Exploratory Visualization for Weather Data Verification

CONFERENCE PAPER · AUGUST 2011

DOI: 10.1109/IV.2011.79 · Source: IEEE Xplore

4 AUTHORS, INCLUDING:

Jimmy Johansson

Linköping University / Norrköping Visualizati…

Exploratory Visualization for Weather Data Verification

Patrik Lundblad*, Hanna Löfving†, Annika Elovsson† and Jimmy Johansson†
*Swedish Meteorological and Hydrological Institute, Sweden
†C-Research, Linköping University, Sweden
patrik.lundblad@smhi.se, {hanlo184,annel047}@student.liu.se, jimmy.johansson@liu.se

Abstract

Today weather forecasts assist in the preparation of roads during the winter season to avoid the accidents resulting from snow, rain and slipperiness. As with any weather forecast there is always a search for improvement. One way to do this is by verification of the forecasted parameters with the actual weather observed at the forecasted area.

To facilitate identification of significant trends and patterns within weather data we have developed an application based on interactive information visualization techniques. The application was created in close collaboration with domain experts from the Swedish Meteorological and Hydrological Institute and initial feedback from a performed user study shows that interactive visualization speeds up the analysis process as well as increases flexibility compared to currently used manual methods.

Keywords— **Information visualization, verification of weather data, exploratory analysis**

1 Introduction

Extreme weather is today a force to reckon with at the same time as the streamlined infrastructure is vulnerable to unexpected events. This creates a demand for both accurate as well as long time forecasts so that preparations may be taken for the effects of weather to improve the safety for people. One area of particular interest is road safety during the winter seasons. Every year the roads are effected by rain, snow and ice that makes them dangerous for driving unless preventive methods are taken. The decreasing respect for bad weather conditions, due to modern aids such as anti-lock and anti-spin systems, is another factor that contributes to many accidents during the winter season. Inaccurate weather forecasts can also mean unnecessary costs for municipalities sending out plough cars when the forecasts wrongfully indicate snow. Preventive methods are taken using forecasts provided by meteorologists specifically assigned to the task of making forecasts and warnings within this domain. Verification is needed to improve the forecast models and help meteorologists find specific scenarios that will aid them in their work. Consequently, it is important that these forecasts are as accurate as possible.

In this paper we present a web-enabled application for interactive exploration of weather data verification for analysis of road weather conditions. The application has been developed in close collaboration with domain experts from the Swedish Meteorological and Hydrological Institute (SMHI). The input for the application consists of hourly measured observations along Swedish roads together with daily forecasts for the observation points.

The interface consists of visual representations that allow both overview as well as detailed analyses. The overview components are used to analyze the full season and enable searching for systematical errors that occur over long time periods. These components can also be used to investigate the accuracy of forecasts for different geographical areas. With the aid of a Self-Organizing Map (SOM) the multivariate data is projected to a lower dimension to facilitate analyses. The detail components are used to explore specific scenarios during the season.

The web-based application, figure 1, has been developed to support the following major exploratory tasks:

- Investigation of significant trends and patterns for verification of the underlying forecast model.
- Communication of accuracy measurements for competitive comparisons.
- Visual analysis on multiple levels of detail; from the full winter season down to individual hours.

The remainder of this paper is structured as follows. Related work is provided in section 2. General description of attributes in the input data and events in section 3. In section 4 the application is described followed by an application scenario in section 5. Section 6 contains a user study and finally the paper is concluded with the general results and a discussion in section 7.

