

Explorative Analysis and Visualization of Multi-Dimensional and Geo Related Fuel Survey Data with KNIME

Dr. Alexander Warta, Robert Bosch GmbH



Diesel Systems

DS/ETM1-Warta | 3/2/2011 | 102-1883 | © Robert Bosch GmbH 2011. All rights reserved, also regarding any disposal, exploitation, reproduction, editing, distribution, as well as in the event of applications for industrial property rights.



BOSCH

Outline

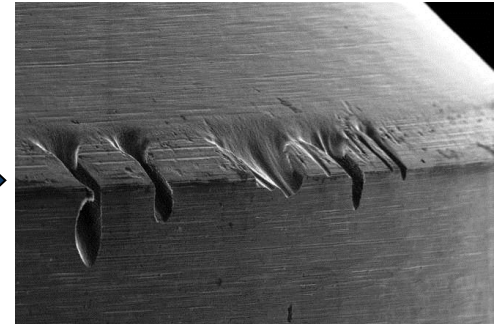
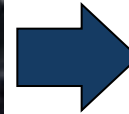
- Context and challenge
 - Evaluation of global fuel surveys

- Solution
 - Designing an open source tool chain for interactive analysis of high dimensional fuel data

- Use cases
 - Interactive boxplotting by integrating R
 - Generating fuel world maps
 - Enhancing fuel world maps with additional data
 - Starting fuel survey data mining

- Summary and next steps

Context: effects of poor fuel quality on the FIE* ...



example:

particles resulting from fuel logistics cause injector wear

* fuel injection equipment

Context and challenge

- In order to design diesel fuel injection systems for global markets Robert Bosch GmbH considers a lot of specific diesel fuel quality parameters of various markets
- For this, fuel samples from almost all countries are chemically analyzed by a service provider regularly – so-called “fuel surveys”
- One survey sample record contains up to 140 attributes, e.g. date, town, country, supplier and the results of chemical and physical analysis like sulfur content, density, viscosity, biodiesel content etc.
- About 10.000 records are currently of relevance
- The previous process integrated Microsoft Excel (plots, histograms, etc.) and PowerPoint (world map) in a non-automated succession
- This procedure is quite time consuming, not interactive, inflexible and not scalable

