



# An overview of interactive visual data mining techniques for knowledge discovery

Frederic Stahl,<sup>1\*</sup> Bogdan Gabrys,<sup>2</sup> Mohamed Medhat Gaber<sup>3</sup> and Monika Berendsen<sup>4</sup>

In the past decade, the analysis of data has faced the challenge of dealing with very large and complex datasets and the real-time generation of data. Technologies to store and access these complex and large datasets are in place. However, robust and scalable analysis technologies are needed to extract meaningful information from these datasets. The research field of *Information Visualization* and *Visual Data Analytics* addresses this need. Information visualization and data mining are often used complementary to each other. Their common goal is the extraction of meaningful information from complex and possibly large data. However, though data mining focuses on the usage of silicon hardware, visualization techniques also aim to access the powerful image-processing capabilities of the human brain. This article highlights the research on data visualization and visual analytics techniques. Furthermore, we highlight existing visual analytics techniques, systems, and applications including a perspective on the field from the chemical process industry. © 2013 Wiley Periodicals, Inc.

## How to cite this article:

*WIREs Data Mining Knowl Discov* 2013, 3: 239–256 doi: 10.1002/widm.1093

## INTRODUCTION

It is predicted that the digital universe in the year 2020 will be 44 times as large as it was in the year 2009, where it was estimated to be 800,000 petabytes in size.<sup>1</sup> Computer systems record our daily lives,<sup>2</sup> for example, CCTV cameras and GPS-enabled navigation devices such as smart phones record our whereabouts; loyalty card and credit card issuers record our consumer behavior; and social media applications such as facebook or email capture our social life.<sup>2</sup> However,

also in the scientific arena, large quantities of data are stored and need to be analyzed. For example, in astronomy, data bases comprise terabytes of image data (see the GSC-II<sup>3</sup> or the Sloan Digital Sky survey,<sup>4</sup>) In bioinformatics, the human genome project stores the genetic blueprint of the entire human body comprising 20,000–25,000 genes and three billion chemical base pairs.<sup>5</sup> Hardware technology allows us to not only capture but also store these data. Hence, the commercial demand in developing and improving analysis techniques to extract meaningful information from the digital universe is growing.<sup>6</sup> This leads to the development of novel and scalable visual analytics techniques.

According to Xie,<sup>7</sup> in the recent years, data mining researchers have started to regard visualization techniques as a critical aspect of the decision making and data analysis process. Visualization can help analysts to visually discover different kinds of patterns such as clusters, relationships, and associations. The visual analysis capabilities of the human brain can be combined with the silicon computer hardware to improve the scalability and accuracy of the discovery of

The authors have declared no conflicts of interest in relation to this article.

\*Correspondence to: F.T.Stahl@reading.ac.uk

<sup>1</sup>School of Systems Engineering, University of Reading, Reading, UK

<sup>2</sup>The School of Design, Engineering & Computing, Bournemouth University, Bournemouth, UK

<sup>3</sup>School of Computing, University of Portsmouth, Portsmouth, UK

<sup>4</sup>Evonik Industries AG, Marl, Germany

DOI: 10.1002/widm.1093

patterns from large and complex data. Keim et al.<sup>8</sup> lay out that this combined usage of silicon and biological hardware should be augmented with the human expert's knowledge in an interactive way.<sup>8</sup> This is described as visual data analytics or short *Visual Analytics*.

The aim of this overview article is to discuss the role of visual analytics in data analysis in the overall knowledge discovery from data (KDD) process. Furthermore, this article highlights existing visual analytics software systems and applications. One particularly relevant application of data visualization and visual analytics techniques is the chemical process industry. This industry offers applications in the research for new products and production methods, as well as for monitoring the current production status and trends on large chemical plants. One of the biggest global players in the chemical industry is Evonik Industries AG and this article provides an overview on the use of visual analysis techniques in this company.

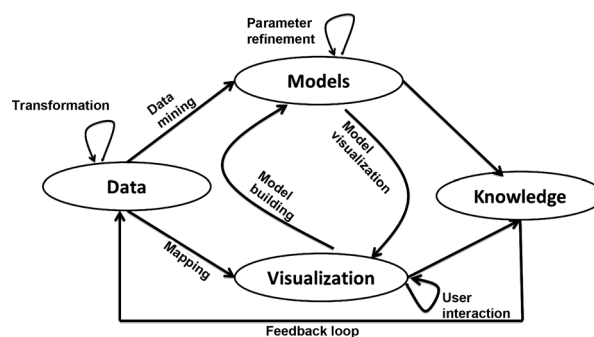
The article is organized as follows: First, the article discusses *visual data analytics* and how it fits into the overall KDD process. Then typical components and concepts often used in such visual data analytics software systems are discussed. Next, some actual visual data analytics software systems are highlighted followed by a review of visualization techniques used in data stream mining. Then, the article describes visual data analytics applications in the chemical process industry among other applications. The article is concluded by highlighting major challenges and providing an outlook into the future of visual data analytics.

## THE INTERACTIVE PROCESS OF VISUAL DATA MINING AND KNOWLEDGE DISCOVERY FROM DATA

This section highlights the basic idea of the combination of visualization and data analytics techniques and discusses the basic principles involved.

Information visualization is defined as '*the use of computer-supported, interactive, visual representations of abstract data to amplify cognition*'.<sup>9</sup> Information visualization aims to gain understanding of data using graphical representations to access the powerful image-processing capabilities of the human brain. This is described in the information seeking mantra as '*overview first, zoom/filter, details on demand*'.<sup>10</sup>

Visual analytics has been defined as the '*science of analytical reasoning facilitated by interactive visual interfaces*'.<sup>11</sup> Visual analytics is based on the same



**FIGURE 1** | The process of visual analytics according to Ref 13.

ideas as information visualization. However, visual analytics also comprises the incorporation of automatic analysis methods prior, after, and during the usage of interactive visual representations.<sup>8</sup>

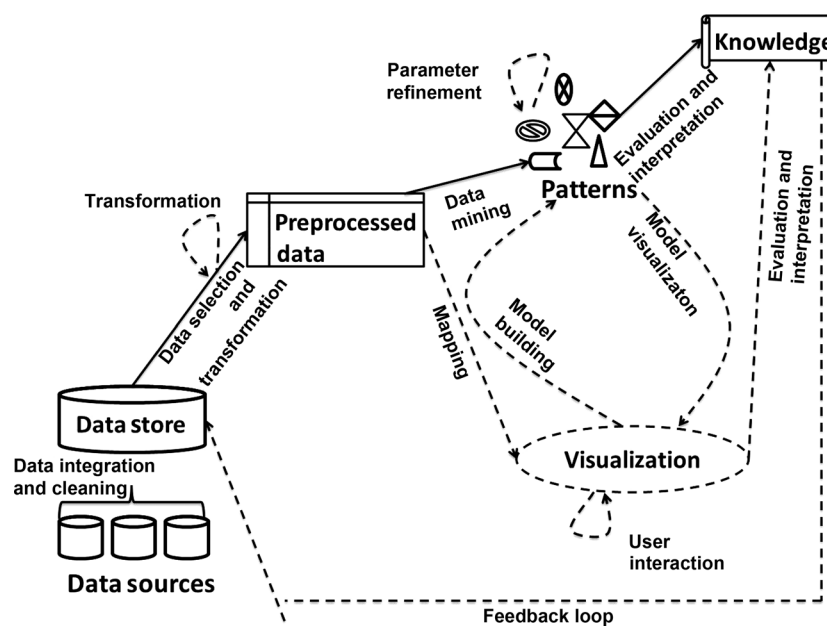
## The Principle of Visual Analytics

According to Ref 12 visual analytics as a field of research started in 2005 with Ref 11 regarded as commencement. Keim et al.<sup>13</sup> describes the visual analytics processes as a feedback loop comprising different stages and transitions<sup>13</sup> as illustrated in Figure 1. The first stage is the gathering of data. The data may come from different data sources and hence need to be integrated before further analysis methods can be applied. This step also comprises preprocessing of data such as data normalization and data cleaning. This preprocessing and integration are represented by the transformation arrow.

After this transformation, the analyst applies either visualization techniques or automated analysis methods. Automated analysis methods can comprise statistical and data mining techniques. Visualizations are then not only used to evaluate the findings, but also to refine the automated analysis methods. This alternation between automatic analysis and information visualization is the main characteristic of visual analytics and the difference when compared with just information visualization. This is also reflected in the mantra of visual analytics '*Analyze First—Show the Important—Zoom, Filter and Analyze Further—Details on Demand*'.<sup>8</sup> It is important to optimize the visualizations for the human visual system to create an effective visual analytics system; also user interaction methods such as focus and context are important for scientific visualization.<sup>14</sup>

## Visual Analytics and KDD

However, data mining for model building is just one step of the overall KDD process, and Figure 2 describes the KDD process augmented with the visual



**FIGURE 2** | The knowledge discovery from data process augmented with the visual analytics process. The dashed components and lines are the augmented elements from visual analytics.

analytics process as discussed in the ‘*The Principle of Visual Analytics*’ section.

The dashed printed components are taken from the visual analytics process, whereas the solid printed elements correspond to the KDD process. Please note that none of the KDD steps have been removed. Also, note that as opposed to the KDD process, the visual analytics process is interactive through interaction with the user of the system and the feedback loop is feeding changes back to the data input<sup>15</sup> as shown in Figure 1.

The following steps are extracted from Figure 2 and describe the main activities involved in the KDD process augmented by the process of visual analytics as it is depicted in Figure 1:

- *Data Integration and Cleaning*: It comprises the integration of different data sources into one data store or warehouse.
- *Data Selection and Transformation*: Before the data mining or analysis task, the data need to be selected from the data store. When selecting data, these data may be transformed, for example, normalized, cleaned, reduced, or converted into a certain format for the data mining step or for the visualization. In the original KDD process, the data analyst is responsible for this. However, after incorporating of visual analytics, this step may be automated with user interaction information

and the feedback used instead. Two versions of the data may be generated, one in an appropriate format for the data mining step and one for visualization purposes.

