Color Coding of Large Value Ranges Applied to Meteorological Data

Daniel Braun*  
University of Cologne  
Department of Mathematics and Computer Science

Kerstin Ebell†  
University of Cologne  
Institute for Geophysics and Meteorology

Vera Schemann‡  
University of Cologne  
Institute for Geophysics and Meteorology

Laura Pelchmann§  
University of Cologne  
Department of Mathematics and Computer Science

Susanne Crewell¶  
University of Cologne  
Institute for Geophysics and Meteorology

Rita Borgo||  
King’s College London  
Department of Computer Science

Tatiana von Landesberger**  
University of Cologne  
Department of Mathematics and Computer Science

1 INTRODUCTION

Data with large value ranges are data sets whose values contain several different exponents using the scientific notation \(v = m \cdot 10^e\) [8]. One example is meteorological data on ice water content in clouds, whose exponents vary between minus eight and minus two. The data includes three variables: time, height and ice water content. Mapping the values to position is not possible in a 2D-visualization. The current meteorological standard is to visualize the data using a scatterplot with the values encoded by a logarithmic scaled colormap [5, 9]. The meteorologists call this time-height series. The challenge is to encode such a large value range on a color scale and a limited number of pixels without loss of information.

The common colormaps used for meteorological scatterplots are Viridis (Fig. 1a) [5] and Rainbow (Fig. 1b) [9]. We present a novel color scheme to visualize data with large value ranges: the order of magnitude colors (OMC), which uses a separate representation of the mantissa and exponent of the data (Fig. 1d) for an easier classification of the orders of magnitude. It uses one hue for each exponent and linear gradient of brightness within for magnitude. Additionally, a variation of this design is shown (Fig. 1e).

We compare our approach with the currently used color scales as well as a state-of-the-art color scheme generated via the tool ColorCrafter [20] (starting color: blue) (Fig. 1c) in an empirical user study. The results show that our new color scheme significantly outperforms the others in interpretation tasks, while showing comparable performances in discrimination tasks. We already received requests from meteorologists for the use of our design.

2 RELATED WORK

Color Perception  
Golebiowska and Coltekin [6] explored the perceptual differences from rainbow to sequential color scales. Their
We present a novel color scheme named order of magnitude colors. This design is inspired by the scientific notation of numbers and is created in the CCC-tool [12]. Inspired by the idea of the recent approaches [2, 7, 8], we use the two parts mantissa $m$ and exponent $e$ of a value $v$ (so that $v = m \cdot 10^e$) for color coding. Every exponent given in the data is mapped to another hue. Within an exponent, the mantissa is mapped to a perceptually linear sequential scale of the respective hue.

During the research process, we created and tried many different color scales to the meteorological data (see the supplementary material). The color selection as well as the data and the study tasks were discussed and agreed upon in several focus groups organized together with the meteorologists. In the final design, the colors were chosen through a manual process so that the color distances between adjacent colors are approximately even. Color blindness would be an additional criteria for further research. We made the HSV-gradient for the different hues as smooth and equal as possible (see Fig. 2).

To achieve a better color nameability [15, 16], we chose unique hues instead of different saturation tones of one specific hue.

Boundaries in colors are easily perceived as changes in the exponents in the data [17]. This is deemed appropriate, as our colormap describes the changes in magnitudes. It helps the meteorologists looking at changes in the data. Nevertheless, we also created a variation of the OMC-colormap that reduces the color distances (i.e. smaller DeltaE-values [11]) (see Fig. 2 and Fig. 3). In this version, we flipped every second sequential color scale so that we have a changing lightness direction comparing two neighboring color scales: The OMC smoothed lightness (Fig. 3).

4 Evaluation

Viridis and Rainbow are common color schemes for visualizing meteorological data. Therefore, we evaluate these two, a state-of-the-art colormap created by ColorCrafter [20] and our two new designs - the OMC and OMCsd - in an exploratory user study.

Data

The data used for the study were real meteorological data sets provided by the Institute for Geophysics and Meteorology at the University of Cologne. Every set of data contains one measurement day with the variables time (0 to 24 hours), height (0 to 12 kilometers) and the ice water content in clouds. The ice water content spans a range of $10^{-9}$ to $10^{-2}$. Due to the large value ranges, the colormaps were scaled logarithmic.

To avoid a learning effect, each question of the study contained visualizations of different measurement data. For consistency, we used data sets with similar properties, e.g., an equal value range.