Search for a solution: catalog of requirements

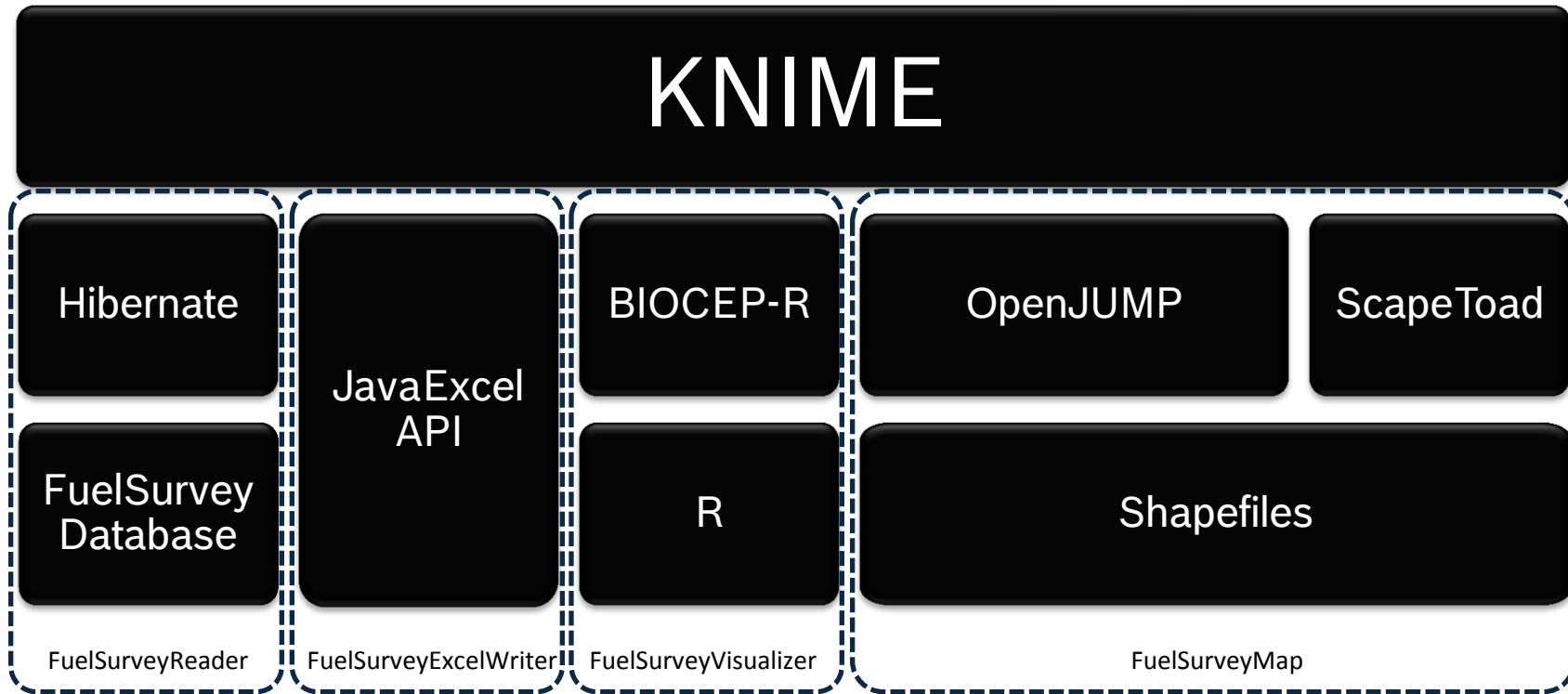
- extract knowledge through interactive exploration
- easy access to all fuel surveys with filter methods
- generate choropleth maps and cartograms
 - show country names
 - show additional diagrams for each country
 - show only selected countries
 - enrich map with external data (like cities of the fuel survey records, locations of oil refineries, etc.)
- generate star plots, parallel coordinates, scatterplots
- apply data mining algorithms for finding new patterns between instances and features (like association rule learning, hierarchical clustering, multidimensional scaling)
- enrich fuel survey data with external data (like new diesel car registrations, failure count of the common rail system, etc.)

Why did we choose KNIME?

