The Iterative Process of Interactive Visual Analysis

Helwig Hauser (Univ. of Bergen)

Thanks & context

- **Thanks** for the invitation to talk at EuroVA 2012! :–)
- "**Order**": to comment on VA ↔ SciVis, …
- **Context**:
  - ≈12 years of res. on interactive visual analysis, mostly at VRVis and at the Univ. of Bergen
  - PhD projects by Helmut Doleisch (–2004), Raphael Fuchs (–2008), Johannes Kehrer (–2011), Çağatay Turkay (2010–), and several others
  - res. cooperation with SimVis (H. Doleisch, *et al.*), VRVis (Krešimir Matković, Harald Piringer, *et al.*), Univ. of Magdeburg (Steffen Oeltze *et al.*), *etc*.
  - related projects, including VisMaster, SemSeg, *etc*.
  - funding from FFG (Austria), EC, UiB, *etc*.

Interactive Visual Analysis

- Given data –
  too much and/or complex to be shown at once,
- an **interactive visualization methodology**
  to facilitate
  - the **exploration** and **analysis** of data
    (not necessarily the presentation of data), including
    - **hypothesis generation** & **evaluation, sense making**,
    - **knowledge crystallization**, *etc.*
  - focusing according to the **user's interest**, *e.g.*,
    by interactive feature extraction,
  - navigating between **overview** and **details**, *e.g.*,
    to enable interactive information drill-down [Shneiderman]
- through an **iterative & interactive visual dialog**

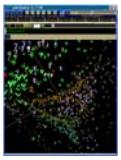
*reminds you of **visual analytics**?*

Visual Analytics ↔ Interactive Visual Analysis

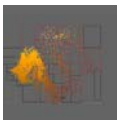
- **IVA** (interactive visual analysis) **since 2000**
- **Tightly related** to visual analytics, of course, *e.g.*,
  integrating computational & interactive data analysis
- **Particular methodology** with specific components
  (*CMV*, *linking & brushing*, *F+C vis.*, *etc.*)
- General enough to work in **many application fields**,
  but not primarily the VA fields (national security, *etc.*),
  in particular "SciVis fields"…
- Really a question of difference?? :–)

Integrating Interaction & Computation

- **Goal**: to combine the *best of two worlds* [Keim et al.]:
  - data **exploration**/analysis by the **user**, based on interactive visualization
  - and **data analysis** by the **computer**, based on statistics, machine learning, *etc.*
- State of the art / **levels of integration**:
  - **mostly no integration**, still
  - some **vis. of results** of computations
  - also: making **comp. semi-interactive** (here called "**inner integration**")
  - **rare: tight integration**
- **Outer integration** (here!): bundling **interaction & computation in a loop**



[Mariyar & Nabney, 2006]



[Williams & Munzner, 2004]

Target Model of "Scientific Data"

- **Characterized** by a combination of
  - **independent variables**, like **space** and/or **time** (aka. **domain**)
  - and **dependent variables**, like pressure, temp., etc. (aka. **range**)
- So we can think of this type of data as given as **d**(**x**) with **x** ↔ domain and **d** ↔ range – examples:
  - CT data $d(\mathbf{x})$ with $\mathbf{x} \in \mathbb{R}^3$ and $d \in \mathbb{R}$
  - time-dep. 2D flow $\mathbf{v}(\mathbf{x},t)$ with $\mathbf{x} \in \mathbb{R}^2$, $t \in \mathbb{R}$, and $\mathbf{v} \in \mathbb{R}^2$
  - num. sim. result $\mathbf{d}(\mathbf{x},t)$ with $\mathbf{x} \in \mathbb{R}^3$, $t \in \mathbb{R}$, and $\mathbf{d} \in \mathbb{R}^n$
  - system sim. $\mathbf{q}(\mathbf{p})$ with $\mathbf{p} \in \mathbb{R}^n$ and $\mathbf{q} \in \mathbb{R}^m$
- **Common property**:
  - **d** is (at least to a certain degree) **continuous** wrt. **x**