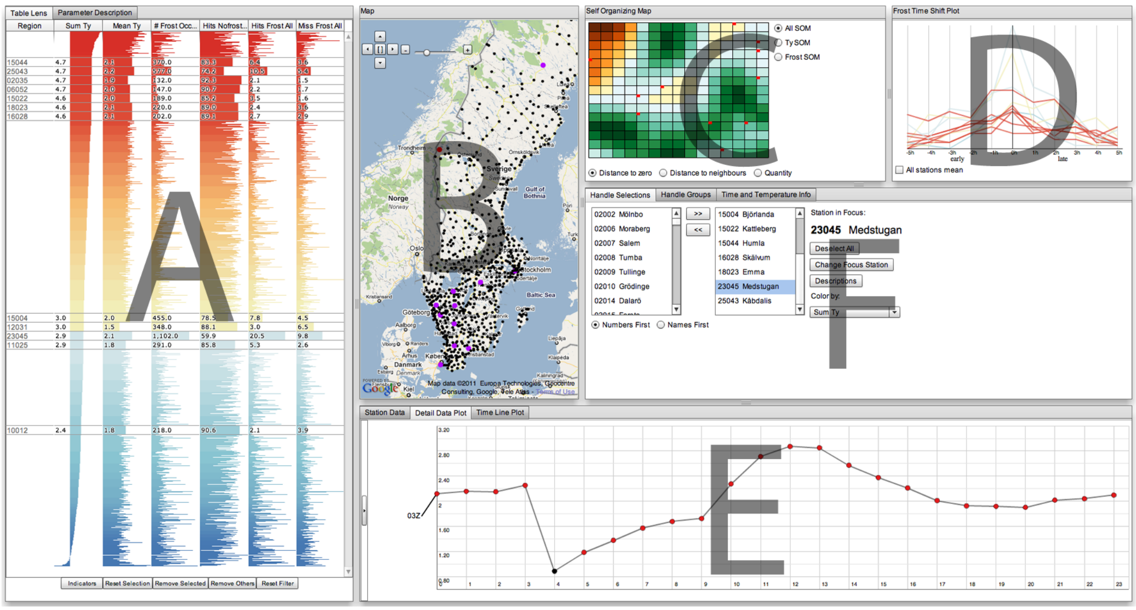


Figure 1: The application is developed to assist the weather forecast verification work at SMHI. Weather data can be examined in different components and on different levels of detail. The application consists of four overview components: table lens (a), glyph map (b), SOM (c) and frost time shift plot (d), and two detail components: detail data plot (e) and time line plot (in the second tab, next to e). In f, three tabs contain extra information about the selected stations.

2 Related Work

Analysis of large, multivariate and spatial data is a very active research area within the information visualization and geo-visualization communitys. In this section we focus on the most related tools that are commonly available for this task.

Several tools for multivariate data analysis are available today, some of which are Protovis [2], Improvise [21], Prefuse [8], GAV Flash [9], InfoVis Toolkit [4], CommonGIS [3], GeoVista Studio [6], VIS-STAMP [7] and CGV [19]. As discussed in Ho et al. [9], it has been a challenge to adapt many of these tools to the Internet. They propose that web-enabled tools are needed for applications with the purpose of communicating analyses of large spatial-temporal and multivariate data. GAV Flash has been adapted for Internet usage through Adobe Flash and is "designed with the intention to significantly shorten the time and effort needed to develop sophisticated and dynamic Web-enabled Geovisual Analytics applications" [5].

RoadVis [14], introduced by Lundblad et al., is a web-enabled application created for road weather visualization. It is a tool for deploying real-time weather forecasts and observations for the winter roads for selected attributes. The tool calculates present and forecasted scenarios for the coming 24 hours and alert the user of risks on the roads, using information visualization components.

Koonar et al. [11] presents a road weather information network built to provide a safer and more sustainable road system in Canada. A network of sensors collect weather data which is used to determine when hazardous road condition occurs. The paper does, however, not describe how the data can be visualized.

Exploratory analysis of multivariate, time-varying and spatial data for weather forecast verification requires specialized visualization and interaction components not readily available in the tools described above. Our proposed solution is influenced by several of these tools and also adopts the concept of coordinated and multiple linked views.

The review of multiple linked view tools, methodologies and models provided by Roberts [15] presents that linking and relating information between views will assist the user's exploration process and provide additional insight to the underlying information. Multiple views can also help stimulate the visual thinking process [13]. Wang Baldonado et al. have developed a set of guidelines for when and how to use multiple views in information visualization [1] to avoid unnecessary complexity of the system. According to Kosara et al. [12] the usage of multiple views on the same data is trivial, but not how to implement it right and to provide it with the necessary interaction.