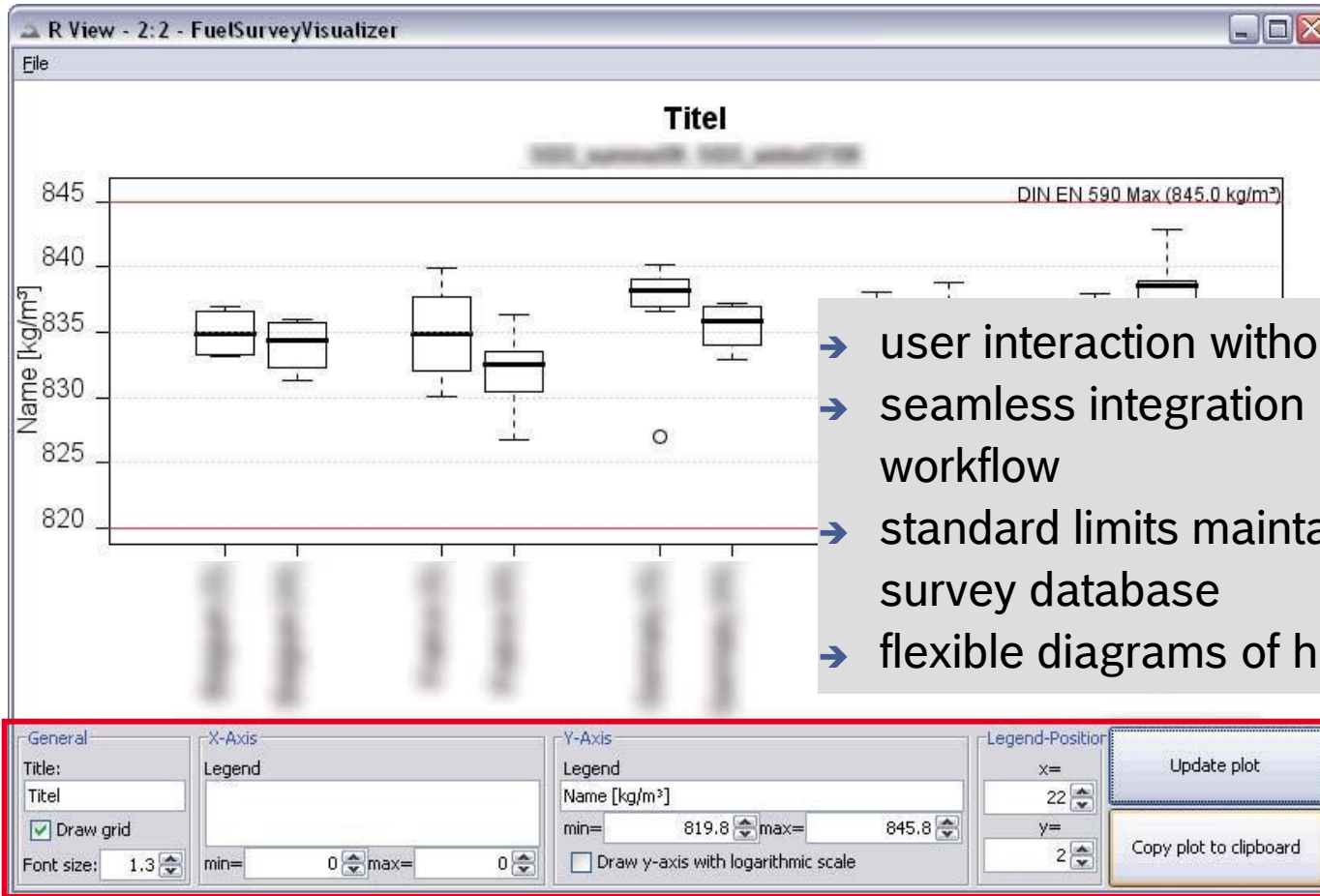
- In development since 2006
- Forum provided especially for developers
- KNIME.com GmbH provides additional support and tutorials. This is also an evidence for further active development
- Easy to use due to the intuitive node concept
- Good documentation. The KNIME SDK, Javadoc of the API and the source code of existing nodes make it easy to develop new nodes
- Based on Eclipse which is already approved by the internal Robert Bosch software clearing process. Additional support through Eclipse community possible
- Provides various nodes by default. A lot of functionality is already available and doesn't have to be implemented
- Numerous preprocessing nodes for aggregation, transformation and nodes for visualizing the data (scatter plot, parallel coordinates, etc.)