Interactive Visual Analysis **of Scientific Data**

- **Interactive visual analysis** (as exemplified in this talk) **works really well with scientific data**, *e.g.*,
  - results from **numerical simulation** (spatiotemporal)
  - imaging / **measurements** (in particular multivariate)
  - sampled **models**
- When used to study scientific data, **IVA employs**
  - methods from **scientific visualization** (vol. rend., …)
  - methods from **statistical graphics** (scatterplots, …), **information visualization** (parallel coords., *etc.*)
  - **computational tools** (statistics, machine learning, …)
- Applications include
  - **engineering**, **medicine**, **meteorology/climatology**, **biology**, *etc.*

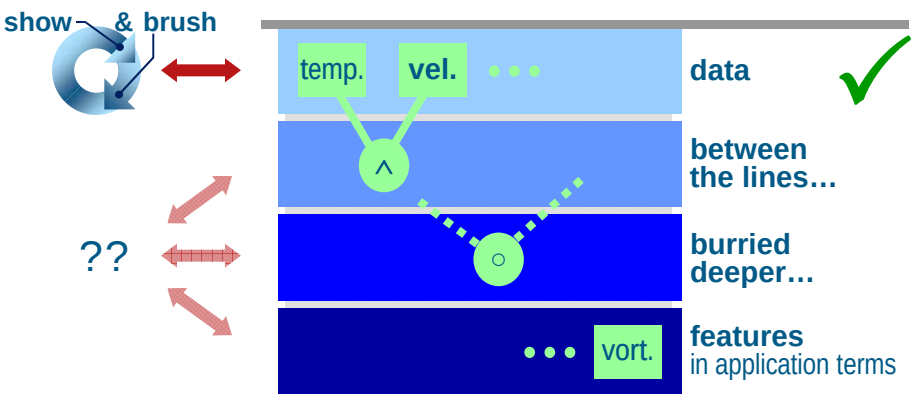
The Iterative Process of IVA

- Loop / bundling of *two complementary parts*:
  - **visualization** – show to the user!
    Something new, or something due to interaction.
  - **interaction** – tell the computer!
    What is interesting? What to show next?
- Basic example (**show – brush** – show – …), cooling jacket context:
  1. show a histogram of temperatures
  2. brush high temperatures (>90°[±2°])
  3. show focus+context vis. in 3D
  4. locate relevant feature(s)

- **KISS-principle IVA**:
  - linking & brushing, focus+context visualization, …

IVA – Levels of Complexity (1/4)

- A lot can be done with KISS-principle IVA! [pareto rule]
- We can consider a **layered information space**: from **explicitly** represented information (the **data**) to **implicitly** contained information, **features**, …

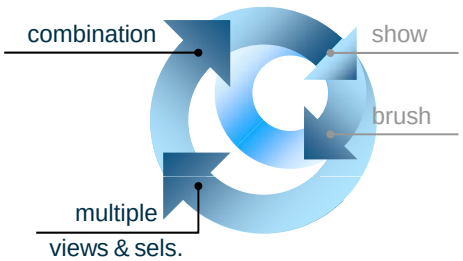


IVA – Levels of Complexity (2/4)

- A lot can be done with KISS-principle IVA! [pareto rule]
- **For more advanced** exploration/analysis **tasks**, we extend it (in seveal steps):
  - IVA, level 2: **logical combinations of brushes**, *e.g.*, utilizing the *feature definition language* [Doleisch et al., 2003]
  - IVA, l. 3: **attribute derivation**; **advanced brushing**, with interactive formula editor; *e.g.*, similarity brushing
  - IVA, l4: **application-specific feature extraction**, *e.g.*, based on vortex extraction methods for flow analysis
- Level 2: like **advanced verbal feature description**
  - ex.: "**hot** flow, also **slow**, near **boundary**" (cooling j.)
  - brushes comb. with **logical operators** (AND, OR, SUB)
  - in a **tree**, or **iteratively** $(((b_0\ op_1\ b_1)\ op_2\ b_2)\ op_3\ b_3)\ \ldots)$