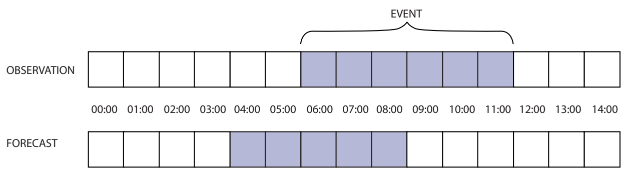


Figure 2: Forecast and observation data is collected for each hour, which is called a time tag. One or more coherent time tags are described as an event. In the figure, two events partly overlap. The forecast event starts two hours before the observation event and ends three hours early.

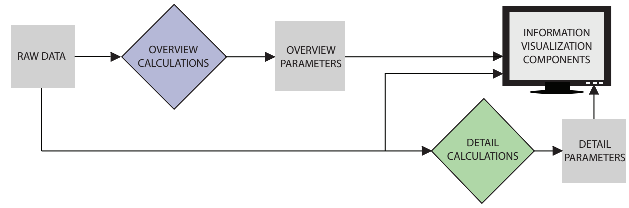


Figure 3: The raw forecast and observation data is used to calculate the overview parameters used as input to the visualization. When a station is in focus, the data for the specified station is read and both shown in its raw format as well as used to calculate detail parameters.

3 Weather Forecasts and Observations

This section describes the data provided by SMHI and the Swedish Transport Administration which is used as input to the application, and introduces the concept of events which is used to describe a snowfall or a frost occasion.

SMHI produces several different forecasts on a daily basis, covering weather information from one hour up to ten days in advance, based on models and calculations. Four times a day, the numerical model HIRLAM (High Resolution Limited Area Model) forecasts the weather for 24 hours [20]. This means that for any given hour there are four forecasts, hereinafter referred to as model runs, predicting a number of weather parameters for that hour. The model runs are called 03Z, 09Z, 15Z and 21Z where the number indicates when (in UTC) the forecast was made.

The Swedish Transport Administration provides SMHI with observations from over 760 weather stations that are located at roads with a high risk of slipperiness [18]. The stations monitor the weather conditions and measure air temperature, surface temperature, air humidity, precipitation, wind speed and wind direction. The forecasts can be compared to the observations to verify the their accuracy.

The majority of the weather stations have an associated environment description. This information can be interesting to examine and compare between different stations since patterns in these descriptions can be found between stations with similar values.

For both the forecasts and the observations, data for a whole winter season is collected at hourly intervals, hereinafter referred to as a time tag. Each time tag contains values for the observation and the four forecast model runs, for all parameters.

3.1 Parameters

The weather parameters focused on in this work are surface temperature, frost and snowfall. Surface temperature and snowfall are both forecasted with calculations from HIRLAM and observed by the weather stations. However, frost can not be measured by the weather stations, but the formation of frost can be calculated from other parameters.

Frost is formed when solid surfaces are cooled to below the dew point temperature of the adjacent air as well as below the freezing point of water (formula 1).

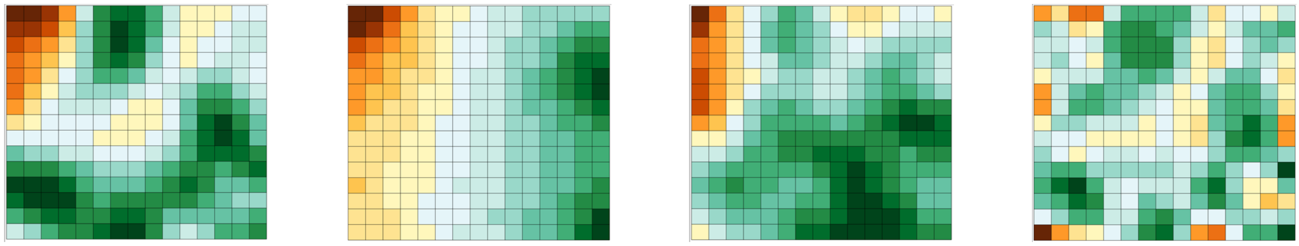
$$\begin{cases} \text{road temperature} < \text{dewpoint temperature} \\ \text{road temperature} < 0^\circ\text{C} \end{cases} \tag{1}$$

The surface temperature is difficult to forecast and can also occasionally be observed incorrectly, e.g. if snow is covering the observation location. In these cases the forecast and observation can differ significantly and this needs to be taken into account when examining the data.