- *Data Mining*: Intelligent methods are applied to the preprocessed data to extract useful patterns such as classification or association rules. However, these models or patterns are refined by the user through interaction with their visual representation.
- *Evaluation and Interpretation*: In the original KDD process, models derived in the data mining step need to be examined to identify the interesting aspects of the extracted patterns and to find a meaningful representation. However, the application of the visualization methods can provide direct insights to the analyst.

Information visualization and visual analytics are similar approaches as they both aim to combine the human brain and silicon hardware to analyze data. However, having made the distinction between both terms and putting visual analytics in the context of KDD, one must note that these terms are often used interchangeably in the literature with information visualization actually referring to visual analytics or visual data analytics. This paper is no exception. A comprehensive book on information visualization is Ref 16 and on visual analytics is Ref 13.

Another confusion often present in the literature is the interchangeable use of the terms visual online analytical processing (visual OLAP) and visual analytics. OLAP refers to the ad hoc exploration of multidimensional data volumes where each feature represents a dimension. One common task in OLAP is to find intersections and relationships between dimensions, analyze and visualize them. Visual OLAP additionally uses visualizations to interact with the analytical algorithms and to refine them. This is very similar to visual analytics. However, visual analytics is not limited to multidimensional data.

The next section highlights some existing data mining software systems that incorporate the ideas of visual analytics and information visualization. Some of these systems are of commercial nature and some are free.

## TYPICAL CONCEPTS AND COMPONENTS OF VISUAL DATA ANALYTICS SYSTEMS

This section discusses typical components and concepts that are often found in software systems that are more focussed on the visual aspects of KDD. The visual concepts outlined in this section can be found in many visual data analysis software systems. As it has been highlighted in the previous section, visual data analysis or visual analytics is more than just static data representations. The user rather manipulates the data and data representations through interactions with multiple and diverse views. With the expression *view*, we refer to a graphical data *representation*. We may use the word *view* and *representation* interchangeably in this article. Some visual and interaction concepts that can be found in many visual data analysis systems are highlighted below. When we talk about views, we talk about a visual component in the graphical user interface that visualizes some aspects of the system such as data, configurations, data mining models, data mining results, etc. In the case when the view displays data, we also call it data representation.

### Typical User Interface Concepts

Visualizations can be as versatile as the number of possible applications or even datasets used. In fact, many data mining and visual data analysis systems such as Refs 17 and 18 are open source or at least allow to integrate new views in the form of plugins, and thus allow users to integrate their own individual application-tailored views. Hence, the section titled '*Typical Concepts and Components of Visual*

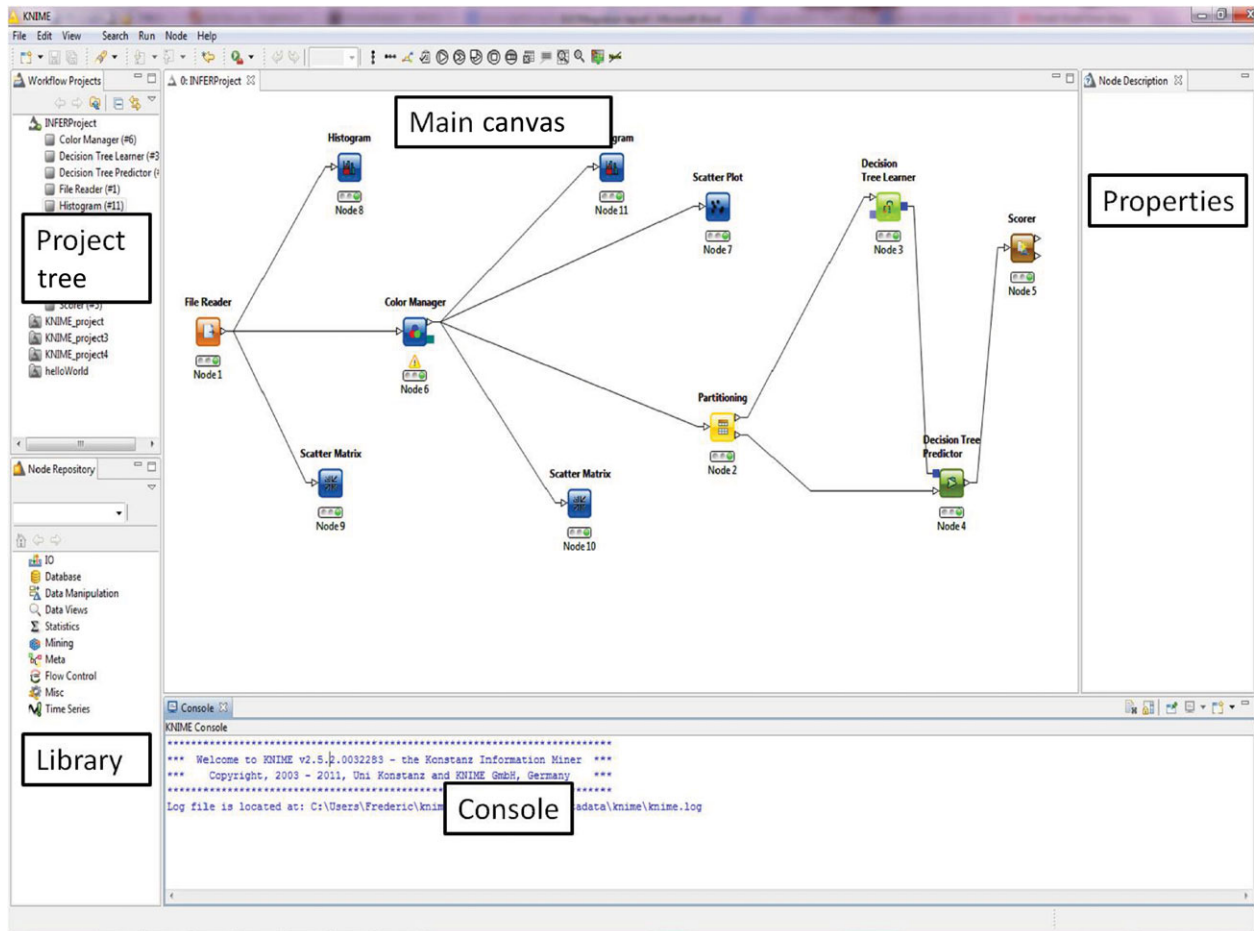
*Data Analytics Systems*' highlights visual concepts and views in a more general sense.

Visual analytics often comprises the usage of multiple views, which requires a well designed and intuitive user interface, taking into consideration the display and arrangement of the visualization and allow the user to interactively parameterize views.<sup>18</sup> Most visual data analysis tools are based on the same or a similar layout compared with the 'Eclipse Rich Client Platform'.<sup>19</sup> This layout or similar ones are used in many data mining or data visualization tools for example in Refs 17, 18, and 20. This has the advantage that the visual layout of the framework and basic interactions are already familiar with many potential users. However, some tools use their individual visualization techniques different to those used in Eclipse, for example, Weka's explorer.<sup>2</sup> Figure 3 shows a typical Eclipse-based GUI comprising the basic views labeled in black letters in the screen shot, however, additional views are possible. The *Project Tree* is usually a hierarchical navigator through the project; the *Library* offers to select tools, views, data, etc. for data mining and visualization projects; the *Properties* view usually displays information about selections in the project tree or in the main canvas; the *Main Canvas* shows a visual representation of the project, usually in the form of a workflow. However, other visualizations are possible in the main canvas such as the graph representation in the CGV tool highlighted in the '*Available Visual Data Analysis Software Systems*' section. The *Console* view produces textual and tabular output such as error messages or warnings.

### Interaction Techniques

According to the authors of Ref 21, interaction in the context of *Human Computer Interaction* can be described as the communication between the user and the system. In the context of Figure 1, user interaction is essential for the visual exploration of data and the adjustment of data mining models. Only well-designed interaction with the view and the data mining models can help the user to browse and select subsets of the data, adjust the visual mapping of the data, visually modify the parameter setting, and adjust the visual mapping of the data mining model and the associated results.

In this sense, the user realizes a *dynamic interaction*, a two-way process of feeding information into the data mining system but also retrieving new or modified information through the visualization. This is in contrast to a *passive interaction* using static images that can be observed, rotated, zoomed, and



**FIGURE 3** | A typical Eclipse RPC graphical user interface taken from the KNIME<sup>17</sup> data mining software.

un-zoomed to enhance the user's mental model on the data.<sup>22</sup> According to the authors of Ref 23, an interaction technique is defined as a way of using a physical input/output device to perform a task on using a human-computer dialogue. Some basic interaction concepts can be found in many visual data analysis systems and are discussed here in general terms rather than looking into application-specific techniques. For more information about data/information visualization, we refer the interested reader to Ref 24.

The authors of Ref 25 conducted an exhaustive review of existing software systems and the literature about information visualization and hence data visualization. In particular, 59 papers, 51 systems, and 311 identified individual interaction techniques have been reviewed.<sup>25</sup> However, the authors further point out that different interaction techniques aim to achieve the same or similar goals and hence they established a categorization of these techniques based on the notion of user intent in using these techniques. These categories are described below, which are (a)

Select, (b) Explore, (c) Reconfigure, (d) Encode, (e) Abstract/Elaborate, (f) Filter, and (g) Connect. For visual interaction examples, we refer to the articles cited within the description of the interaction techniques.

