# Decision Analytics with Heatmap Visualization for Multi-step Ensemble Data

## An Application of Uncertainty Modeling to Historical Consistent Neural Network and Other Forecasts

With today's computing power, it is easy to generate huge amounts of data. The real challenge lies in adequately condensing the data in decision making processes. Here, the focus is on ensemble data that typically arises when distributions of forecasts are generated for several time steps in the future. Often a distribution is aggregated by taking an ensemble's mean or median. This results in a single line that is easy to interpret. However, this single line may be seriously misleading when the ensemble splits into two or more different bundles. The mean or median may also lie in a region where there are only very few ensemble members. To remedy this, a heatmap visualization to better represent ensemble data for decision analytics is proposed. Heatmap visualization provides an intuitive way to identify regions of high and low activity. The regions are color-coded according to the (weighted) number of ensemble members in a specific region.

DOI 10.1007/s12599-014-0326-4

### The Authors

**Dipl.-Math. Cornelius Köpp**
**Prof. Dr. Hans-Jörg von Mettenheim (✉)**
**Prof. Dr. Michael H. Breitner**
Institute for Business Information Systems
Leibniz University Hannover
Königsworther Platz 1
30167 Hannover
Germany
mettenheim@iwi.uni-hannover.de

Received: 2013-07-15
Accepted: 2014-01-31
Accepted after two revisions by the editors of the special focus.
Published online: 2014-04-26

This article is also available in German in print and via http://www.wirtschaftsinformatik.de: Köpp C, von Mettenheim H-J, Breitner MH (2014) Decision Analytics mit Heatmap-Visualisierung von mehrschrittigen Ensembledaten. Eine Anwendung von Unsicherheitsmodellierung für Historical Consistent Neural Network und andere Prognosetechniken. WIRTSCHAFTSINFORMATIK. doi: 10.1007/s11576-014-0417-3.

**Electronic Supplementary Material**
The online version of this article (doi: 10.1007/s12599-014-0326-4) contains supplementary material, which is available to authorized users.



## 1 Introduction and Motivation

One of the typical tasks of a decision support system is to produce forecasts and – more importantly – to help people interpret those forecasts. However, single value forecasts may be misleading. For this reason ensemble forecasts – based on a collection of several individual forecasts – can be used to largely improve the forecasting accuracy; see for example (Zhang and Berardi 2001). With ensemble forecasts, quite often only a single value is (or a few points of the distribution are) used. This is useful for automatic usage in information systems but the real shape of the other features of the distribution remains unclear; see for example (Welch 2001; Hansen 2008). It is not necessarily the case that the forecast distribution is unimodal. Generally it could be multimodal. That means that we cannot characterize the forecast by a single number (for example the mean). Several different models make up the forecast and we cannot expect our forecast to have a single peak. In this case the mean or median can be misleading, as we will show. Multi-step means that our forecast does not just forecast, for example, tomorrow's value. Rather, the forecasts we are looking at will stretch over several time-steps. In essence, common data aggregation techniques either lose information or do not scale well. In both cases this greatly reduces the usefulness of the forecast.

We present first steps towards answering the research question: "*How can an adequate visualization enable decision analytics for today's ensemble forecast methods?*"

Ensembles of artificial neural network (ANN) models are a typical case where we obtain forecasts that consist of several hundred individual paths. In the present paper we look at a 20-day forecast for the price of natural gas in US dollars. The figures are computed from an ensemble of 200 networks. We are using a new class of ANN, the Historical Consistent Neural Network (HCNN) introduced by Zimmermann et al. (2010). See also von Mettenheim and Breitner (2010) for a detailed presentation and performance evaluation. For the goals of the present paper it is sufficient to be aware of the fact that HCNNs use a simple state equation to compute the following state from the immediately preceding state. Multi-step forecasts are therefore easy to generate. We generally use HCNNs when we have to model several time-series and their distribution simultaneously. When we train different HCNNs with randomly initialized weights, we obtain a diverse ensemble of forecasts.

The exact forecast asset is not central to the following discussion and it is not our goal to evaluate the forecast performance of the ANN. Rather, our focus is on making the forecast output easier to interpret for the human decision maker. Our approach is therefore not limited to ANNs. Other model types that might produce an unlimited number of forecast paths based on good historical performance could also be used in this context. These include, for example, Support Vector Machines, Evolutionary Programming, and Monte Carlo Simulations. For this reason we outline the ANN model only briefly. In our paper we present a heatmap visualization of the resulting ensemble forecasts. This is a step towards visually supporting the human decision maker, because heatmaps aggregate information but conserve the essence of the forecast, even if the distribution is multimodal.