3.2 Event

Measuring differences between temperatures is normally done by subtracting the forecast value from the observation value, giving the difference between them, and thus the error. This method has been used for measuring the errors of the surface temperature forecasts. When verifying parameters such as frost or snowfall, however, different methods need to be used.

One or more coherent time tags are described as an event. If a forecast event and an observation event fully overlap, the forecast is perfect regarding time. However, if they only partly overlap, see figure 2, it is unclear how to measure the accuracy of the forecast. The values can be compared for each time tag and give a hit/miss value that can be used to calculate different hit rates, but they do not indicate whether the forecast was on time or not. A time shift is calculated by comparing the start for the observation and forecast, regarding time. The difference between the start times can be from -5 to 5 hours (chosen as a limit by the domain experts), where a value of zero indicates that the forecast and the observation frost events usually starts at the same time and a negative value represents that the forecast events usually are too early.



(a) All parameters (b) Surface temperature (c) Frost (d) Distance to neighbours

Figure 4: In this application three different SOMs are calculated. In (a) the map is calculated from all parameters, in (b) it is based on the surface temperature parameters and in (c) the frost parameters. In (d) colour is used to convey similarities between neighbouring nodes, from green (small values) to brown (large values).

4 Application

The following section describes the functionality and interactivity of the application (figure 1). The application contains two component categories: overview components and detail components. Multiple stations can be selected in the overview components and the first selected station, hereinafter referred to as the station in focus, is presented in the detail components.

4.1 Pre-processing

The raw input data is pre-processed into overview and detail parameters, see figure 3. The following section gives a short description of the calculated parameter categories.

- The road surface temperature is examined in two ways. Sum Ty indicates if forecasts for a station tend to be too warm or cold and Mean Ty describes the mean error between the forecast and the observation.
- A root mean square value of the frost time shift distribution is calculated to indicate the deviation of the start for the frost event.
- For the frost forecasts accuracy a number of hit rates (hit, miss, false alarm) are calculated.

4.2 Overview Components

The overview components are used to analyze the full winter season and consists of a table lens, a glyph map, a SOM and a frost time shift plot.

4.2.1 Table Lens

A table lens is a graphical spreadsheet that can be used when examining multivariate data sets. It combines overview and detail, can manage both categorical and quantitative data and has the advantage that many people are familiar with tables [17].

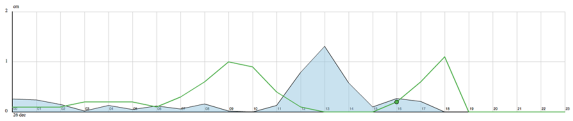
The table lens, figure 1a, is central in the application, rows represent the stations and the columns represent the overview parameters. Using this component, stations with high or low values can be found, the rows can be reordered to find correlations between parameters and a number of interesting stations can be selected for further examination. The user can edit the visible parameters to adjust the view for easier examination. The colours of the stations represent the values for a specified parameter. Which parameter to colour after can be changed, which enables further relationships to be found.

4.2.2 Glyph Map

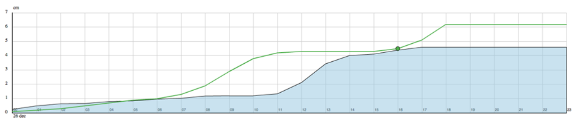
The geographical position of a weather station can affect its values, and to be able to extract geographical patterns the application contains a glyph map, see figure 1b. The map consists of a Google Map with an associated glyph layer, where the weather stations are represented by black glyphs. Selected stations are represented by purple glyphs, and to distinguish the station in focus from the rest it is represented by a red glyph. The map is zoomable and a level of detail function controls the size of the glyphs so that they are visible on all zoom levels without overlapping.

4.2.3 Self-Organizing Map

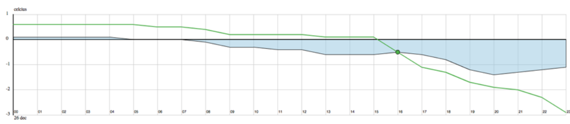
To examine and find patterns in multivariate data, the data is often projected to a lower dimension using reduction methods [7]. A SOM is an artificial neural network that represents a multidimensional data set with a low-dimensional map [10]. The visual representation of the multivariate data becomes easier to understand since the data is represented with spatial clustering which facilitates exploratory analyses [13]. The link between the attribute space visualization tools and maps in multiple views can provide multiple perspectives for exploration, evaluation and interpretation of patterns and ultimately support for knowledge construction.