(a) **Select**: This technique of interaction enables the user to mark interesting data items in a view. This allows the user to easily identify and keep track of the data items of interest if the representation of the data is changed, especially if there are many items represented in the view. An example of select that can be found in Weka and KNIME is the scatter matrix.<sup>2,17</sup> A scatter matrix is a matrix of plots of each feature against each other. Selected data points in one plot can be highlighted across all plots in the scatter matrix. *Select* can be combined with other techniques, i.e., *Reconfiguration* to see where data items move when rearranging representations. A well-known example of *Select* used together with *Reconfiguration* is the 'placemark' feature of Google Earth.<sup>26</sup> 'Placemark' allows to select a geographical location, rotate

and zoom the view, and easily return to the marked location.

(b) Explore: This technique enables the user to view and inspect a different subset of the data. This is useful especially if the data are very large and hence screen sizes as well as the processing capabilities of the human brain limit the data representation and processing as a whole. The user can explore a subset, gain insights, and move on to explore further data. Yi et al.<sup>25</sup> further highlight two commonly used approaches to *Explore* interaction, namely *Panning* and *Direct-Walk*. In photography, panning refers to the movement of a scene in front of a fixed camera. In the context of information visualization or visual analytics, the scene is the data and the camera is the user's eye. The simplest way of moving the scene could be via scroll bars.<sup>25</sup> Direct walk allows the user to move the focus of the view from one position in the data structure to another, i.e., using hyperlinks such as in web browsers.

(c) Reconfigure: This technique enables the user to change the data representation by changing its spatial arrangement. Thus, the user is presented with different views of the data that helps to uncover hidden relationships. For example, in scatter plots some of the data elements may have similar or the same visualized numerical values, which results in the data elements to overlap in the plot. In this case, jitter can be used to make hidden data elements visible. Jitter refers to shifting the data elements randomly in the display space by a small amount to avoid the overlap of data elements.<sup>27</sup> For example, the jitter technique is used by the authors of Ref 27 and in the Spotfire<sup>28</sup> visual analytics software. However, also the rotation of 3D scatter plots may reveal overlapping data elements.

(d) Encode: This technique enables users to alter the representation of data elements. For example, altering their size, color, or shape. Altering the representation is intended to increase the human cognition in terms of understanding relationships in the data and the distribution of data elements. In Yi's review, two widely used encoding approaches are highlighted, *changing the type of the data representation* and *interaction to change the encoding of the data elements* such as color, size, or shape.<sup>25</sup> Changing the type of the data representation is intended to reveal new aspects of relationships between data elements. A tool that allows to change the data representation is Ref 28. Changing the encoding of data elements through user interaction is intended for the user to find color schemes most suitable to discover distributions of multiple variables or features. Tools that allow the user to change the encoding system based

on color or similar encoding mechanisms are for example.<sup>29,30</sup>

(e) Abstract/Elaborate: This technique enables the user to adjust the level of detail of the data representation. This allows the user to view the data in a wider context or in a more detailed view on demand often with many different levels of detail in between the context and the individual data elements. This can be realized in many different ways, i.e., tool tip texts that appear when the user moves the mouse cursor over a particular data element. However, tool tip texts usually only allow two levels of detail. More sophisticated approaches are needed, i.e., graph or tree views. The expansion and collapsing of graph or tree nodes to reveal and hide subtrees or subgraphs is, for example, implemented in the CGV system.<sup>18</sup>

(f) Filter: This interaction technique enables the user to define conditions that change the set of data items displayed in the current view. The data items that are 'filtered out' remain unchanged but are hidden or displayed in a different way. Usually, the filter interaction is complemented with a reset facility that allows to recover the hidden or differently displayed data items. A popular approach of performing filter operations in visualization tools is the usage of dynamic query controls to select conditions/ranges such as check boxes.<sup>25,31</sup> More specialized controls exist, for example in the *TimeSearcher* tool<sup>32</sup> that allows the data analyst to define conditions graphically using 'timeboxes' and 'angular queries'. 'Timeboxes' are used for filtering multiple univariate time-series profiles according to time and value ranges; and 'angular queries' are used for filtering according to the time-series value rate of change in a given time frame.<sup>32</sup>

(g) Connect: This interaction technique enables the user to highlight relationships between data items. Highlighting can happen within the same view or across different data representations. An example of highlighting between different views is KNIME's<sup>17</sup> *line plot* and *scatter plot*, data items highlighted in either of these plots are cross highlighted in the other active plot(s) as well. An example for highlighting within the same view is CGV's magic eye view.<sup>18</sup> The magic eye projects a graph's hierarchy in the form of a tree onto a 3D hemisphere. CGV extends the magic eye view to visualize cross edges among selected hierarchy nodes by spanning an arch around the 3D hemisphere.

At this stage, we would like to highlight that not all interaction techniques can be allocated to one of the categories outlined in Ref 25. Some visualizations are highly specialized on the application domain and may therefore not fit well in one or more of the categories.

## Use of Animation

According to the authors of Ref 33 animation has become popular in graphical user interfaces because of its engaging nature. Animation can be used to facilitate perception of changes in the data and data graphics.<sup>33</sup> However, the authors of Ref 33 also claim that animation has to be used with care as it may contribute to distraction if used inappropriately. Inappropriate use of animation could be for example animating irrelevant information or change that may grasp the users attention and thus misleads the users biological visual analysis system. This potential danger has resulted in research that aims to direct the usage of animation in general.

In this section, we highlight three general kinds of animations for data and information visualization that can be found in the literature and visual analytics systems. The kinds of animation are *animation of viewports*, *animated transition of graphs*, and *animated time*. With the term viewport, we refer to a typically rectangular viewing area that is of interest to the observer. *Animated viewpoints* are visualizations of the navigation space with respect to changes to the current viewport. These changes can happen, for example, through user interactions such as discussed in 'Interaction Techniques' section. For example, filtering of data items, zooming into the display area, changing the level of detail, etc. If these changes happen abruptly, then they may be too difficult to be comprehended by the observer. Animation can be used to allow the user to maintain an overview during interaction.<sup>34</sup> The authors of Ref 34 presented a generic model for the smooth animation of such changes that takes, among other aspects, the optimal animation velocity into account. This smooth viewport animation has been implemented and used, for example, by the authors of Refs 35 and 36. *Animated transitions of graphs* is the animation of switches from one data representation to another. For example, the transition from a bar chart to a pie chart to see the relative percentages of the data represented in the bar chart. Animation of such transitions is intended to allow the user to identify elements across diverse representations.<sup>33</sup> The authors of Ref 33 have investigated animated transitions between statistical data representations and derived guidelines for the design of animated transitions, which they applied in their *DynaVis* visualization system. With *animated time*, we refer to animated changes over time in the data rather than changes on the viewpoint or the representation. The representation of time can be simply facilitated by using the display space, also known as static mapping, or through animation by using the physical time, also known as *dynamic mapping*.<sup>37</sup> Both static

representation of time and dynamic representation of time are important for visual analytics. For example, the static representation of time allows the user to observe all available information and to compare the data with respect to the time, whereas dynamic mapping allows to observe the general development of the data over time.<sup>38</sup>

However, the disadvantage of animation in general is that the data may simply be too complex to be perceived by the viewer,<sup>39</sup> this is also the reason why the authors of Ref 35 refrain from animating more than one view simultaneously.