Tasks

In line with literature and after consultation with the meteorologists, we decided to focus on four main tasks, divided into reading [1, 3, 13, 18, 22] and comparison [3, 13, 18, 22] tasks:

- **Reading Tasks:**
  - **Extrema:** name the maximum and minimum exponent of a given day
  - **Value:** specify a value range for the marked region of data points
- **Comparison Tasks:**
All participants had to solve the same tasks. There was one task for each colormap have the same level of difficulty. The comparison tasks are single choice, multiple options questions. In the reading tasks, the participants had to insert their answered numbers manually. Additionally, we asked the participants about their confidence in the given answers of the reading tasks.

Experimental Setting and Procedure A total of 53 participants (36 male, 13 female and 4 prefer not to say) took part in the study. The age was distributed from 20 up to 60, but the majority of the participants were between 20 and 40 years (88%). We filtered out 5 participants, who did not correctly answer to the Ishihara tests for color blindness [4] (this was necessary because the rainbow scale is not suitable for color blindness).

The expertise of the participants was broad and ranged from a degree in mathematics (25%) to a degree in physics (13%) to a degree in computer science (10%) and others (4%), with the majority of the participants having a degree in meteorology (48%).

The study was conducted online and was set up with LimeSurvey [10]. There was one task per page and the participants had to click on a button manually to get to the next page, so we could store the given answers and the response time per task. The visualizations used had a resolution of 1291×500px. All of the participants used a computer screen with a size of at least 13”.

The processing time of the study was approximately 30 minutes. All participants had to solve the same tasks. There was one task for every of the five designs for every type of task, i.e. in total we had 5 [colorschemes] × 4 [tasks] = 20 trials.

After a short training phase introducing the visualization type and the types of tasks, the sorting of the tasks was randomized to reduce a learning effect. At the end of the study the participants were asked to give feedback. Study documentation is in annex.

4.1 Analysis
The analysis of the results is performed in R [14]. For every task of the study, we measure accuracy and response time. In the reading tasks, the specified confidence of the participants is also examined (on a Likert scale from 1 – very unsure to 5 – very confident).

Correctness measurement For every task, we compare the number of correct answers to the number of incorrect answers. In the extreme reading task, both exponents have to be correct to be considered as a correct answer. For value reading an answer is correct, if the values of the marked region lie in the answered range.

In addition, the sizes of the specified ranges in the reading value task are compared. To exclude logarithmic scaling, the size of the ranges is calculated using Equation 1:

\[
\text{range size} = \frac{(\text{Exp}_{\text{max}} - \text{Exp}_{\text{min}}) \cdot 10 + (\text{Mant}_{\text{max}} - \text{Mant}_{\text{min}})}{10}
\] (1)

Significance Tests To perform our analysis, a three-stage significant test for each task was used. Since we could not assume that the data is normally distributed, we first ran a Shapiro-Wilk test on the given answers and response times, with the result that none of the data is normally distributed.

In the second stage of the analysis, we used a \( \chi^2 \)-test (and a Fisher-test if the frequencies are too low) to investigate significance in the amounts of correct and wrong answers. For the quantitative data like the range sizes or response time, we used the Kruskal-Wallis test.

In the third stage, as post-hoc analysis we used a Wilcoxon-test respectively a \( \chi^2 \)-test for pairwise comparison of the color schemes for tasks for which significance was found.

All tests were performed with the significance level \( \alpha = 0.05 \).

- Extrema: compare two measurement days and decide which one contains the global extremum
- Value: compare the values of two marked regions of data points in one day

We manually selected the regions to ensure that the tasks for each colormap have the same level of difficulty. The comparison tasks are single choice, multiple options questions. In the reading tasks, the participants had to insert their answered numbers manually. Additionally, we asked the participants about their confidence in the given answers of the reading tasks.

4.2 Results
Fig. 4 shows a summary of the results including the percentages of correct answers for the different colormaps in the four tasks and the response time of the participants.

Reading Extrema Our OMC design (71% correct answers) and the rainbow color scheme (69%) perform best. The \( \chi^2 \)-test shows a significant dependence between the color scales and the amount of correct/ false answers (\( \chi^2 = 52.466, \text{p-value} = 1.102e-10 \)). The post-hoc pairwise test confirms that OMC and Rainbow deliver significantly better results than ColorCrafter (33%) and Viridis (10%) (Table 1). OMC\(_{sl}\) is third best with 60% of correct answers.