(a) Snow



(b) Accumulated snow



(c) Surface temperature

Figure 5: The raw data of the forecast and observation is shown in the time line plot. For a specified range of dates the snow (a), accumulated snow (b) or surface temperature (c) can be examined. The green line represents the 15Z model run and the dot indicates the first hour of the forecast. The observed values are represented by the blue surface.

The SOM (figure 1c) consists of 15 x 15 nodes, which can be adjusted if the number of stations change. There are three different maps in this application: all parameters (figure 4a), surface temperature parameters (figure 4b), frost parameters (figure 4c). This enables the user to examine clusters using a specific parameter category. Initially the colours of the nodes represent the distance between the weight vector and a zero vector. The brown nodes are further away from zero than the green nodes. The nodes can also be coloured using the distance between the weight vectors and its neighbours (figure 4d), or coloured by the station quantity in each node. If one or more, but not all, stations in a node are selected the node is marked with a red rectangle in the upper left corner. If all stations in a node are selected it is indicated with a red bounding box.

4.2.4 Frost Time Shift Plot

In the frost time shift plot (figure 1d) each selected station is represented by a line in the plot. The x-axis represents the number of hours the forecast event is shifted in time compared to the observation, from -5 to 5 hours, and the y-axis represents the number of times each case has happened. To match the selected stations with the lines in the plot, the lines have the same colour as the stations in the table lens. The black line represents the mean value of all the stations to indicate the trend of the frost forecast accuracy.

4.3 Detail Components

A detail data plot and a time line plot are used to explore specific scenarios.

4.3.1 Detail Data Plot

The detail data plot distributes the overview parameters over the 24 hours of the day (figure 1e). Each measurement represents a mean value for the specific hour for the whole season. The plot allows the user to further examine the parameters and find if the accuracy of the weather forecasts differ during certain parts or hours of the day.

The parameters can be examined separately in the plot, where the x-axis concern the time and the y-axis the actual values. The model runs can be examined separately or all at the same time. It is also possible to view a line showing the merged values from all the model runs for comparisons between the overall trend and a specified model run.

4.3.2 Time Line Plot

While the overview components can provide the user with an image of what the data looks like, there are times when the data needs to be examined in very high detail. The time line plot was implemented with the intention to facilitate the creation of the plots when the meteorologists need to manually examine the data. The plot is generated by selecting surface temperature, snowfall or accumulated snowfall, and a time span, and displays the raw data (figure 5). The model runs are presented as lines and the observation values are shown as a blue surface. The dots represent the first hour of the forecasts.

4.4 Implementation

The application has been developed using GAV Flash [5] which includes a number of visual components, analytic algorithms, data providers and other tools for developing web based interactive systems. A pre-processing programme was written in C# to process the raw data and calculate the overview parameters. The SOM was calculated in MATLAB using the function package SOM Toolbox [16].

5 Application Scenario

The following is a possible application scenario where a meteorologist explores the data on different levels of detail to find areas where the model may be improved.

The meteorologist starts exploring the data in the table lens and selects the station with the highest mean surface temperature error, Mean Ty (figure 6, left), to examine possible reasons to why the forecasts for this station are less accurate than for other stations. The geographical position of this station can be seen in the glyph map (figure 6, middle). To find stations with similar values for the surface