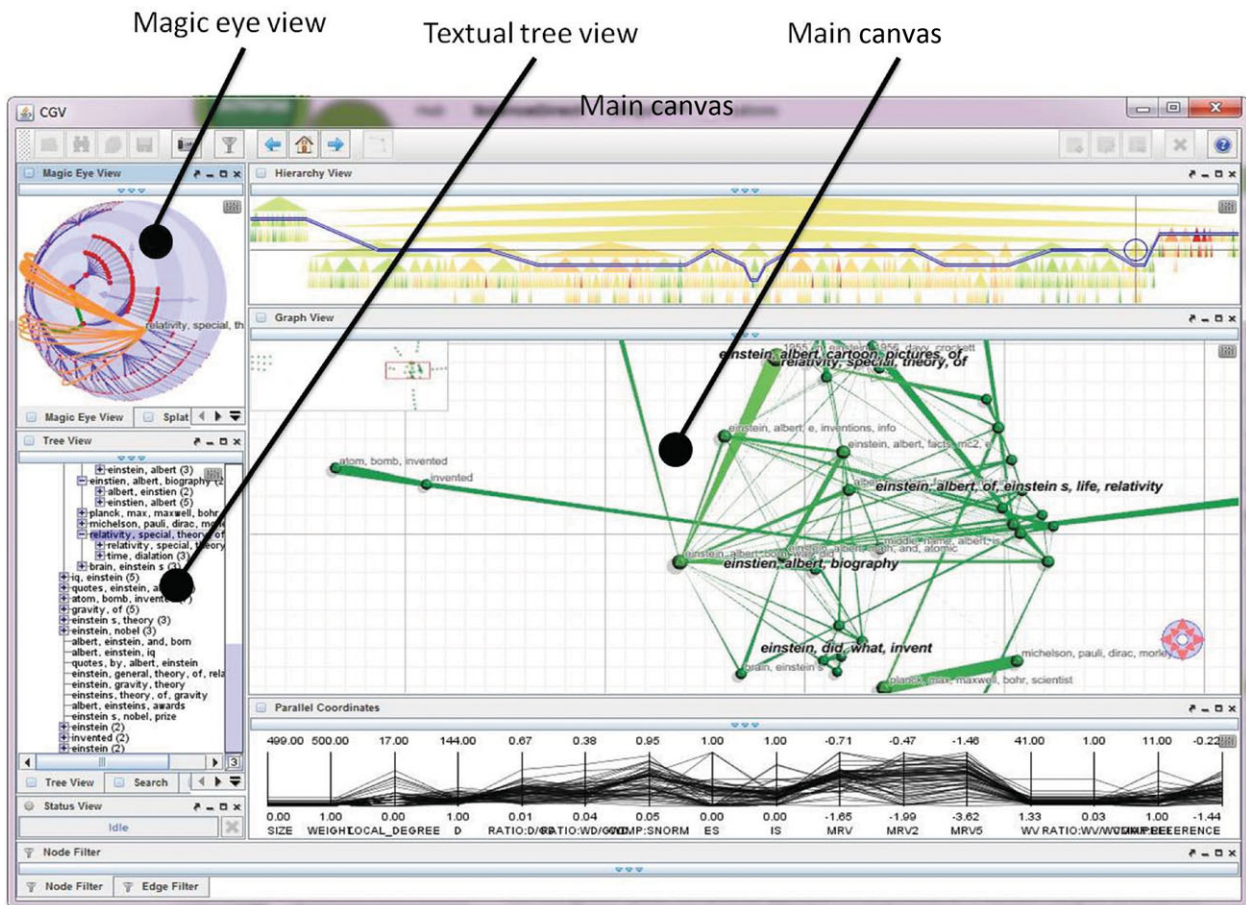
## AVAILABLE VISUAL DATA ANALYSIS SOFTWARE SYSTEMS

Standard data analysis systems such as SPSS<sup>40</sup> and Weka<sup>2</sup> already provide a wide range of interactive visualization techniques and data views. However, their visualization techniques' aim is more for reporting and the interpretation of the data mining results and data statistics rather than for the actual analysis of the data and the interactive adjustment of the data mining models. Below, we highlight some open source and some commercial data mining tools that allow to visually analyze data, some of which are tailored to a particular kind of data analysis application.

Bak et al.<sup>41</sup> use a combination of interactive visual analysis methods to effectively analyze multivariate datasets with demographic data. The authors use a self-organising map (SOM)<sup>42</sup> to visualize each cluster as a radial parallel coordinate plot.<sup>43</sup> Opacity bands<sup>44</sup> within this plot illustrate the variance within a cluster. The background color coding is used to correlate the cluster with a target value. User interaction is facilitated at different stages of the analysis as well as within the visualization.<sup>41</sup> For example, the user can modify the color maps and their scaling and can choose different types of scaling such as logarithmic scaling. The visualization also provides a 'mouse-over', which the user can utilize to identify individual cluster members and to display their data. The same techniques have also been used to analyze minority or ethnic residential patterns.<sup>45</sup>

A complete implementation of an interactive graph visualization system, called coordinated graph visualization (CGV), has been presented in Ref 18 and a test version can be obtained from Ref 35.

Figure 4 shows a screenshot of the system. The main focus of the CGV system is on user interaction, addressing in particular the problem of how to interact with a graph that is too large to be displayed on a computer screen. CGV is based on hierarchy tree



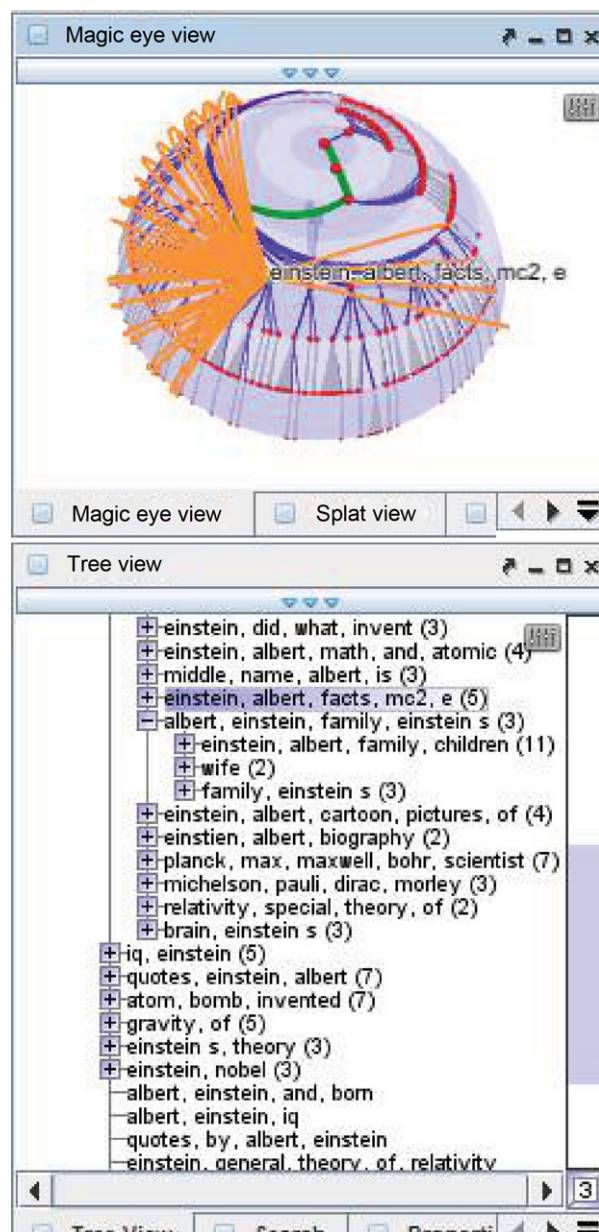
**FIGURE 4** | The graphical user interface of the Coordinated Graph Visualization (CGV) system.

computations and graph macro-views that can be displayed in the available display space.<sup>46</sup> Multiple views allow the user to explore, interact, and access the data from different perspectives simultaneously. Figure 4 shows the basic graphical user interface of CGV. The main canvas of *main graph view* uses a node-link representation of a macro-view graph. To represent the graph hierarchy, CGV uses several coordinated views. The *hierarchy view* (at the top of Figure 4) uses a superimposed polyline to visualize the current antichain and a color coding for illustrating selected node attributes.

The complementary *textual tree view* and the *magic eye view* are displayed enlarged in Figure 5. The *textual tree view* (at the left-hand side of Figure 4 and at the bottom of Figure 5) is collapsible and complements the hierarchy view by providing textual labels. The textual tree view can also be magnified by a fisheye transformation. The *magic eye view* (at the top left corner of Figure 4 and at the top of Figure 5) is based on the Walker Layout<sup>47</sup> and is extended by cross-edges among nodes. These

cross-edges are visualized using arcs spanning around the eye. However, there are further views in the CGV system, and we refer the interested reader to Ref 18 for further details. CGV distinguishes between *coordinated* and *uncoordinated interactions*. Uncoordinated interactions are specific to the particular view such as color or font size, etc., whereas coordinated interactions are changes to the global perspective of the data and hence are propagated to all views, so that all views are consistent in their representation of the data. CGV supports basic user interactions such as zooming, panning, fisheye magnification, identifying objects (coordinated), locate objects (coordinated), lock/unlock and brushing the focus (coordinated), expand/collapse graph or tree structure (coordinated), and visual parameter adjustment (uncoordinated). CGV implements filtering strategies that aim to filter out irrelevant nodes and thus enhance the clarity of the graph representation.<sup>18</sup> Furthermore, CGV implements view space and data navigation techniques, visual augmentation techniques, and undo operations that are not discussed further





**FIGURE 5** | The textual tree view and magic eye view of the Coordinated Graph Visualization (CGV) system.

here. However, a comprehensive discussion can be found in Ref 18.

The NFlowVis system<sup>48</sup> is a visual analytics system for the visualization of NetFlow data (IP traffic information) in a computer network to detect network attacks. NFlowVis uses two principal visualization techniques, *home centric flow visualization* and *graph-based flow visualization*. In the home centric approach, the local hosts that are related to attacking hosts are visualized in a TreeMap<sup>49</sup> and attacker hosts are placed at the borders. Flows between attack-

ers and local hosts are visualized using splines. The color of the local hosts and their size and the color of the splines can be used to represent various properties such as packets or bytes transferred.<sup>50</sup> Thresholds can be used to hide splines with a low traffic and highlight splines with a high traffic to the attackers as shown in the screenshot of NFlowVis in Figure 6.

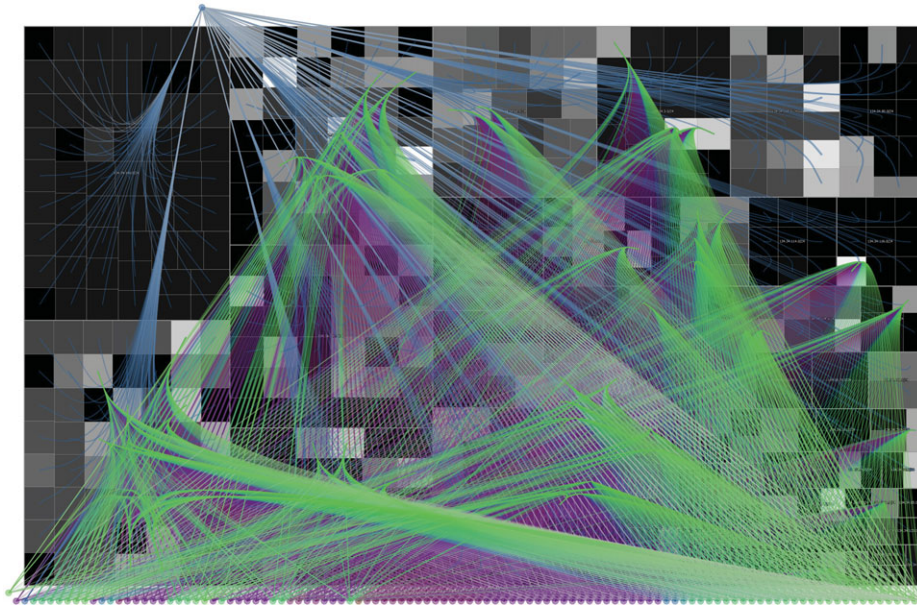
Some commercial visual analytics tools exist such as *Miner3D*<sup>51</sup> for 3D data visualization of multidimensional data, *Panopticon*<sup>52</sup> that allows to visualize real-time streams of data, and *SAS Visual Analytics*.<sup>53</sup> SAS visual analytics also emphasizes on the system's computational scalability to large data volumes, which is achieved through in-memory processing. Another tool for visual data analytics specialized for large quantities of data is *Spotfire*.<sup>28</sup> Also, long-established data analytics tools such as Matlab<sup>54</sup> have recognized the importance of data visualization and model-based visualizations components such as in Matlabs Simulink extension.<sup>55</sup>

As mentioned before, some tools such as *Panopticon* are tailored for the visualization of data streams. The following section addresses the challenges in visualizing data streams and data stream analysis techniques.