Development of different KNIME nodes

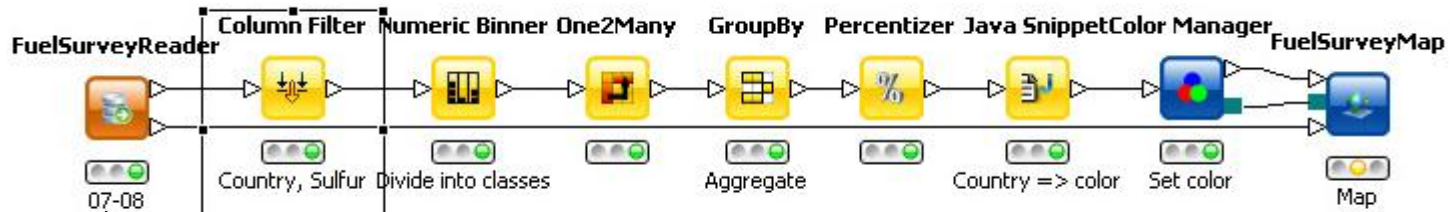


Use case: interactive boxplotting using R

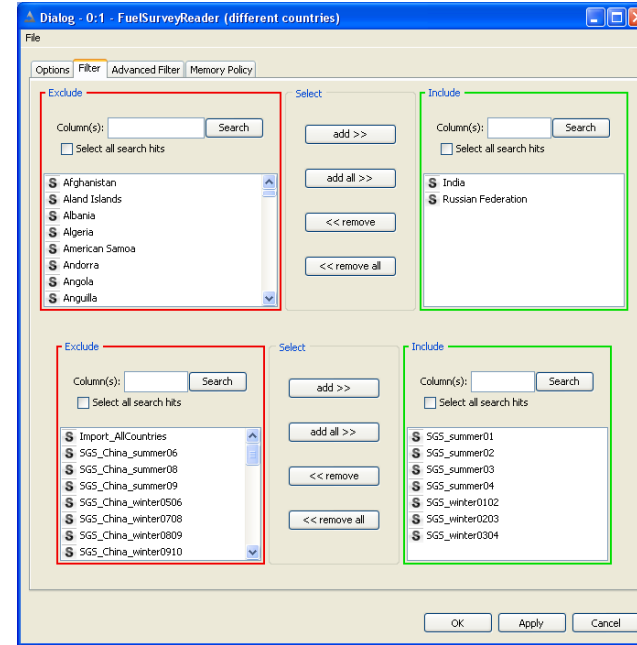
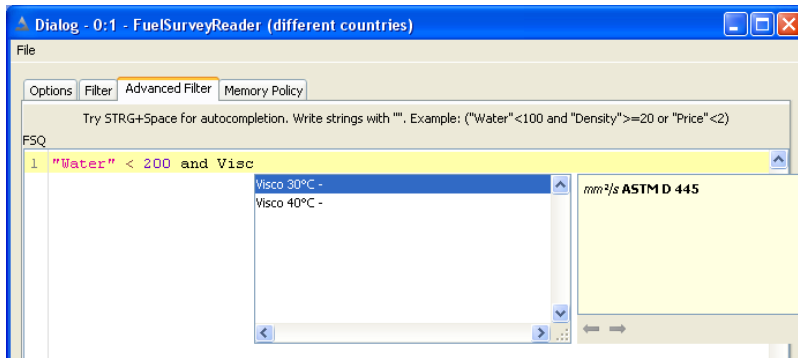


- user interaction without R know-how
- seamless integration in KNIME workflow
- standard limits maintained in the fuel survey database
- flexible diagrams of high quality