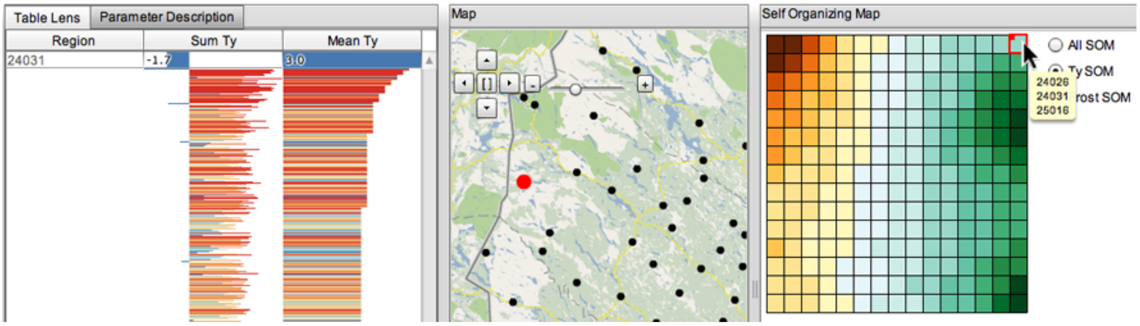


Figure 6: When a station is selected in the table lens (left) it is also selected in the glyph map (middle) and the SOM (right). The cell containing the selected station is marked in the SOM, which contains two additional stations with similar values for the surface temperature parameters.

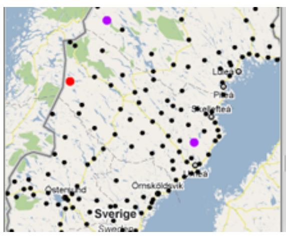


Figure 7: Three selected stations in the glyph map. The red glyph represents the station in focus.

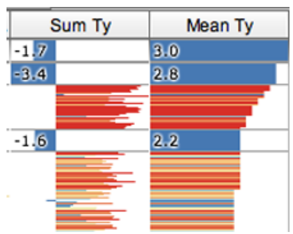
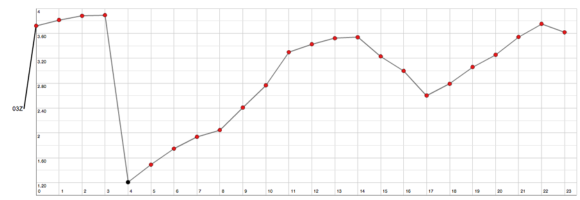
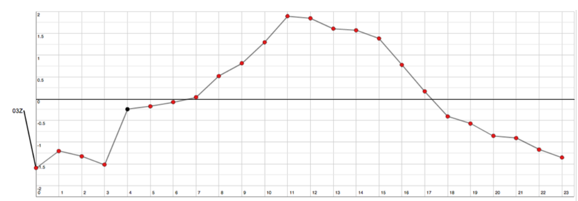


Figure 8: Detailed information about the three selected stations in the table lens.



(a) Mean Ty - the mean surface temperature error.



(b) Sum Ty - the sum of the surface temperature error.

Figure 9: The station in focus in the detail data plot, showing the overview parameters distributed over the 24 hours of the day.

temperature parameters, the user selects the node belonging to the selected station in the Ty SOM (figure 6, right). These stations are now also selected in the glyph map (figure 7) and since all three stations are located in the northern part of Sweden the user can suspect a possible correlation between geographical position and a high value for Mean Ty. In the glyph map, the station in focus is represented by a red glyph and the other selected stations are represented by purple glyphs. The three selected stations can also be further examined in the table lens (figure 8) where the user finds that these stations also have very low values for the parameter Sum Ty, which indicate that the forecasts for these stations are usually too cold.

The detail data plot is used to examine the surface temperature for a station on a detailed level. In the plot for Mean Ty (figure 9a), the user can see that the 03Z forecast is better during the first hours from when the forecast was made and also better during the afternoon. In the plot for Sum Ty (figure 9b), the user can see that the forecasts are usually too cold during the night and too warm during the day. These conclusions can be subject to further analysis.

To see if the found results have any correlation with other parameters, the user examines the starts of the frost events in the frost time shift plot (figure 10). The three selected stations are represented with three lines in the plot and the user can draw conclusions whether the frost forecasts are early or late.

Another examination possibility could be if the user knew that the forecasts had difficulties during a particular time period. This period can be selected in the time line plot where the raw data is graphically represented and the user can study the data in very high detail. Figure 11 shows the snow forecast for the 2009 Christmas Day and reveals that the 03Z forecast was approximately correct in amount but incorrect with respect of time.