## VISUAL DATA MINING FOR DATA STREAMS

Mining data streams have attracted a big deal of attention over the last decade. A large number of techniques have been proposed to address the problem of analyzing high-rate streaming data in real time such as incremental learning algorithms.<sup>56</sup> Gaber et al.<sup>57</sup> have provided a review of clustering, classification, and frequent pattern mining techniques adopted for mining of streaming data. In a more recent review, a taxonomy of notable techniques in the area has been provided by Gaber in Ref 58. This taxonomy divided stream mining techniques to (1) two phase techniques that start with an online stage, followed by an offline one; (2) Hoeffding bound techniques that provide statistical guarantee on the approximation of the data mining output; (3) symbolic approximation that proved to be the state-of-the-art approach to time series analysis; and (4) granularity-based techniques that provide a framework for resource-adaptive data stream mining.

Limited work has been proposed for visualization of data stream mining results. This is because of the dynamic nature of the results and also to the potential large knowledge structures produced by the process (number of clusters, number of levels in a



**FIGURE 6** | The figure shows a brute force SSH attacks from the Internet to computers located at the University of Konstanz.<sup>48,50</sup> The figure has been reproduced with accordance to the Creative Commons Attribution-ShareAlike 3.0 Unported License ([http://creativecommons.org/licenses/by-sa/3.0/deed.en\\_US](http://creativecommons.org/licenses/by-sa/3.0/deed.en_US)).

decision tree, etc.). Massive Online Analysis (MOA)<sup>59</sup> is a comprehensive data stream mining tool that provides, among its functionalities, dynamic visualization of data stream clustering.

Visualization of the results of data stream mining becomes more challenging when the process run on mobile devices with small screen real-estate like smartphones. Mobile data stream mining has been well researched over the last few years.<sup>60</sup> Open Mobile Miner is a stream mining tool that is tailored to function on mobile platforms.<sup>61</sup> The screen of the mobile device can get cluttered rapidly when trying to visualize the results of a data stream mining process, running onboard the device. Addressing the problem adopting clutter reduction techniques has been reviewed in Ref 62. The reviewed techniques addressed the problem in a general way, not for mobile devices specifically. Ellis and Dix<sup>62</sup> have categorized the clutter reduction techniques into three categories: (1) representation change of data items, (2) distance change among data items, and (3) animation when a temporal dimension of the data exist. Although these approaches deal with the clutter effectively in different cases, they are not self-adaptive.

Adaptability to screen clutter for mobile devices has been proposed by Gaber.<sup>58</sup> The approach is based on a theoretical underpinning that balances perception, clutter, and amount of information of the visual image, having two observations: (1) the more

information, the greater the clutter of the visualized image; (2) the higher the clutter, the less the perception of the image. With the objective of increasing the information and reducing the clutter to enhance the perception, Gaber<sup>58</sup> has developed Adaptive Clutter Reduction (ACR).<sup>63</sup> ACR is a generic theory that states an optimization function that works on maximizing both the information and perception while reducing the amount of clutter.

ACR has been applied to data stream clustering on the smartphone. The technique has been coined clutter-aware cluster visualization (CACV). The process is divided into two steps. The first is applying the data stream clustering technique (RA-Cluster).<sup>64</sup> The visualization of the evolving clusters is done in the second step. Applying the ACR theory, all clusters are presented to satisfy two requirements: (1) maximum allowable screen coverage by the clusters and (2) maximum allowable percentage of cluster overlapping. Accordingly, a heuristic-based technique is applied. One of four levels of visualization is applied at any point in time to satisfy the two requirements. The levels are ranked according to the amount of information. Thus, the higher the level, the more information is presented on the screen. When a clutter situation occurs by having at least one of the two requirements not satisfied, the level of visualization changes to the lower ranked in the list. These levels are (1) normal: clusters are presented on the screen having the size

representing the number of points in the cluster; (2) scaling: clusters are scaled down in an attempt to satisfy the two requirements; (3) coloring/shading: all clusters would have the same size, having the darker ones represent those clusters with larger number of points; (4) active mode: only active clusters are presented on the screen. It can be observed that each level of visualization represents the level of information. Thus, CACV satisfies the requirements of the ACR theory, by keeping the highest information and perception level, whereas minimizing the screen clutter.

Visualization of data stream mining results is still in its infancy state. ACR theory represents an attempt to generalize the solution for small screen real-estate of smartphones. Nevertheless, the approach is applicable for larger screens. However, more work on the application of ACR to other data mining techniques is needed.

## VISUAL DATA MINING APPLICATIONS IN THE CHEMICAL PROCESS INDUSTRY

The application of visual analytics to commercial as well as to scientific problems is becoming popular because of the growing availability of commercial as well as scientific visual analytics tools as highlighted in the 'Available Visual Data Analysis Software Systems' section. This section discusses the usage of visual data analytics from the perspective of the chemical process industry as an example. However, further applications of visual data analytics are highlighted in the 'Further Applications of Visual Data Analysis' section.

As an example in this article, we use the case of Evonik Industries AG, which is one of the biggest global players in the chemical process industry. As of 2012, Evonik Industries AG employed roughly 33,000 employees and operates production plants in 24 countries. This and more information on Evonik Industries AG can be found at <http://evonik.com>.

Evonik Industries AG generates massive amounts of data about production processes and on laboratory-scale research. Typical applications in this industry that collect data are Process Control Systems (PCSs), Process Information Systems (PIMs), and Laboratory Information Management Systems (LIMSs). The collection of these systems is used to define Key Performance Indicators (KPIs). There are two main user groups of these systems, the *production* and the *research and technical support*. The aforementioned systems are generally used for reporting

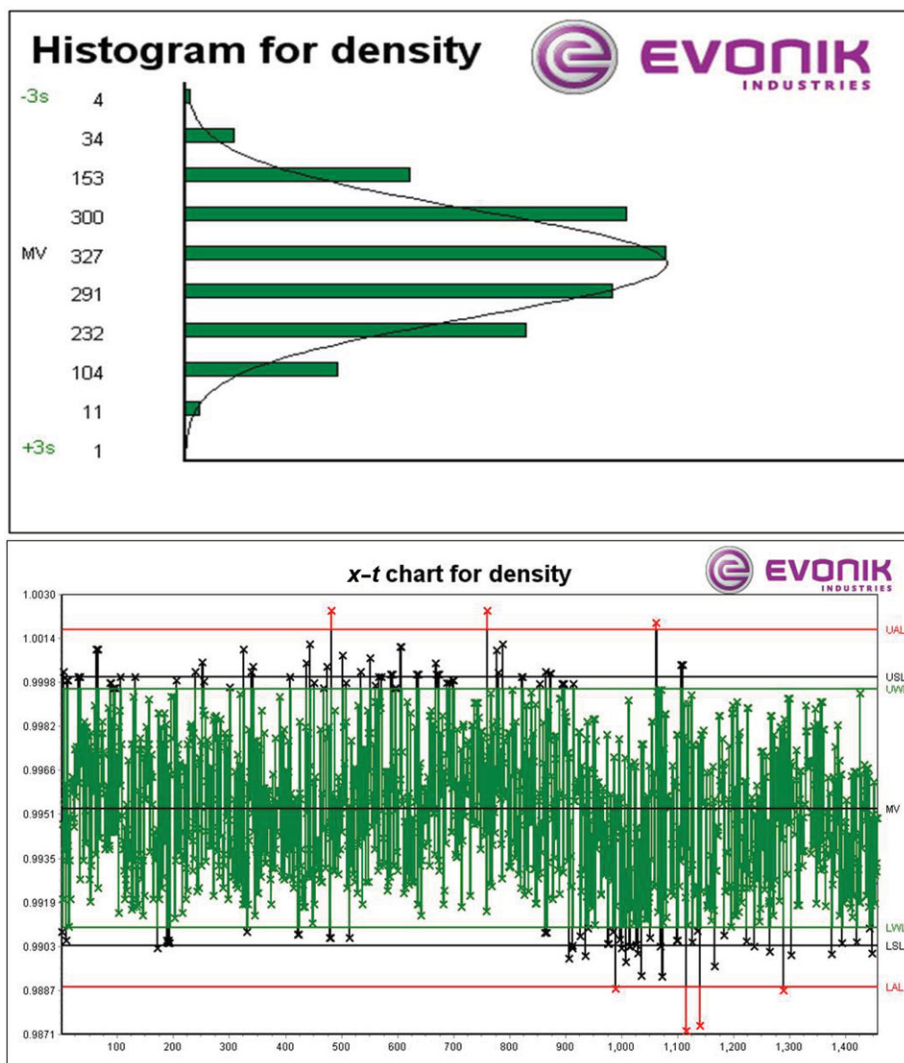
and for the improvement of the production in terms of reducing cost and improving the product quality.

