Visual Analytics in Visplore

Harald Piringer
VRVis Research Center
Vienna, Austria
Part 1: Visplore
Background

- Software for exploratory data analysis developed at VRVis
- The beginning: analysis of multi-run simulations
  - Application context: Development of car engines
  - Collaborative project with AVL List GmbH
  - Distributed as part of their commercial software suite
- Now: general-purpose analysis for numerous applications
  - Extensions for time-dependent data, categorical data, etc.
  - Applied to data from finance, telecommunication, business, facility management, pharmaceutics, infrastructure, etc.
Supported Tasks

- **Exploratory analysis**
  - Get a feeling and understanding of large and complex data
  - Find patterns of any kind (e.g., clusters, outliers, trends, ...)
  - Find relationships

- **Black-box analysis**
  - Try „what-if“ scenarios concluding from input to output and vice versa without delay

- **Optimization**
  - Analyze trade-offs for multi-criteria objectives

- **Interactive statistics**
  - Statistically describe local features
Technicals

- written in C++
- OpenGL used for rendering
- gtk+ used for GUI
- designed for large data
  - highly multi-threaded
  - quick preview during interaction
  - memory management
- open architecture
  - views / importers / etc. are plugins
Interested...?

- Internship / diploma thesis
  - Contribute to visplore!
  - **Interesting topics** oriented towards practice
  - Large-scale real-world project
  - Help and supervision
  - Payment
  - Ask me or write us an email: kehrer@vrvis.at, hp@vrvis.at

- Questions?
Part 2: Application Scenarios
Application Background

- Development of powertrain systems
  - Computational Fluid Dynamics (CFD) simulations (1D)
    - based on physical equations
    - slow
  - Surrogate models
    - based on statistical regression
    - estimated results in real-time
- Identification of regression models
  - complex task
    - selection of training and validation data
    - selection of attributes as inputs
    - numerous training parameters
  - critical: involving engineers in identification process
Validation of Regression Models

- Important part of the model identification process

- Three levels of detail
  - L1: Global prediction quality
    - statistical summaries (e.g., maximal residual)
  - L2: Local prediction quality
    - derived attributes (e.g., residual for each validation point)
    - Identification without explanation
  - L3: Model in context of validation data
    - Most information (e.g., gradient)
    - Does not trivially scale to higher dimensionality
Validation of Regression Models

- **Combined visualization of regression model and validation data**
  - Model: scalar function \( y = f(x_1, x_2, ..., x_n) \)
  - Validation data: set of \( n+1 \)-tuples

- **Layout: projection to low-dimensional space**
  - Paraxial slices around focal point
  - Matrix of plots
    - Parameter plots
    - Surface plots
    - Parity plot

---

Harald Piringer
Visual Analytics in Visplore
Validation of Regression Models

- Different kinds of information
  - Fit by a model
  - Coverage by validation data
- → Relevant range around slice of each plot
- Changing relevant range
Sensitivity Analysis

- Sensitivity to changes of model parameters
- Variation graphs along range of active dimension
  - Grid lines as meaningful steps of variation
  - Coupled to color
  - Family of graphs
Thank you for your attention!

Questions welcome!
hp@vrvis.at