Reading Values The OMC designs perform best with respect to the number of correct answers. The \( \chi^2 \)-test indicates significance (\( \chi^2 = 19.833, \text{p-value} = 0.001 \)). Due to the pairwise test, only the Viridis colormap is significantly worse than other color schemes.

The positive effect of the OMC colormap can be seen better in the comparison of the given range sizes (Fig. 5). The OMC colormap has by far the smallest range size (mean = 0.36), followed by OMC\(_{sl}\) (0.68). Viridis (0.91), ColorCrafter (0.99) and Rainbow (1.15) are at a similar level. The Kruskal-Wallis-test confirms the significant differences in the means (\( \chi^2 = 50.518, \text{p-value} = 2.814e-10 \)). The pairwise Wilcoxon-test shows, that the OMC color scale leads to significantly smaller range sizes (i.e. a better identification of the searched values) than all others (Table 2). OMC\(_{sl}\) also performs significantly better than ColorCrafter and Rainbow. This can be attributed to our design approach, where different hues enable fast recognizing of magnitudes.

Comparison of Extrema Very few errors were made. ColorCrafter (3 errors) and OMC\(_{sl}\) (3) have the most errors, followed by Rainbow (2) and OMC (1). Viridis even produced no error at all. Due to the low number of errors, no significant dependence can be determined using Fisher’s exact test (p-value = 0.471).

Comparison of Values ColorCrafter performs worst for this task (58% correct answers). OMC\(_{sl}\) has 94% of correct answers, while the other three even have 98%. Due to the \( \chi^2 \)-test there is a significant effect (\( \chi^2 = 66.679, \text{p-value} = 1.332e-9 \)). The post-hoc pairwise Fisher-test shows that ColorCrafter leads to significantly more errors than all other color scales (Table 3).

4.3 Extended Analysis and Results
Confidence For the reading extreme task, the analysis of the participants’ confidence shows that the sequential color schemes (Viridis, ColorCrafter) lead to significantly more uncertainty in the answers than the multicolored schemes (OMC, OMC\(_{sl}\), Rainbow).

For reading value task, the participants were significantly more confident with our OMC color scale than with all others, both for the exponents and the mantissas of the specified range.
Figure 4: Percentages of correct answers (the more the better) and response times in seconds (the less the better) for tasks and colormaps.

Figure 5: Sizes of the value ranges answered for value reading.

Table 3: p-values of the pairwise Fisher-test for value comparison.

| p-value  | OMC   | OMC_{sl} | ColorCrafter | Rainbow |
|----------|-------|----------|--------------|---------|
| OMC_{sl} | 0.617 | -        | -            | -       |
| ColorCrafter | 0.000 | 0.000   | -            | -       |
| Rainbow  | 1.000 | 0.617   | 0.000        | -       |
| Viridis  | 1.000 | 0.617   | 0.000        | 1.000   |

Table 3: p-values of the pairwise Fisher-test for value comparison.

Time Only for the value comparison task, the Kruskal-Wallis-test shows a significant mean effect in the response time ($\chi^2 = 13.398$, p-value = 0.009). As the pairwise Wilcoxon-test shows, participants used significantly more time with ColorCrafter than with all others. Notably, the participants did not need any additional time to understand our new order of magnitude colors design.

Free Text 31 participants gave free text feedback. Almost all of them were positive regarding our OMC color scale. We already received requests to use our design. It was stated that the color scheme supports value comparison and identification by its clear borders: "The color scale with different colors between every power of 10 is the one that allowed me to better identify the values and the differences between points." Suggestions for improvement were the color selection and the perception of qualitative patterns.

The sequential colormaps were perceived as more aesthetically pleasing. Rainbow and ColorCrafter got the most negative feedback, especially for their color gradients.

Expertise Comparing the results of participants with a meteorological background to the others, there was no significant difference in the correctness of their answers and in the response time. This indicates that our color scheme can be used broadly, without specific expertise or background.

Limitations There are some limitations to our color scheme. The approach is not suitable for comparing values that are close to the borders of the exponents or for data containing positive and negative values. Additionally, further experiment may be needed to assess the scalability of the proposed scheme for value ranges with different exponent distributions. In order not to make the user study too extensive, we focused on a subset of tasks most relevant to our collaborators. Therefore, some types of task could not be dealt with sufficiently. For example, an evaluation of the OMC colormap for high level tasks would be interesting.