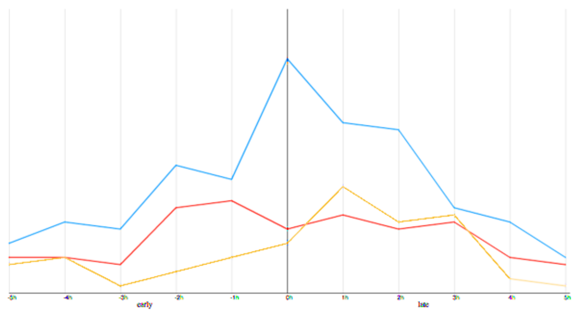


Figure 10: The distribution of occurrences of frost forecasts that are early or late, from -5 to 5 hours, compared to the observations. The station with a peak at 0 (blue line), indicates that the forecasts often are on time, while the station with a peak at +1 (yellow line), indicates that the forecasts often are one hour late. For the station represented by the red line a significant peak is missing and thus the forecasts are usually both early and late.

6 User Study

An informal user study was performed as a first attempt to evaluate the application. The main purpose was to get feedback on the functionality, interface, and the overall navigation of the application.

6.1 Procedure

The user study was conducted with four master students and two meteorologists. The students were familiar with information visualization, but had no previous experience with meteorological data. The meteorologists had domain knowledge but were inexperienced with visualization.

According to Koua and Kraak [13] a user-based evaluation is the most suitable approach to assess usability and usefulness of geovisual environments. Their presented performance-oriented approach was however replaced with a more satisfaction-oriented test. Since no one in the test group was familiar with both the weather data and the visualization techniques, the focus lay on whether the components and user interface felt intuitive and easy to use.

To begin with, the application was demonstrated and explained, and the users were provided with a parameter and component description. Two test leaders were present to answer questions and document the result. The users were asked to complete a number of tasks, which were designed to encourage them to try all the visualization components.

Some representative tasks were as follows:

- ”*Identify one station with more than 800 frost hours that is not located in the northern part of Sweden.*”
- ”*Is there any correlation between the nodes in the SOM and the geographical position of stations?*”

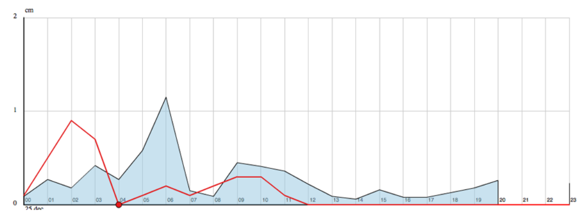


Figure 11: The time line plot shows the 03Z snow forecast for Christmas Day in 2009. The forecast is approximately correct in amount but incorrect with respect of time.

- ”*During Christmas weekend when was there most precipitation for the station MS4 SMHI?*”

When the tasks were completed the test ended with a discussion about the overall experience of using the applications, which tasks had been difficult to complete and why, and what improvements could be made.

6.2 Result

Most of the tasks were completed by the test group with little or no difficulty and the general opinion was that the applications feels intuitive and easy to use after the given introduction. The test group pointed out suggestions for concept design improvements and changes in the layout.

In the time line plot, one of the participants wanted to be able to adjust the time span with a slider in addition to the calendars. This would make the creation of weather data plots easier and more intuitive, and would help improve the interactivity and efficiency of the component.

Some of the participants confused the overview and detail parameters and tried to change the parameter for the detail data plot component in the table lens.

Since many stations are represented in the table lens it is hard to examine them in a time-efficient manner, but as one of the participants pointed out a fish-eye feature could make the searching in the table lens more efficient.

7 Conclusions and Future Work

In this paper we describe a web-based interactive application for exploration of weather data, focusing on verification and analysis of road weather conditions. The application has been developed together with meteorologists from the Swedish Meteorological and Hydrological Institute. It was designed to simplify the process of improving weather forecasts models and to identify inaccurate weather stations.

The results from an informal user study with meteorologists showed that the application has great potential to help them in their verification work. The use of interactive visualization was a new way of working with weather data verification which was seen as a good compliment to the traditional manual methods currently used.

The presented application was designed as a proof of concept to demonstrate interactive visualization and its possibility to assist the process of finding patterns and relationships in weather data. Possible future work includes investigating the usability of more advanced information visualization components as well as further evaluation of the existing application. Furthermore, the user study showed that additional interaction, allowing even more detailed exploration, were highly desired by the meteorologists.