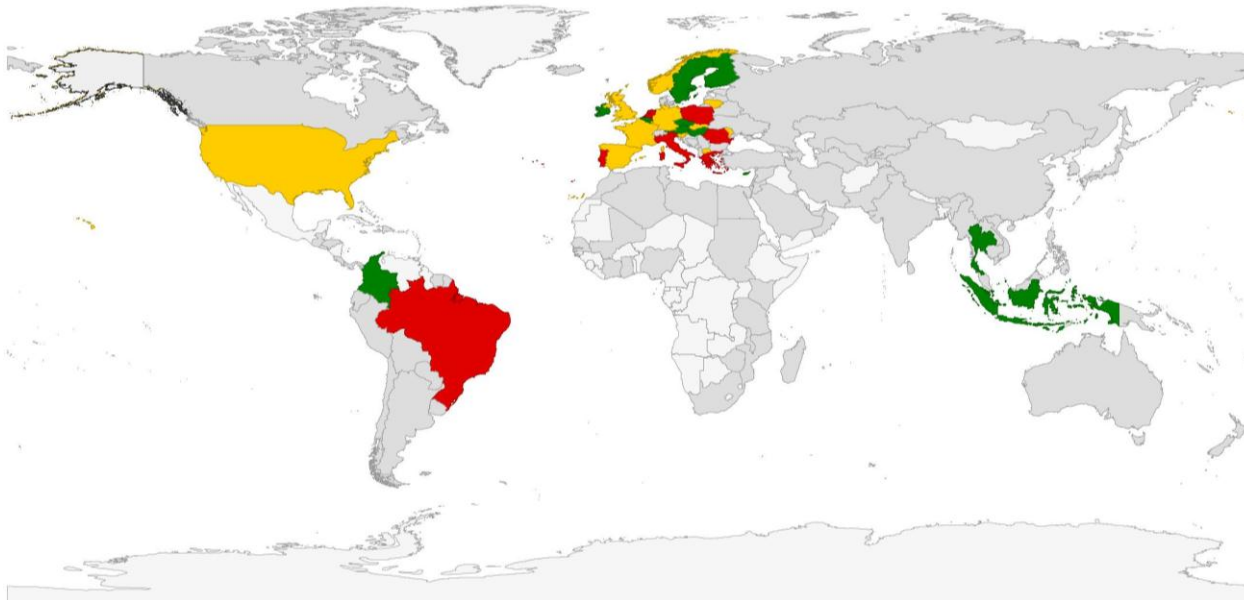
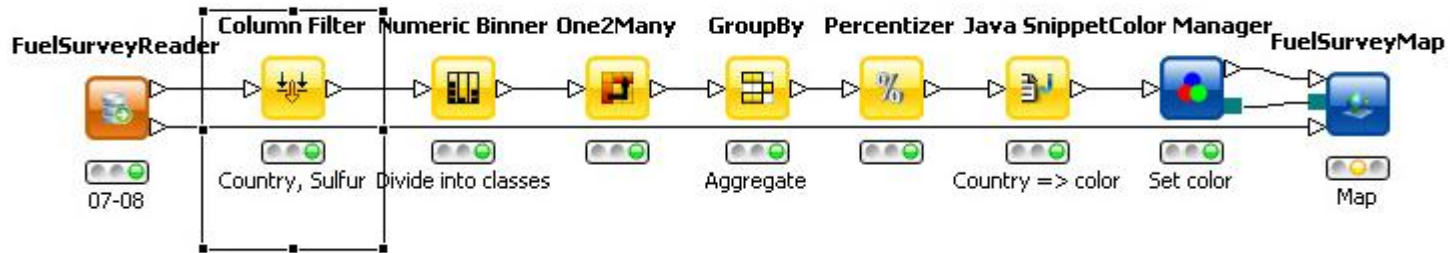
Use case: generating fuel world maps



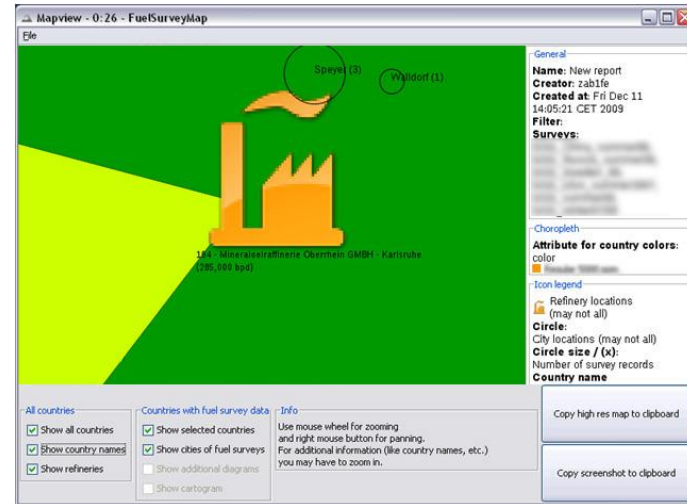
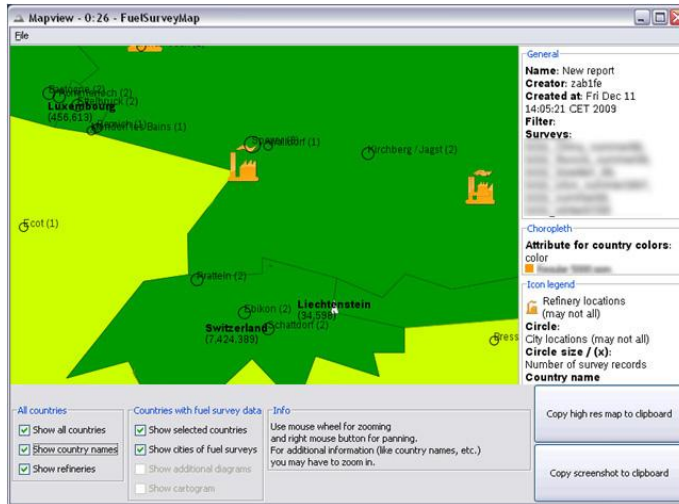
- DB connector
- survey prefiltering (survey, country, attribute values)