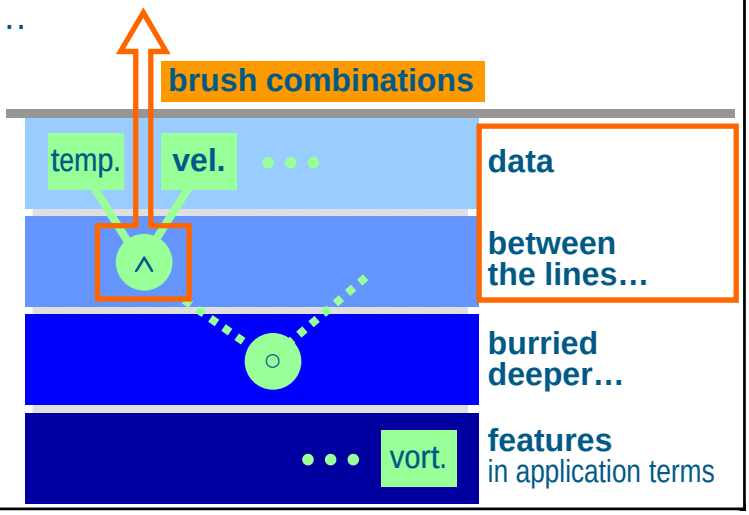
IVA – Levels of Complexity (2/4)

- A lot can be done with KISS-principle IVA! [pareto rule]
- **For more advanced** exploration/analysis **tasks**, we extend it (in seveal steps):
  - IVA, level 2: **logical**
    utilizing the *feature*
  - IVA, l. 3: **attribute d**
    with interactive form
  - IVA, l4: **application-**
    based on vortex extr
- Level 2: like **advanced**
  - ex.: "**hot** flow, also **s**
  - brushes comb. with
  - in a **tree**, or **iteratively** $(((b_0\ op_1\ b_1)\ op_2\ b_2)\ op_3\ b_3)\ \ldots)$

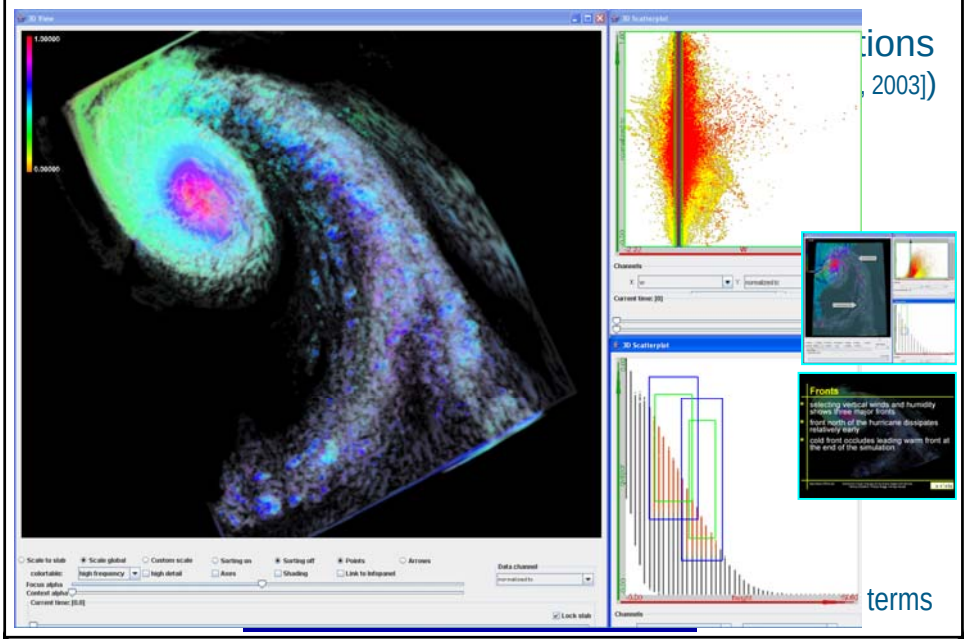


IVA (level 2) **Synopsis**

- Multiple views, multiple brushes, brush combinations via logical ops. (feature definition language [Doleisch et al., 2003])
- Example…



