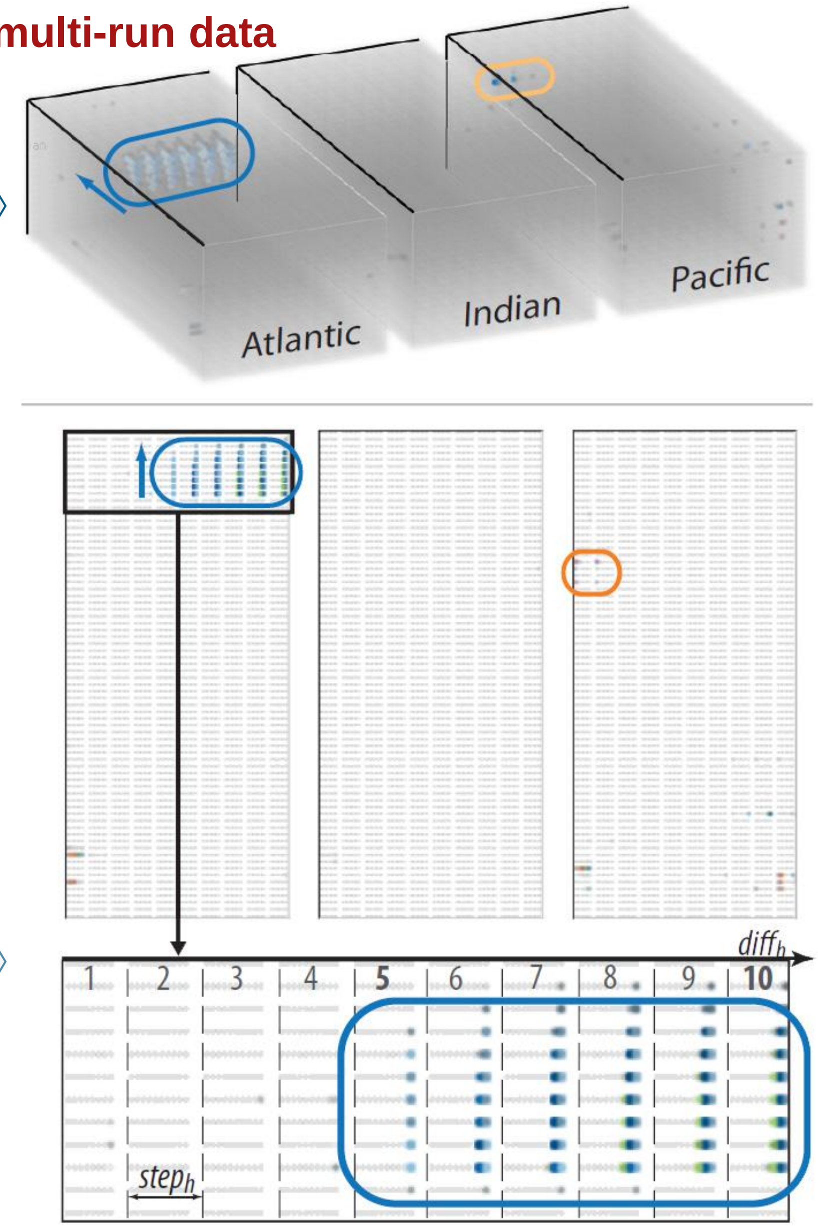


Sensitivity Analysis

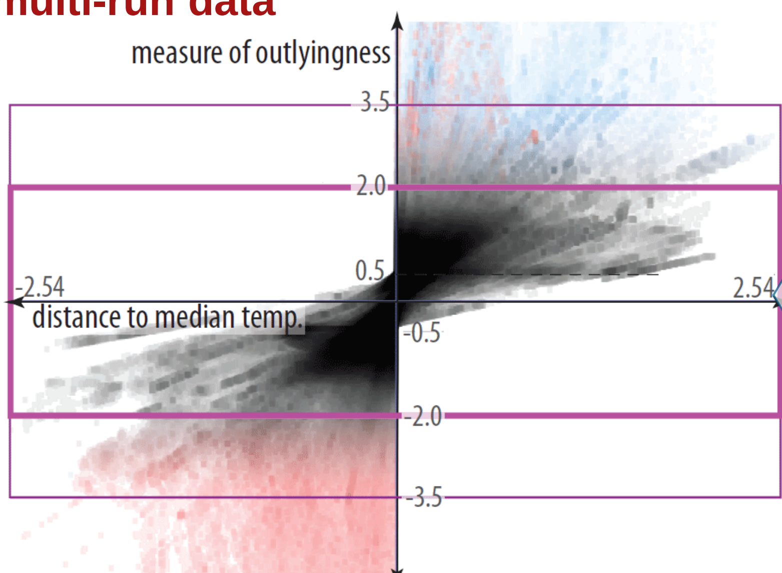
aggregated data



multi-run data

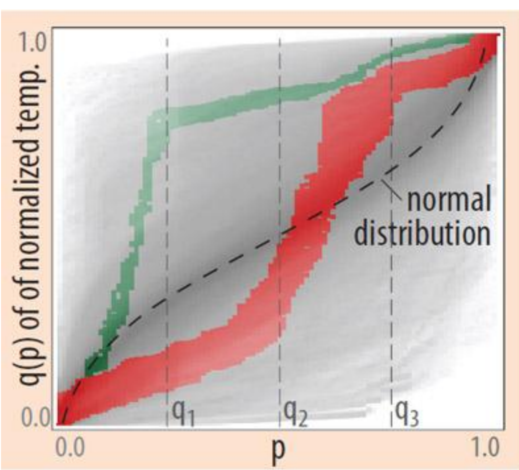


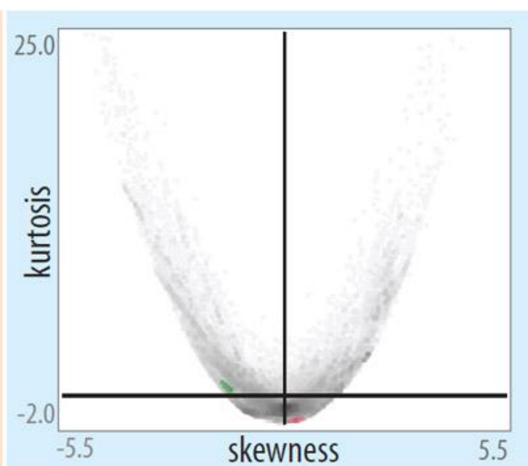
multi-run data

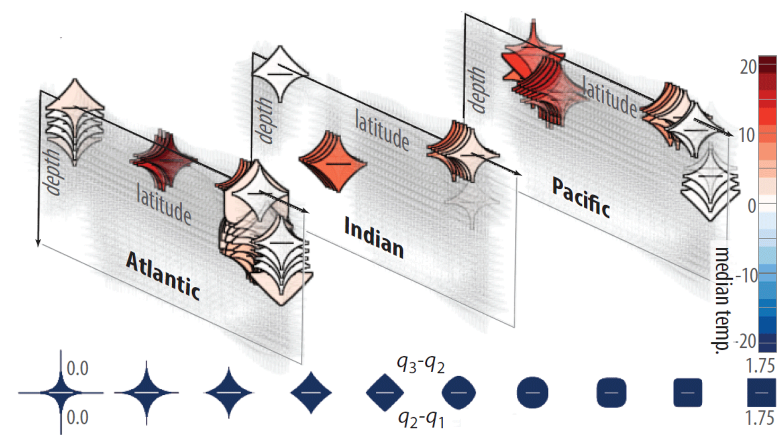


Statistics-based Visual Analysis

- J. Kehrer, P. Filzmoser, and H. Hauser. **Brushing moments in interactive visual analysis.** *CGF, 29(3):813–822, 2010.*
- J. Kehrer, P. Muigg, H. Doleisch, and H. Hauser. **Interactive visual analysis of heterogeneous scientific data across an interface**. *IEEE TVCG, 17(7):934–946, 2011.*







Conclusions

- Study **higher-dimensional scientific data** (multi-run)
- IVA across **2 data parts**
  - relating multi-run data ⇔ aggregated statistics
  - analyst can work with both parts (e.g., check validity)
- Integration of **statistical moments**
  - traditional vs. robust statistics, outliers
  - iterative view transformations
  - interactive statistical plots (linking & brushing)
- Workflow for **hypothesis generation**

Acknowledgements

- Helwig Hauser, VisGroup @ UiB
- Helmut Doleisch, Philipp Muigg, Wolfgang Freiler
- Florian Ladstädter, Andrea Steiner, Bettina Lackner, Barbara Pirscher, Gottfried Kirchengast
- Peter Filzmoser, Andreas Lie, Ove Daae Lampe
- Thomas Nocke, Michael Flechsig
- Armin Pobitzer, C. Turkay, Stian Eikeland
- many others