Use case: generating fuel world maps – the result

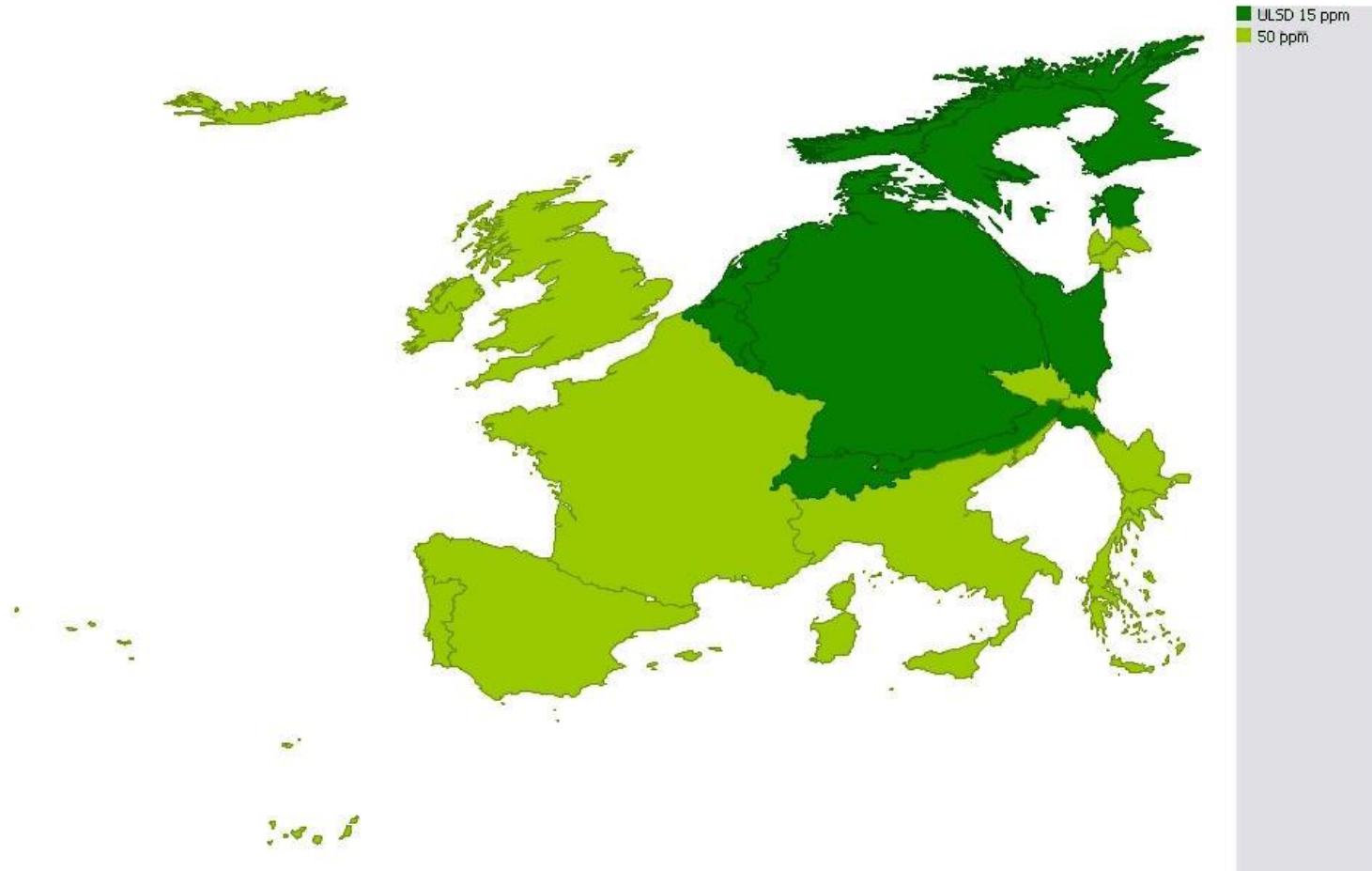


Use case: Enhancing fuel world maps



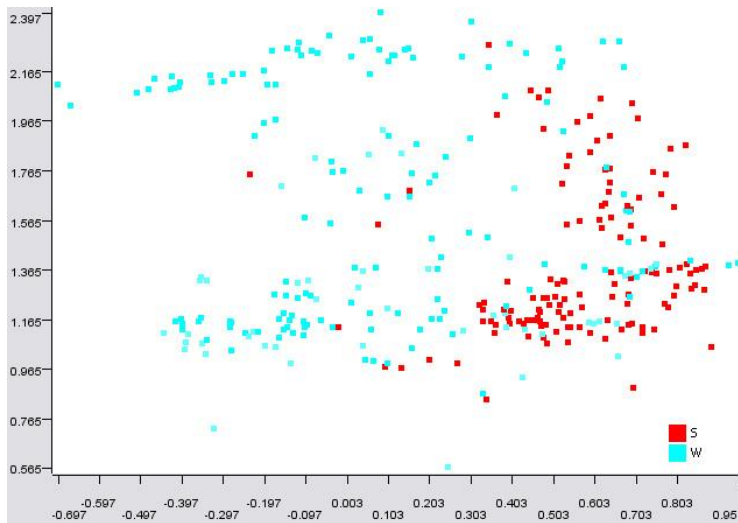
- realization of “focus & context” principle (“details on demand”)
- map processing with Java OpenJUMP GIS
- plotting sample and other locations via web service (persisted in file)
- plotting of icons according to additional statistical attributes (e.g. car registration in a country)