IVA (level 2) **Example**



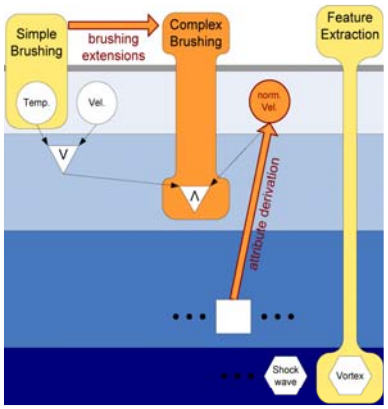
…tions
…, 2003])
…terms

IVA – Levels of Complexity (3/4)

- A lot can be done with KISS-principle IVA! [pareto rule]
- **For more advanced** exploration/analysis **tasks**, we extend it (in seveal steps):
  - IVA, level 2: **logical combinations of brushes** [e.g. utilizing the *feature definition language* [Doleisch et al. 2003]]
  - IVA, l. 3: **attribute derivation**; **advanced brushing**, with interactive formula editor; *e.g.*, similarity brushing
  - IVA, l4: **application-specific feature extraction**, *e.g.*, based on vortex extraction methods for flow analysis
- Level 3: using **general info extraction** mechanisms, two (partially complementary) approaches:
  1. **derive additional attribute**(s), then show & brush
  2. use an **advanced brush** to select "hidden" relations

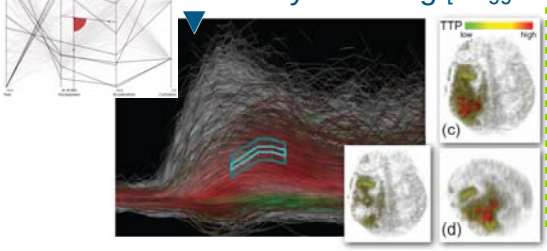
IVA – Levels of Complexity (3/4)

- A lot can be done with KISS-p
- **For more advanced** explorat
  we extend it (in seveal steps):
  - IVA, level 2: **logical combin**
    utilizing the *feature definitior*
  - IVA, l. 3: **attribute derivatio**
    with interactive formula editc
  - IVA, l4: **application-specifi**
    based on vortex extraction n
- Level 3: using **general info extraction** mechanisms, two (partially complementary) approaches:
  1. **derive additional attribute**(s), then show & brush
  2. use an **advanced brush** to select "hidden" relations

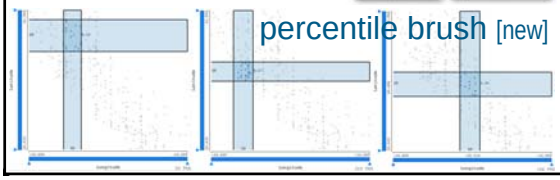


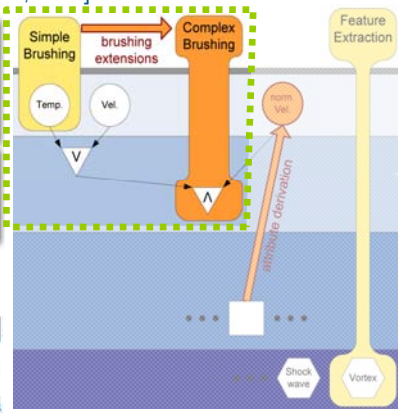
IVA (level 3): **Advanced Brushing**

- **Std. brush**: brush 1:1 what you see
  Adv. brush: executes additional function ("intelligent"?)
- **Examples**:
  - angular brushing [Hauser et al., 2002]
    similarity brushing [Muigg et al., 2008]



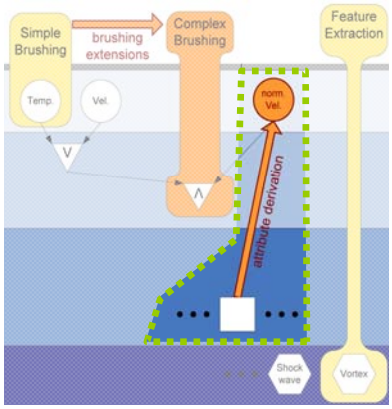
percentile brush [new]