For doing so, the production uses the KPIs but also aims to predict the future product quality using data mining techniques. However, within the production group, the view on visual data mining differs between plant operators and plant managers. For example, the actual plant operator desires detailed information about the actual process' state of a single active plant; the operator uses this information for supervising the current production. The plant manager desires a less detailed overview of the production but for several plants and over a longer period of time. This condensed information for the manager often needs to be displayed in so-called information dashboards. Information dashboards are easy to read condensed data representations often only showing the current status and historic and future trends. The manager uses this information for reporting and improvements of the production processes.<sup>65</sup> Yet, the data for both, the plant operators and the plant managers, are the same. Hence, visual data analytics needs to support filtering and other interaction techniques accordingly. The *Examples of Visualizations in the Production* section gives examples for the use of visual data analytics in Evonik Industries AG within the production user group.

The research user group's goal is to improve the products' recipes to improve the production quality and lower the cost of raw material. This group aims to automatize their laboratory work by using High Throughput Screening (HTS) and robotics to conduct a multitude of tests in a short period of time, examples can be found here.<sup>66,67</sup> To analyze the experimental data, this group uses information visualization and visual data analytics techniques. The 'Examples of Visualizations in Research' section gives examples for the use of visual data analytics in Evonik Industries AG within the production user group.

### Examples of Visualizations in the Production

The visualization of real-time data during a production process aims to identify the process's 'capability index', which is a measure used to estimate the optimal production capacity. There is a balance between changing production parameters (which may destabilize the process) and keeping the process stable. Statistical information and data visualization help to decide about interventions. At Evonik Industries AG, an application called ChemSPF,<sup>68</sup> an in-house development, is used for the statistical evaluation in process analysis and protection of process capability.



**FIGURE 7** | Charts used by Evonik Industries AG to display quality attributes during the production process. The chart on the bottom half of the figure is called a ChemSPFchart and the chart on the top half of the figure is a histogram.

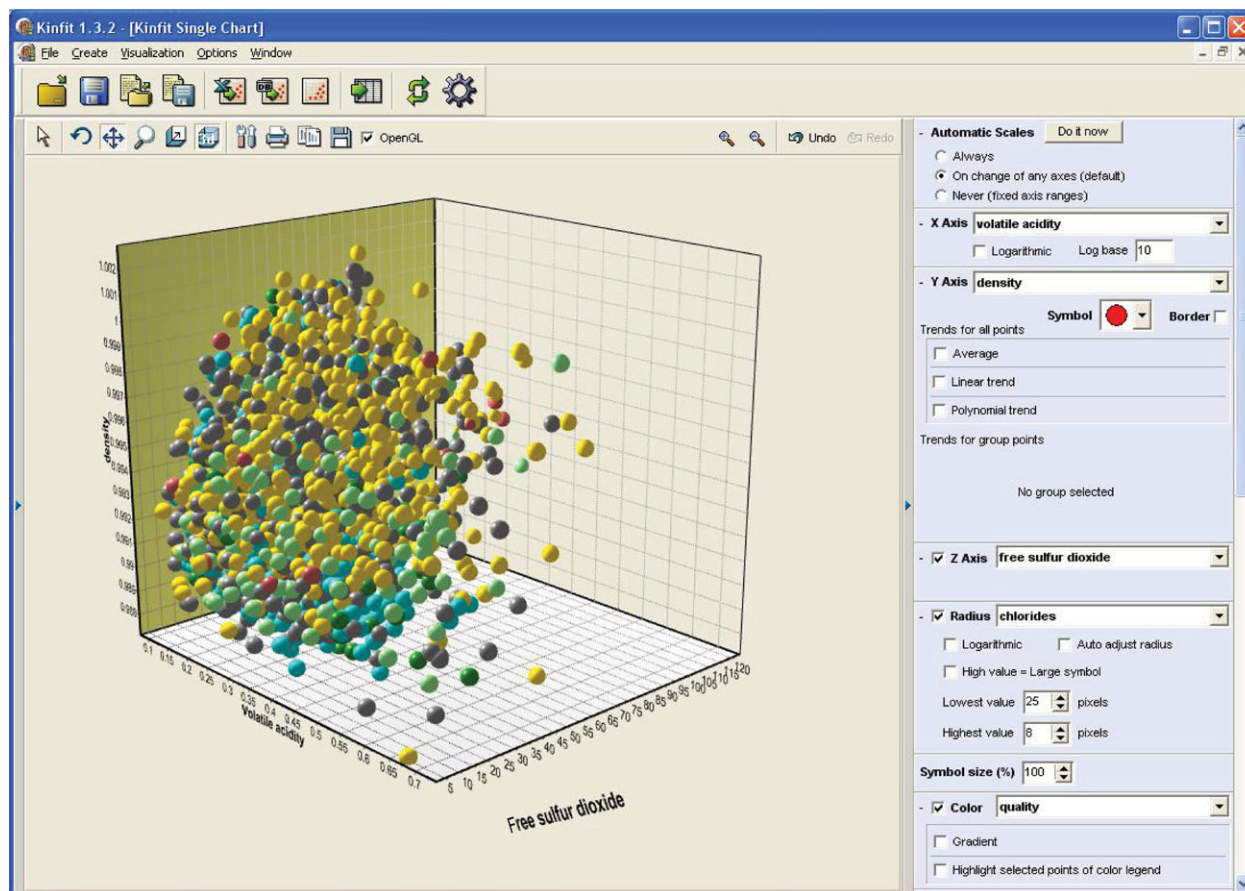
The ChemSPF-chart and histogram in Figure 7 show some of the visualizations used to display quality attributes of an ongoing production. The ChemSPF-chart is on the bottom half of Figure 7, it also displays the mean value (MV), upper (USL) and lower limits (LSL) shown for triggering a warning, as well as upper and lower limits for triggering interventions (UAL and LAL). Please note that the acronyms used in Figure 7 are based on the German language and hence the acronym expansions are omitted here. Also, the name ChemSPF charts comes from a German expression and is also known as Statistical Process Control (SPC) charts in the English language.

These are just two examples of standard visualizations used in the production. However, there are many more including text-based information. What

is important to note is that the charts in Figure 7 display only one variable. However, in a production process, there may be hundreds of variables with different value ranges that have to be observed in real time. Hence, these charts have to be redrawn in adequate time intervals and need to convey large amounts of information simultaneously.

### Examples of Visualizations in Research

The production group uses mostly basic visualizations, whereas the research group needs to observe as many variables of historic productions or laboratory-scale productions within the same visualization. The task for the researcher is to extract as much information as possible from the data. Evonik Industries AG



**FIGURE 8** | Multidimensional representation of five variables using KinFit.

uses an application that makes use of visual data analytics techniques, which is used by researchers for analyzing large quantities of data. The application is called KinFit<sup>69</sup> and is an in-house development. One task of visual analytics in research at Evonik Industries AG, is to find correlations between variables. Therefore, KinFit gives a data overview by showing each dimensions plotted against each other, which is also known as scatter matrix in other data mining tools such as WEKA.<sup>2</sup> Within Evonik Industries AG the visual analytics techniques in KinFit are used for exploring production data and for general research purposes. Information visualizations in KinFit are also used for reporting.

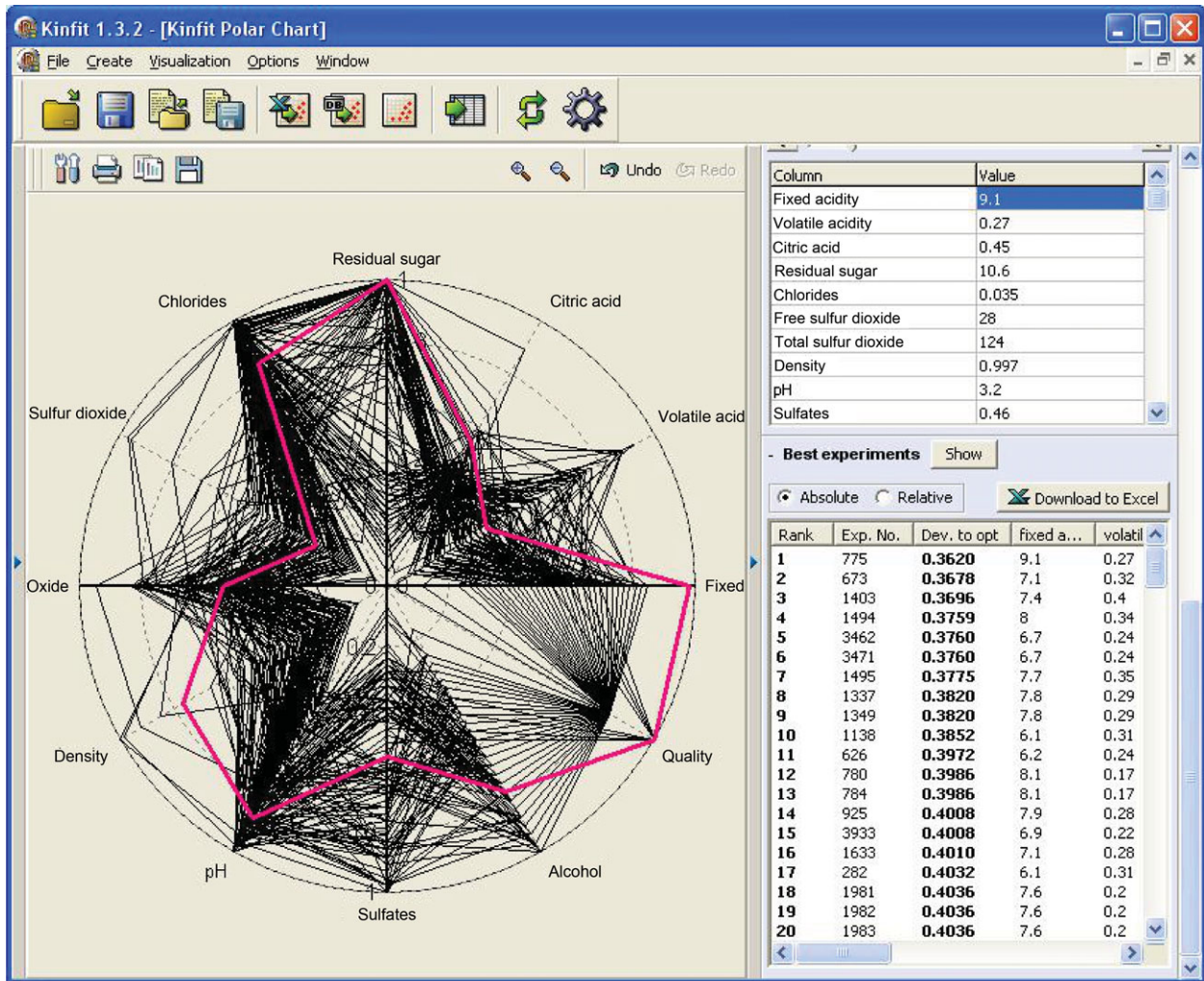
Many more interactive visualization approaches are used. For example, to visualize multiple dimensions (variables) in the same chart, multi-dimensional charts such as displayed in Figure 8 are frequently used. The user can change the encoding of the data elements such as colors, symbols, and the symbol's size as well as moving, zooming, or rotating the graph. KinFit also allows to animate the graph to show a sixth dimension such as time.