- different shape files (simple, high resolution, Mercator)

Use case: enhancing fuel world maps – cartograms

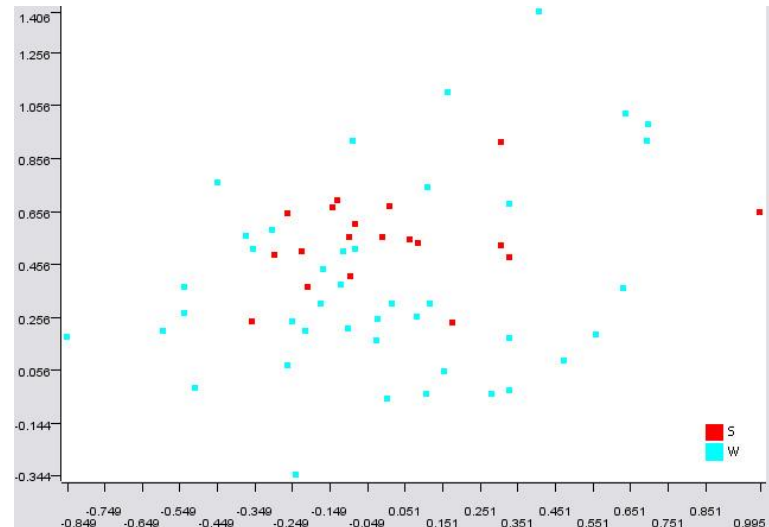


Use case: starting fuel survey data mining I

- difference of summer and winter fuel in Russia and USA
- selection 14 relevant fuel attributes
- technique: multidimensional scaling (MDS)
- explanation of the result: extreme cold winters in Russia require fuel additives