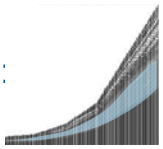
IVA (level 3): **Attribute Derivation**

- **Principle** (in the context of iterative IVA):
  - **see** some data feature Φ of interest in a visualization
  - **identify** a **mechanism** T **to describe** Φ
  - **execute** (interactively!) an **attribute derivation** step to represent Φ explicitly(as new, synthetic attribute[s] d_φ)
  - **brush** d_φ to get Φ
- **Tools** T to describe Φ from:
  - **numerical mathematics**
  - **statistics**, **data mining**
  - *etc*.
  - ➢ **scientific computing**
- **IVA** w/ **T** ↔ **visual computing**



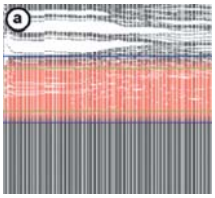
Attribute Derivation ↔ **User Task** / example

- The tools T, available in an IVA system, must reflect/match the **analytical steps of the user**:
- **Example**:
  - **first vis.**:

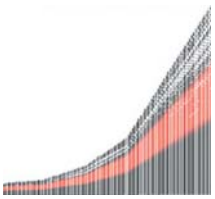


↔ user wishes to select the "band" in the middle

  - **so?** :–) an advanced brush? a lasso maybe?
  - **ah!** → let's normalize *y* and then brush (a)



  - **leading to the wished selection**:



What user wishes to reflect?

- Many **generic wishes** – users interest in:
  - something **relative** (instead of some absolute values), example: show me the *top-15%*
  - **change** (instead of current values), ex.: show me *regions with increasing temperature*
  - some **non-local property**, ex.: show me regions with *high average temperature*
  - **statistical properties**, ex.: show me *outliers*
  - **ratios/differences**, ex.: show me population per area, difference from trend
  - *etc*.
- **Common characteristic** here:
  - **questions/tools generic**, not application-dependent!

How to reflect these user wishes?

- Many **generic wishes** – users interest in:
  - something **relative** (instead of some absolute values), example: show me the *top-15%* ⇒ ***use,*** *e.g.* ***, normalization***
  - **change** (instead of current values), ex.: show me *regions with increasing temperature* ⇒ ***derivative estimation***
  - some **non-local property**, ex.: show me regions with *high average temperature* ⇒ ***numerical integration***
  - **statistical properties**, ex.: show me *outliers* ⇒ ***descriptive statistics***
  - **ratios/differences**, ex.: show me population per area, difference from trend ⇒ ***calculus***
  - *etc*. ⇒ ***data mining*** (fast enough?)
- **Common characteristic** here:
  - **questions/tools generic**, not application-dependent!

Some useful tools for 3rd-level IVA

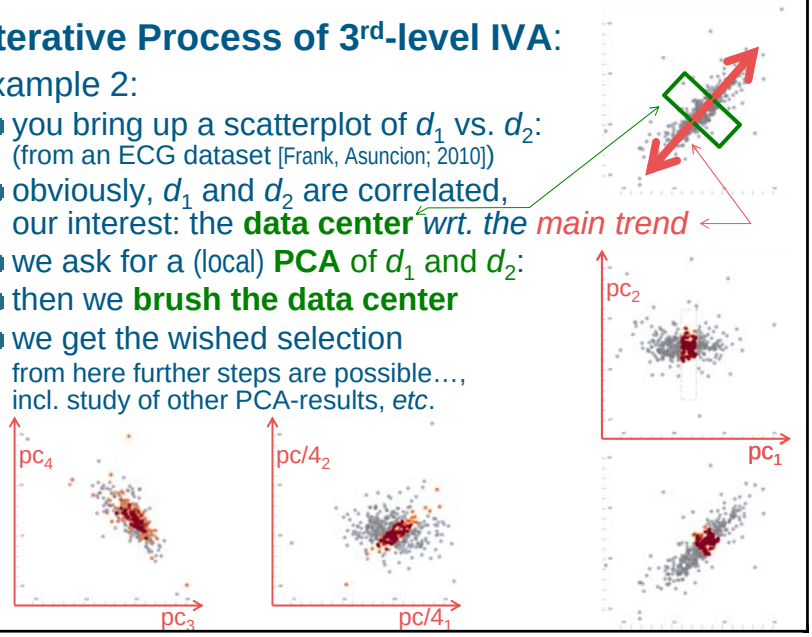
- From **analysis**, **calculus**, **num. math**:
  - **linear filtering** (convolve the data with some linear filter on demand, e.g., to smooth, for derivative estimation, *etc.*)
  - **calculus** (use an interactive formula editor for computing simple relations between data attributes; +, −, ·, ⁄, *etc.*)
  - **gradient estimation**, **numerical integration** (*e.g.*, wrt. space and/or time) ⇒ *example*
  - **fitting/resampling** via **interpolation/approximation**
- From **statistics**, **data mining**: ⇒ *example*
  - **descriptive statistics** (compute the statistical moments, also robust, measures of outlyingness, detrending, *etc.*)
  - **embedding** (project into a lower-dim. space, *e.g.*, with PCA for a subset of the attribs., *etc.*) ⇒ *example* ⇒ *example*
- **Important**: **executed on demand**, after prev. vis.