Acknowledgements

This work was partly supported by the Swedish Research Council in the Linnaeus Centre CADICS and by funding from the Visualization Programme coordinated by the Swedish Knowledge Foundation. The authors wishes to thank the involved meteorologists at SMHI for their feedback during the project and their dedicated contribution that made this project possible.

References

[1] M. Q. Wang Baldonado, A. Woodruff, and A. Kuchinsky. Guidelines for using multiple views in information visualization. In *Proceedings of the working conference on Advanced visual interface*, 2000.

[2] M. Bostock and J. Heer. Protovis: A graphical toolkit for visualization. *IEEE Transactions on Visualization and Computer Graphics*, 15:1121–1128, 2009.

[3] CommonGis. url: http://www.commongis.de.

[4] J.-D. Fekete. The infovis toolkit. In *Third International Conference on Information Technology and Applications*, pages 167–174, 2005.

[5] GAV Flash. url: http://ncva.itn.liu.se/tools.

[6] Geovista. url: http://www.geovista.psu.edu.

[7] D. Guo, J. Chen, A. M. MacEachren, and K. Liao. A visualization system for space-time and multivariate patterns (vis-stamp). *IEEE Transactions on Visualization and Computer Graphics*, 12(6):1461–1474, 2006.

[8] J. Heer. Prefuse: a toolkit for interactive information visualization. In *In CHI 05: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 421–430. ACM Press, 2005.

[9] Q. Ho, P. Lundblad, T. Åström, and M. Jern. A web-enabled visualization toolkit for geovisual analytics. In *SPIE: Electronic Imaging Science and Technology, Visualization and Data Analysis, Proceedings of SPIE, San Francisco*, 2011.

[10] T. Kohonen. *Self-Organizing Maps*. Springer; 3rd edition, 2000.

[11] A. Koonar, P. Delannoy, and D. Denault. Building a road weather information network for integrating data from heterogeneous sources. In *Third International Conference on Information Technology and Applications*, pages 501–507, 2005.

[12] R. Kosara, H. Hauser, and D. L. Gresh. An interaction view on information visualization. In *Eurographics*, pages 123–138, 2003.

[13] E. L. Koua and M-J Kraak. A usability framework for the design and evaluation of an exploratory geovisualization environment. In *Proceedings of the Eighth International Conference on Information Visualisation*, pages 153–158, 2004.

[14] P. Lundblad, J. Thoursie, and M. Jern. Swedish road weather visualization. In *Information Visualisation, International Conference on*, volume 0, pages 313–321, Los Alamitos, CA, USA, 2010. IEEE Computer Society.

[15] J. C. Roberts. Exploratory visualization with multiple linked views. In *Exploring Geovisualization*, pages 150–170. Amsterdam: Elseviers, 2004.

[16] Self-Organizing Map Toolbox for Matlab. url: http://www.cis.hut.fi/somtoolbox.

[17] R. Spence. *Information visualization design for interaction*. Prentice Hall, Harlow, second edition, 2007.

[18] The Swedish Road Administration's brochure about Road Weather Information Systems (VViS). url: http://publikationswebbutik.vv.se/upload/5455/89370_vagvaderinformationssystem_vvis.pdf.

[19] C. Tominski, J. Abello, and H. Schumann. CGV - An Interactive Graph Visualization System. *Computers & Graphics*, 33(6):660–678, 2009.

[20] P. Unden, L. Rontu, H. Järvinen, P. Lynch, J. Calvo, G. Cats, J. Cuxart, K. Eerola, C. Fortelius, J. A. Garcia-Moya, C. Jones, G. Lenderlink, A. Mcdonald, R. Mcgrath, B. Navascues, N. Woetman Nielsen, V. Degaard, E. Rodriguez, M. Rummukainen, K. Sattler, B. Hansen Sass, H. Savijarvi, B. Wichers Schreur, R. Sigg, and H. The. Hirlam-5 scientific documentation. Technical report, 2002.

[21] C. Weaver. Building highly-coordinated visualizations in improvise. *Information Visualization, IEEE Symposium on*, 0:159–166, 2004.