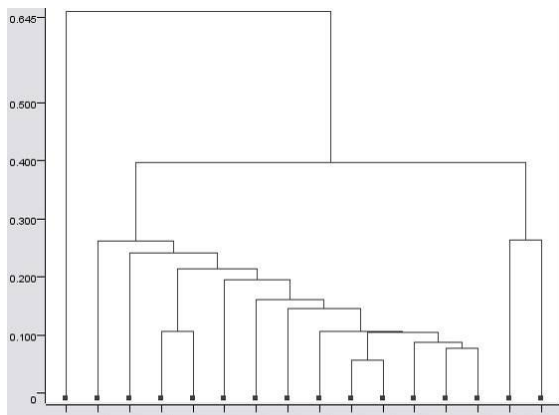
Russia



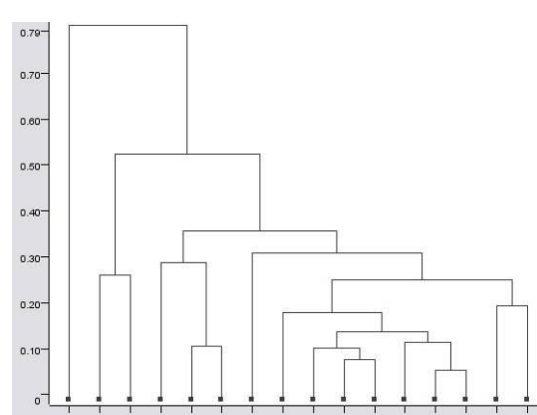
USA

Use case: starting fuel survey data mining II

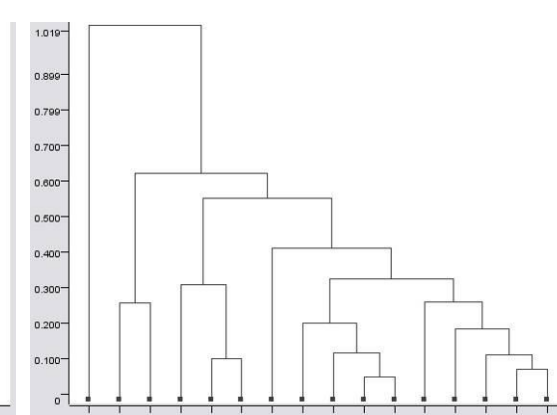
- similarity of fuel markets
- k-means clustering to determine representative datasets for each market
- hierarchical clustering, distance functions: single, complete and average linkage
- visualization with dendrograms, main results as expected



single linkage



average linkage



complete linkage

Summary and next steps

- KNIME's adaptability to different application areas is convincing:
 - easy to use,
 - good to extend and
 - with a wide range of base functions and data mining algorithms

- Some current and future use cases:
 - similarity of fuel samples near to refineries
 - clustering of boiling curves
 - outlier detection
 - prediction of missing values
 - application to other contexts (vehicle field data)

- Sharing of KNIME nodes

References

- Zaiser, Benjamin: Neue Visualisierungen hochdimensionaler Kraftstoff-Daten, Stuttgart, Hochschule der Medien, Studienbereich Medieninformatik, Master Thesis, 2010
- Zaiser, Benjamin; Maucher, Johannes; Warta, Alexander: Open Source Tool Chain for Interactive Analysis of High Dimensional Data, Working Paper, 2010

Thank you very much for your attention!

The screenshot displays the FuelSurveyMiner application interface. On the left, a KNIME workflow is visible with various nodes like 'Column Filter' and 'Table Row to Column'. The main area shows a 'Column Filter' dialog box with 'Include' and 'Exclude' options. Below it, a world map is displayed with color-coded regions. To the right, several panels show data analysis results, including a box plot and a network graph. The title 'FuelSurveyMiner' is prominently displayed in the upper right corner of the interface.

Below the screenshot, a row of logos is displayed: KNIME, Biocep-R, OPEN JUMP, HIBERNATE, ANTLR, launch4j, and BOSCH.

Contact

Dr. Alexander Warta

Robert Bosch GmbH
Diesel Systems, Engineering
Test Methods (DS/ETM1)
Postfach 30 02 20
70442 Stuttgart
Germany
www.bosch.com

Tel. +49 711 811-30554
alexander.warta@de.bosch.com