3rd-level IVA – Sample Iterations (1/2)

- The **Iterative Process of 3rd-level IVA**:
  - Example 1:
    - you look at some *temp. distribution over some region*
    - you are *interested raising temperatures*, ***but not temperature fluctuations***
    - you use a **temporal derivate estimator**, for ex., central differences $t_{change} = (t_{future} - t_{past})/\text{len}(future - past)$
    - you plot t_{change}, e.g., in a **histogram** and **brush** what ever change you are interested in
    - maybe you see that some frequency amplification due to derivation, so you go back and
    - **use an appropriate smoothing filter** to *remove high frequencies from the temp. data*, leading to a derived, new $\tau = t_{smooth}$ data attribute
    - selecting from a **histogram** of τ_{change} (computed like above) is then less sensitive to temperature fluctuations

3rd-level IVA – Sample Iterations (2/2)

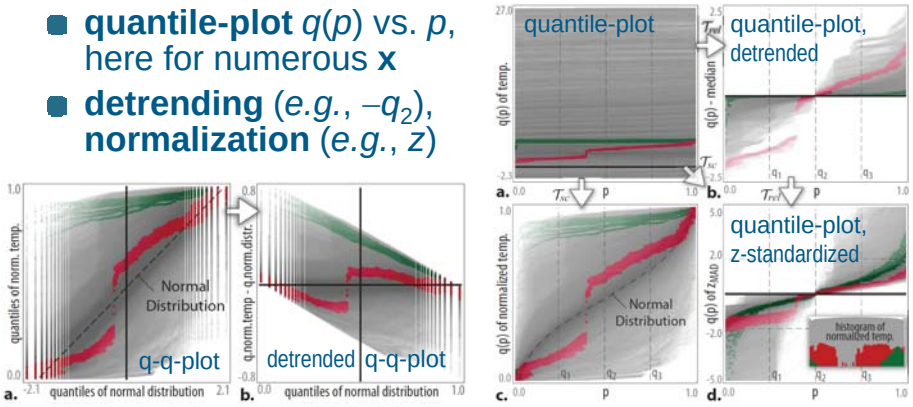
- The **Iterative Process of 3rd-level IVA**:
  - Example 2:
    - you bring up a scatterplot of d_1 vs. d_2: (from an ECG dataset [Frank, Asuncion; 2010])
    - obviously, d_1 and d_2 are correlated, our interest: the **data center** *wrt. the main trend*
    - we ask for a (local) **PCA** of d_1 and d_2:
    - then we **brush the data center**
    - we get the wished selection
    - from here further steps are possible…, incl. study of other PCA-results, *etc*.