We can now make our research question more concrete: *"How can we intuitively present the complete forecast information to a decision maker, but also exploit all distribution information?"*

We propose heatmap visualization. A heatmap allows us to differentiate between more active and less active regions of the forecast space by color coding (see **Figs. 1**(**d**)–**1**(**f**) and examples in **Figs. 3**, **4**, and **5**). A detailed discussion of **Fig. 1** will follow below, especially in Sect. 6.

Whereas simply plotting aggregate values in **Figs. 1**(**a**) and **1**(**b**) loses information, it is also not possible to only plot each forecast individually as is done in **Fig. 1**(**c**). We cannot distinguish individual forecasts anymore and the output is useless. **Figures 1**(**d**) and **1**(**e**) show our heatmap approach and **Fig. 1**(**f**) is an even more useful presentation. We can clearly see splitting paths in **Figs. 1**(**d**)–**1**(**f**). This is a warning signal from the model: the forecast is dubious.

Heatmaps also help us gauge the quality of the forecast. Depending on the width of the forecast and the number of peaks we see in the heatmap, we can qualify the forecast as more or less reliable This offers an alternative to the usual binary output of most of today's forecasting methods. It is quite common that a forecast model either outputs "up" or "down". This is not entirely honest. There should be a third output possibility: "don't know". Heatmaps offer just that: rather than hiding information behind a single number (which will invariably be wrong) they present the entire forecast spectrum to the decision maker. Problematic areas are then easy to identify.

## 2 Research Design

To answer the research question, we were inspired by the Design Science Research approach from Hevner et al. (2004, p. 83). **Figure 2** shows the implementation of our research design. In the following we place our contribution in the context of selected research guidelines:

The presented research is relevant (*Problem Relevance*), as our visualization approach tackles distributional forecasts that for example arise in the context of ANN ensembles. The same visualization could be useful for any other forecast models that generate ensembles, or time-series. We review existing concepts in the area of ensemble and time-series visual representation and show the limitations of commonly used visualization techniques in Sect. 3. In Sect. 4 we present the formal model of our method for heatmap generation. We argue that this approach is an artifact that "extends the boundaries of human problem solving" (Hevner et al. 2004) because it can help decision makers arrive at better decisions by providing additional information, which is invisible when using statistic aggregations (*Design as an Artifact*).

The first step of our evaluation is to build a prototype implementation. The prototype is described in Sect. 5. We use the prototype to demonstrate the utility of our approach, in the form of qualitative information gain. One example scenario is given by the above-mentioned gas-price forecast scenario; another is artificially created for clearer description of the information gained by interpolation,following the concept of Descriptive *Design Evaluation* with Scenarios (Hevner et al. 2004, p. 86).

An earlier stage of this work was presented to the scientific and practice audience at a conference (von Mettenheim et al. 2012) and a workshop (*Communication of Research*). We discussed the results with ANN experts and incorporated their feedback. From this discussion an independent implementation of our approach emerged, conducted by expert users in a large international company. Hevner et al. (2004) states that "the objective of design-science research is to develop technology-based solutions to important and relevant business problems". We consider the existence of an independent implementation to be a strong indicator for the importance and business *relevance* of the problem we address with our approach.

## 3 Related Work

Potter et al. (2009b) underline the "enormous power" of ensemble data sets, but also the "formidable challenge" of ensemble visualization due to their complexity. Andrienko and Andrienko (2005) focus on spatially distributed time-series data as used in cartographic and geovisualization applications. They criticize the combined plotting of many localized time-series, as cluttering and overlapping lines result in a hardly legible display, and the concept becomes "completely unusable" for a large number of hundreds and more time-series. As an alternative they propose map- and aggregation-based visualization. Aggregation is realized by plotting the minimum, maximum, median and quartiles. Andrienko et al. (2010) focus on event detection support in multiple time-series. They present a toolkit with interactive user-controlled data visualization, using mean or median for statistical summary of multiple time-series. Bade et al. (2004) use minima, maxima, median and the 25 % and 75 % percentile for the presentation of aggregated high-frequency data streams. Hao et al. (2009) describe a visual support framework for time-series prediction. Their tool uses a one dimensional heatmap style "visual accuracy indicator" to show over, under, and close predictions.

Aigner et al. (2007) present a "conceptual visual analytics framework for time-oriented data". They describe aspects