A polar or radar chart such as the one displayed in Figure 9, also called spider chart, shows in each spoke one dimension. Each line is representing one data record. This chart can be used for detecting outliers and clusters. The user can interact with the graph parameters and set for each dimension an optimal value and a weight. The application then produces the range of the optimal records as displayed in Figure 9.

However, one challenge of visual analytics that remains to be solved in the chemical process industry is to convey simultaneously thousands of variables with different and changing values ranges in real time.

### Further Applications of Visual Data Analysis

The last two subsections showed that visual data analytics is one of the emerging technologies for data analysis in the chemical process industry. Also, a range of applications has been mentioned throughout the article, especially in the 'Available Visual Data Analysis Software Systems' section. However, further



**FIGURE 9 |** KinFit’s polar chart implementation for displaying data with multiple dimensions, typically used for identifying optimal production process states.

applications exist. For example, the authors of Ref 70 propose a two-stage visualization approach for fraud detection in stock market trading. In the first stage, they use 3D tree maps to observe real-time stock market performance and detect unusual patterns; and in the second stage, they use social network visualization to analyze the behavior of the suspected pattern to detect imminent attacks. In Ref 71, the authors presented a visual analytics tool that visualizes the aggregate risk alongside a traditional wealth–time plot. The system is aimed at nonexpert users to help with personal finance decisions. The system has been used in a laboratory environment by 27 volunteers and been evaluated using several economics methods. For example, the results shows that the users gain greater returns for similar levels of risk. Also, it shows that users of the system explore and learn

more than the control group, which is reflected in a higher number of investment modifications of system users. Also 23% of the users of the system claimed that they feel more confident in understanding financial planning, whereas 0% of the control group made this claim. In general, the results show that the system improves the decision-making process. The authors of Ref 72 applied visual analytics on multidimensional gene expression datasets. Their approach is to visualize the gene expression conjoined with statistical data retrieved from the original data. The authors of Ref 73 uses visual analysis techniques in the physics arena to gain insights into simulations of magnetically confined burning plasmas. The authors of Ref 74 have developed a visual analytics system that has been used in a variety of tasks in the context of analyzing hyperspectral images of historical documents.

## CONCLUSIONS AND MAJOR CHALLENGES

As discussed in the previous sections of this paper, visual analytics as a field of study is only just emerging. We have attempted to explain that this new field is a combination of interactive visualization and advanced data and predictive analytics that includes building, verifying, and using predictive models. As shown in Figures 1 and 2, it is not only the visualization of the data and its exploratory analysis, the process that is often mistakenly given the name of visual analytics, but also an interactive method available for semiautomated clustering, classification, prediction, dimensionality reduction, information retrieval, feature extraction, etc, which leads to knowledge and actionable decision generation.

In the process of such a visual analytics session, one would expect the underlying predictive models to be searching for and exploiting user feedback in many different forms. Such feedback could take a form of extracting user preferences by monitoring what is used and what is focused on, which in turn could be translated into modified optimization criteria by the underlying models to adapt to the user's current needs. We would also expect that the users should be provided with powerful wizard-driven ways of exploiting and building data processing methods without a need to understand or set any of the technical parameters of such advanced methods. This represents the model building and model visualization interactive loop in Figures 1 and 2.

The visualization of extracted knowledge in turn drives further refinements. This focus on the data or how the data or the existing models should be transformed closes the loop that can be repeated a number of times. This process becomes even more challenging and interesting when the data arrive continuously (i.e., streaming) and the predictive models must operate in a (semi-)automated manner. Visualization of the changing knowledge that is brought to

the attention of the user and suggestions of what may be interesting in the recent data are then also a part of the picture, a picture that may be changing very quickly.

There is a growing number of visual data analysis software tools such as the ones highlighted in the 'Available Visual Data Analysis Software Systems' section and various components required for building of visual analytics systems and tools are also being developed and discussed in related fields as highlighted in the 'Typical Concepts and Components of Visual Data Analytics Systems' section, but what remains a formidable challenge is to bring all of the components together in robust tools that can be used by users with a wide range of technical abilities and know-how. As illustrated in a snapshot of visualization tools that are used by one of our partners in the process industry (see Section 'Visual Data Mining Applications in the Chemical Process Industry'), we are far from such visual analytics scenarios as the current practises very often end on interactive visualization for which the tools are available. It is the role of the data analysts to explore the data and look for things that may be interesting, unusual, dangerous, and promising. We are still very heavily relying on the creativity and curiosity of humans who in addition must have quite formidable technical abilities and knowledge of data and predictive analytics tools to realize their full potential.

In developing a next generation of visual analytics systems, one would hope to relax (if not remove completely) this need for the user's deep knowledge of the predictive analytics methods and enhance the ability of visual analytics systems so they in part can become more autonomous and 'creative' and 'curious' in looking for things that are interesting in some sense. To reach such a state, a lot of research and development effort are still needed and one of this publication's goals is to bring this fact to the attention of wider communities with keen interest in what is emerging as the visual analytics field.

---

## ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Commission within the Marie Curie Industry and Academia Partnerships and Pathways (IAPP) programme under grant agreement no 251617.

## REFERENCES

1. Gantz J, Reinsel D. The digital universe decade, are you ready? In: *IDC 2009; 2010*, 1–16.
2. Witten IH, Eibe F. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. 2nd ed. San Francisco, CA: Morgan Kaufmann; 2005.
3. McClean B, Hawkins C, Spagna A, Lattanzi M, Lasker B, Jenkner H, White R. New horizons from multi-wavelength sky surveys. In: *Proceedings of the 179th Symposium of the International Astronomical Union held in Baltimore, 1998*.
4. Sloan digital sky survey. Available at: <http://www.sdss.org/>. (Accessed May 13, 2013).
5. Human genome project. Available at: [http://www.ornl.gov/sci/techresources/Human\\_Genome/home.shtml](http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml). (Accessed May 13, 2013).
6. Stahl F, Gaber MM, Bramer M. Scaling up data mining techniques to large datasets using parallel and distributed processing. In: Rausch P, Sheta AF, Ayesha A, eds. *Business Intelligence and Performance Management. Advanced Information and Knowledge Processing*. London: Springer; 2013, 243–259.
7. Xie Z. *Exploratory Visualization of Data Pattern Changes in Multivariate Data Streams*. PhD thesis, Worcester Polytechnic Institute, 2011.
8. Keim DA, Mansmann F, Schneidewind J, Ziegler H. Challenges in visual data analysis. In: *Proceedings of the Conference on Information Visualization, IV '06*. Washington, DC: IEEE Computer Society; 2009, 9–16.
9. Card SK, Mackinlay JD, Shneiderman B. Readings in information visualization: using vision to think. *Morgan Kaufmann Series in Interactive Technologies*. San Francisco, CA: Morgan Kaufmann Publishers; 1999.
10. Shneiderman B. The eyes have it: a task by data type taxonomy for information visualizations. In: *Proceedings of the 1996 IEEE Symposium on Visual Languages, VL '96*. Washington, DC: IEEE Computer Society; 1996, 336–343.
11. Thomas JJ, Cook KA, Eds. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. National Visualization and Analytics Center; 2005.
12. Thomas J, Kielman J. Challenges for visual analytics. *Inf Vis* 2009, 8:309–314.
13. Keim DA, Kohlhammer J, Ellis G, Mansmann F, Eds. *Mastering the Information Age—Solving Problems with Visual Analytics*. Goslar, Germany: Eurographics Association; 2010.
14. Kruger J, Schneider J, Westermann R. Clearview: an interactive context preserving hotspot visualization technique. *IEEE Trans Vis Comput Graph* 2006, 12:941–948.
15. Keim DA, Mansmann F, Schneidewind J, Thomas J, Ziegler H. Visual data mining. In: *Visual Analytics: Scope and Challenges*. Berlin, Heidelberg: Springer-Verlag; 2008, 76–90.
16. Chen C. *Information Visualization: Beyond the Horizon. Computer Science*. New York: Springer, 2006.
17. KNIME.com AG. Knime—professional open-source software. Available at: <http://www.knime.org/>. (Accessed May 13, 2013).
18. Tominski C, Abello J, Schumann H. CGV—an interactive graph visualization system. *Comput Graph* 2009, 33:660–678.
19. KNIME.com AG. Eclipse rich client platform. Available at: <http://www.eclipse.org/home/categories/rcp.php>. (Accessed May 13, 2013)
20. SAS. Sas enterprise guide. Available at: [http://www.sas.com/technologies/bi/query\\_reporting/guide/](http://www.sas.com/technologies/bi/query_reporting/guide/). (Accessed May 13, 2013)
21. Dix A. *Human–Computer Interaction*. Harlow, England: Pearson/Prentice-Hall; 2004.
22. Spence R. *Information Visualization: Design for Interaction*. Harlow, England: Pearson/Prentice Hall; 2007.
23. Foley JD, van Dam A, Feiner SK, Hughes JF. *Computer Graphics: Principles & Practice*. In: C. 2nd ed. Addison-Wesley; 1996.
24. Ware C. *Information Visualization: Perception for Design. Interactive Technologies*. Waltham, MA: Elsevier Science; 2012.
25. Yi JS, Kang YA, Stasko J, Jacko J. Toward a deeper understanding of the role of interaction in information visualization. *IEEE Trans Vis Comput Graph* 2007, 13:1224–1231.
26. Google Inc. Google earth. Available at: <http://www.google.com/earth/index.html>. (Accessed May 13, 2013).
27. Migut M, Worring M. Visual exploration of classification models for various data types in risk assessment. *Inform Vis* 2012, 11:237–251.
28. TIBCO Software. Spotfire. Available at: <http://spotfire.tibco.com/>. (Accessed May 13, 2013).
29. Macrofocus GmbH. Infoscope. Available at: <http://www.macrofocus.com/public/products/infoscope/>. (Accessed May 13, 2013).
30. Yi JS, Melton R, Stasko J, Jacko JA. Dust magnet: multivariate information visualization using a magnet metaphor. *Inform Vis* 2005, 4:239–256.
31. Ahlberg C, Williamson C, Shneiderman B. Dynamic queries for information exploration: an implementation and evaluation. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '92*. New York: ACM Press. May 3–7, 1992; 619–626.
32. Hochheiser H, Shneiderman B. Dynamic query tools for time series data sets: timebox widgets for interactive exploration. *Inform Vis* 2004, 3:1–18.
33. Heer J, Robertson GG. Animated transitions in statistical data graphics. *IEEE Trans Vis Comput Graph* 2007, 13:1240–1247.