Visualizing / analyzing lots of statistics

[Kehrer et al., TVCG 2011]

- Useful statistical measures include:
  - **moments** (μ, σ, …), **robust versions** (median, IQR, …)
  - **quartiles**, **octiles**, and **quartiles** $q(p)$
- Useful views allow the interactive visual analysis
  - **quantile-plot** $q(p)$ vs. p, here for numerous **x**
  - **detrending** (*e.g.*, $-q_2$), **normalization** (*e.g.*, z)



[IEEE Vis, 2008]

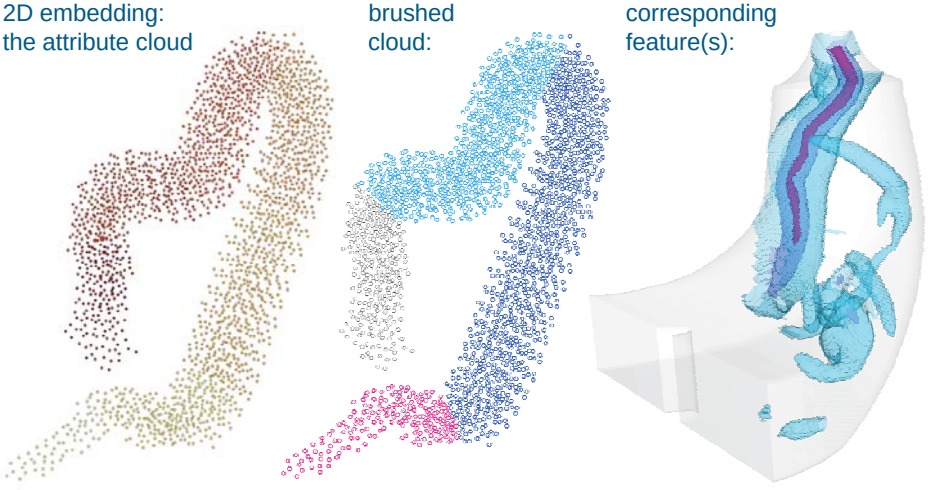
Brushing of Attribute Clouds for the Visualization of Multivariate Data

Heike Jänicke, Michael Böttinger, and Gerik Scheuermann, *Member, IEEE*

2D embedding: the attribute cloud

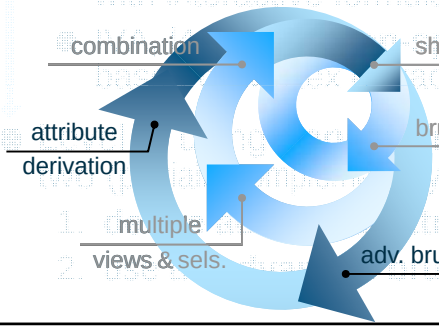
brushed cloud:

corresponding feature(s):



IVA – Levels of Complexity (4/4)

- A lot can be done with KISS-principle IVA! [pareto rule]
- **For more advanced** exploration/analysis **tasks**, we extend it (in seveal steps):
  - IVA, level 2: **logical combinations of brushes** e.g. utilizing the *feature definition language* [Doleisch et al. 2003]
  - IVA, l. 3: **attribute derivation**; **advanced brushing**, with interactive formula editor; *e.g.*, similarity brushing
    - **feature extraction**, *e.g.*, ethods for flow analysis
    - **traction** mechanisms, oaches:
      - s), then show & brush
      - select “hidden” relations

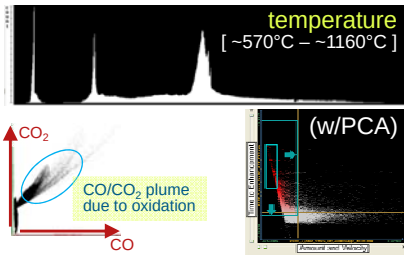


IVA – Levels of Complexity (2/n)

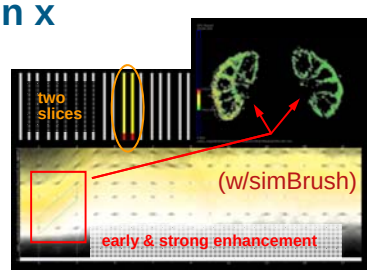
- A lot can be done with KISS-principle IVA! [pareto rule]
- **For more advanced** exploration/analysis **tasks**, we extend it (in seveal steps):
  - IVA, level 2: **logical combinations of brushes** e.g. utilizing the *feature definition language* [Doleisch et al. 2003]
  - IVA, l. 3: **attribute derivation**; **advanced brushing**, with interactive formula editor; *e.g.*, similarity brushing
  - IVA, l4: **application-specific feature extraction**, *e.g.*, based on vortex extraction methods for flow analysis
- Level 4: **application-specific procedures**
  - tailored solutions (for a specific problem)
  - “deep” information drill-down
  - *etc*.

Interactive Visual Analysis – delivery

- Understanding data **wrt. range d**
  - which distribution has data attribute $\mathbf{d}_i$?
  - how do $\mathbf{d}_i$ and $\mathbf{d}_j$ relate to each other? (**multivariate analysis**)
  - which $\mathbf{d}_k$ discriminate data features?



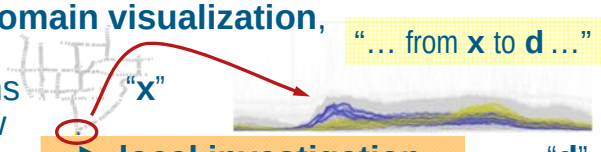
- Understanding data **wrt. domain x**
  - **where** are relevant features? (**feature localization**)
  - **which** values at specific **x**? (**local analysis**)
  - how are they **related to parameters**?



Three Patterns of SciData IVA

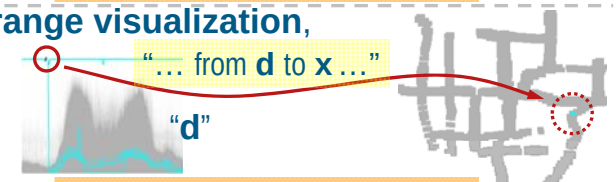
- Preliminary: domain **x** & range **d** visualized (≥2 views)

❶
- **brushing on domain visualization**, e.g., brushing special locations in the map view



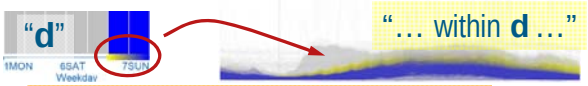
▶ **local investigation** "d"

❷
- **brushing on range visualization**, e.g., brushing outlier curves in a function graph view



▶ **feature localization** "x"

❸
- relating multiple range variates



▶ **multi-variate analysis** "d"

The Iterative Process of IVA…

…leads to an **interactive** & **iterative** workbench for **visual data exploration & analysis** (compare to **visual computing**, again)

- Different **levels of complexity** (show & brush, logical combinations, advanced brushing & attribute derivation, *etc.*)…

…lead to according **iteration frequencies**:

- on level 1: **smooth interactions, many fps,** for example during linking & brushing
- on level 2: **interleaved fast steps of brush ops.,** for example when choosing a logical op. to cont. with
- on level 3: **occasionally looking at a progress bar,** for example when computing some PCA, *etc*.

- These frequencies **limit the spectrum** of usable tools

➢ New res. work will help to **extend this spectrum**!

The Iterative Process of IVA…

…is a **very useful methodology** for **data exploration & analysis**

…is **very general** and can be (has already been) applied to **many different application fields** (in this talk the focus was on scientific data)

…**meets scientific computing** as a complementary methodology (with the **important difference** that in IVA the **user** with his/her **perception/cognition** is **in the loop** at **different frequencies**, also many fps)

…is **not yet fully implemented** (we've done something, e.g., in the context of **SimVis**, **ComVis**, *etc*.) – from here: different possible paths, incl. InteractiveVisualMatlab, IVR, *etc*.)

Acknowledgements

- **You!**
- **Krešimir Matković** & **Giuseppe Santucci**!
- Helmut Doleisch, Raphael Fuchs, Johannes Kehrer, Çağatay Turkay, *et al*.!
- Collaboration partners (St. Oeltze, Fl. Ladstädter, G. Weber, *et al*.)
- All around SimVis and ComVis and …
- Funding partners (FFG, AVL, EU, UiB, …)

Vis/IVA PhD in Bergen?
Apply until 10.6. or 10.8.!

... see www.ii.UiB.no/vis !!