34. van Wijk JJ, Nuij WAA. A model for smooth viewing and navigation of large 2d information spaces. *IEEE Trans Vis Comput Graph* 2004, 10:447–458.
35. CGV—Coordinated graph visualization. Available at: <http://www.informatik.uni-rostock.de/~ct/software/CGV/CGV.html>. (Accessed May 13, 2013).
36. Cockburn A, Savage J, Wallace A. Tuning and testing scrolling interfaces that automatically zoom. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '05. ACM Press; 2005, 71–80.
37. Rind A, Neubauer B, Aigner W, Miksch S, Matkovic K, Santucci G. Static and dynamic visual mappings to explore bivariate data across time. In: *Proceeding of the 3rd International Eurovis workshop on Visual Analytics (EuroVA)*. Vienna, Austria; 2012.
38. Aigner W, Miksch S, Mueller W, Schumann H, Tominski C. Visualizing time-oriented data—a systematic view. *Comput Graph* 2007, 31:401–409.
39. Tversky B, Morrison JB, Betrancourt M. Animation: can it facilitate? *Intl J Hum-Comput Stud* 2002, 57:247–262.
40. IBM. Spss software. Available at: <http://www-01.ibm.com/software/analytics/spss/>. (Accessed May 13, 2013).
41. Bak P, Omer I, Schreck T. Visual analytics of urban environments using high-resolution geographic data. In: Painho M, et al. eds. *Geospatial Thinking, Lecture Notes in Geoinformation and Cartography*. Berlin: Springer; 2010, 25–42.
42. Kohonen T. *Self-Organizing Maps*. Springer Series in Information Sciences. Berlin: Springer; 2001.
43. Van Long, T, Linsen L. Multiclustertree: interactive visual exploration of hierarchical clusters in multidimensional multivariate data. *Comput Graph Forum* 2009, 28:823–830.
44. Fua Y-H, Ward MO, Rundensteiner EA. Hierarchical parallel coordinates for exploration of large datasets. In: *Proceedings of the conference on Visualization '99: celebrating ten years*, VIS '99. Los Alamitos, CA: IEEE Computer Society Press; 1999, 43–50.
45. Omer I, Bak P, Schreck T. Using space-time visual analytic methods for exploring the dynamics of ethnic groups' residential patterns. *Int J Geo Inform Sci* 2010, 24:1481–1496.
46. Abello J, van Ham F, Krishnan N. Ask-graphview: a large scale graph visualization system. *IEEE Trans Vis Comput Graph* 2006, 12:669–676.
47. Buchheim C, Jünger M, Leipert S. Improving walker's algorithm to run in linear time. In: *Revised Papers from the 10th International Symposium on Graph Drawing*, GD '02; London, UK. UK: Springer-Verlag; 2002, 344–353.
48. Mansmann F, Fischer F, Keim DA, North SC. Visual support for analyzing network traffic and intrusion detection events using treemap and graph representations. In: *Proceedings of the Symposium on Computer Human Interaction for the Management of Information Technology*, CHiMiT '09. New York, NY: ACM Press; 2009, 3:19–3:28.
49. Shneiderman B. Tree visualization with tree-maps: 2-d space-filling approach. *ACM Trans Graph* 1992, 11:92–99.
50. Fischer F, Mansmann F, Keim DA, Pietzko S, Waldvogel M. Large-scale network monitoring for visual analysis of attacks. In: *Proceedings of the 5th international workshop on Visualization for Computer Security*, VizSec '08. Berlin, Heidelberg: Springer-Verlag; 2008, 111–118.
51. The miner3d data visualisation software. Available at: <http://www.miner3d.com/>. (Accessed May 13, 2013).
52. Panopticon real-time visual data analysis. Available at: <http://www.panopticon.com/>. (Accessed May 13, 2013).
53. SAS. Sas visual analytics. Available at: <http://www.sas.com/technologies/bi/visual-analytics.html> (Accessed May 13, 2013).
54. Explore, model, and visualize data. Available at: <http://www.mathworks.co.uk/data-analysis/>. (Accessed May 13, 2013).
55. Simulation and model based design. Available at: <http://www.mathworks.co.uk/products/simulink/>. (Accessed May 13, 2013).
56. Bouchachia A, Gabrys B, Sahel Z. Overview of some incremental learning algorithms. In: *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'2007): Proceedings*. London, UK; July 2007; 2007; 1–6.
57. Gaber MM, Zaslavsky A, Krishnaswamy S. Mining data streams: a review. *ACM Sigmod Record* 2005, 34:18–26.
58. Gaber MM. Advances in data stream mining. *Wiley Interdiscip Rev Data Mining Knowledge Discov* 2012, 2:79–85.
59. Bifet A, Holmes G, Kirkby R, Pfahringer B, Seidl JT, T Kranen P, Kremer H. Moa: massive online analysis. *J Mach Learn Res* 2010, 11:1601–1604.
60. Krishnaswamy S, Gama J, Gaber MM. Advances in data stream mining for mobile and ubiquitous environments. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*. Glasgow, UK: ACM Press; 2011, 2607–2608.
61. Haghghi P, Krishnaswamy S, Zaslavsky A, Gaber M, Sinha A, Gillick B. Open mobile miner: a toolkit for building situation-aware data mining applications. *J Org Comput Electron Com*. In press.
62. Ellis, G, Dix A. A taxonomy of clutter reduction for information visualisation. *IEEE Trans Vis Comput Graph* 2007, 13:1216–1223.
63. Gaber MM, Krishnaswamy S, Gillick B, AlTaiar H, Nicoloudis N, Liono J, Zaslavsky AB. Interactive

- self-adaptive clutter-aware visualisation for mobile data mining. *J Comput Syst Sci* 2013, 79:369–382.
64. Gaber, MM, Yu PS. A holistic approach for resource-aware adaptive data stream mining. *New Gen Comput* 2006, 25:95–115.
  65. Adams M, PSI AG WK. Din en 62264. die neue norm zur interoperabilitaet von produktion und unternehmensfuehrung-teil 1., *Automatisierungstechnische Praxis* 2007, 49:S.52–S.57.
  66. Heddle, C, Mazaleyrat SL. Development of a screening platform for directed evolution using the reef coral fluorescent protein zsgreen as a solubility reporter. *Prot Eng Design Select* 2007, 20:327–337.
  67. Koppal T. Ask the expert setting up high-throughput screening laboratory. Available at: <http://www.labmanager.com/?articles.view/articleNo/3574/title/Ask-the-Expert-Setting-up-High-Throughput-Screening-Laboratory/>. (Accessed May, 2013).
  68. Vornholt G. *User-manual chemspf 4.0. Manual*. Evonik Industries AG; 2008.
  69. Porwol D. *User-manual kinfit 2.0. Manual*. Evonik Industries AG; 2012.
  70. Huang ML, Liang J, Nguyen QV. A visualization approach for frauds detection in financial market. In: *Proceedings of the 13th International Conference on Information Visualisation*. Barcelona, Spain; 2009, 197–202.
  71. Rudolph S, Savikhin A, Ebert DS. Finvis: Applied visual analytics for personal financial planning. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology, VAST 2009*. Atlantic City, New Jersey, USA; 2009, 195–202.
  72. Dietzsch J, Heinrich J, Nieselt K, Bartz D. Spray: A visual analytics approach for gene expression data. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology, VAST 2009*. Atlantic City, New Jersey, USA; 2009, 179–186.
  73. Sanderson A, Chen G, Tricoche X, Pugmire D, Kruger S, Breslau J. Analysis of recurrent patterns in toroidal magnetic fields. *IEEE Trans Vis Comput Graph* 2010, 16:1431–1440.
  74. Kim SJ, Zhuo S, Deng F, Fu C-W, Brown M. Interactive visualization of hyperspectral images of historical documents. *IEEE Trans Vis Comput Graph* 2010, 16:1441–1448.