5 Conclusion and Future Work

We presented a new color coding approach to visualize data featuring large value ranges with the application to meteorological data. The empirical study has shown that our order of magnitude colors design performs very well for all tasks examined. It has no significant difference to the best performing colormaps in the comparison tasks and improves the accuracy of value identification significantly. The strict borders of our color scheme support the perception of the different orders of magnitude.

In summary, our results suggest the following ranking for each proposed design:

- **Reading Extrema**: OMC > Rainbow > OMC_{sl} > ColorCrafter > Viridis
- **Reading Values**: OMC > OMC_{sl} > Viridis > ColorCrafter > Rainbow
- **Comparison of Extrema**: Viridis > OMC > Rainbow > OMC_{sl} ~ ColorCrafter
- **Comparison of Values**: OMC ~ Viridis ~ Rainbow ~ OMC_{sl} > ColorCrafter

In general, our study results confirm findings from Golebiowska and Coltekin [6] that multi-colored scales like OMC and Rainbow are more suitable for reading tasks while sequential scales like Viridis work better for comparing tasks.

Our design was developed on the application example of multi-dimensional meteorological cloud data. Our evaluation was data agnostic making our results extendable to other data sets. The choice of color scheme however was informed by our collaboration with domain expert therefore, in the future, we would like to test the performance of our approach on other data types from various application areas, potentially exploring new color schemes. Furthermore, we have to investigate color blindness for our new design.

Acknowledgments

The authors would like to thank all study participants and the reviewers, whose suggestions helped improve this paper.
REFERENCES

[1] R. Amar, J. Eagan, and J. Stasko. Low-level components of analytic activity in information visualization. *IEEE Symposium on Information Visualization*, 2005. INFOVIS 2005, pp. 111–117, 2005. doi: 10.1109/INFOVIS.2005.1532136

[2] R. Borgo, J. Dearden, and M. W. Jones. Order of magnitude markers: An empirical study on large magnitude number detection. *IEEE transactions on visualization and computer graphics*, 20(12):2261–2270, 2014. doi: 10.1109/TVCG.2014.2346428

[3] M. Brehmer and T. Munzner. A multi-level typology of abstract visualization tasks. *IEEE transactions on visualization and computer graphics*, 19(12):2376–2385, 2013. doi: 10.1109/TVCG.2013.124

[4] J. H. Clark. The ishihara test for color blindness. *American Journal of Physiological Optics*, 5:269–276, 1924.

[5] Finnish Meteorological Institute. Cloudnet data portal. https://cloudnet.fmi.fi/search/visualizations.

[6] I. Golebiowska and A. Coltekin. Rainbow dash: Intuitiveness, interpretability and memorability of the rainbow color scheme in visualization. *IEEE transactions on visualization and computer graphics*, PP, 2020. doi: 10.1109/TVCG.2020.3035823

[7] M. Hlawatsch, F. Sadlo, M. Burch, and D. Weiskopf. Scale-Stack Bar Charts. *Computer Graphics Forum*, 32(3pt2):181–190, 2013. doi: 10.1111/cgf.12105

[8] M. Höhn, M. Wundertlich, K. Ballweg, and T. von Landesberger. Width-scale bar charts for data with large value range. *EuroVis 2020 - Short Papers*, 2020. doi: 10.2312/evs.20201056

[9] Institute for Geophysics and Meteorology, University of Cologne. Measurements and visualizations of meteorological cloud data. https://atmos.meteo.uni-koeln.de/~hatpro/dataBrowser/.

[10] LimeSurvey Project Team / Carsten Schmitz. Limesurvey: An open source survey tool. http://www.limesurvey.org, 2022.

[11] R. McDonald and K. J. Smith. Cie94—a new colour-difference formula*. *Journal of the Society of Dyers and Colourists*, 111(12):376–379, 1995. doi: 10.1111/j.1478-4408.1995.TB01688.X

[12] P. Nardini, M. Chen, M. Böttinger, G. Scheuermann, and R. Bujack. Automatic improvement of continuous colormaps in euclidean colorspaces. *Computer Graphics Forum*, 40(3):361–373, 2021. doi: 10.1111/cgf.14313

[13] G. J. Quadri and P. Rosen. A survey of perception-based visualization studies by task. *IEEE transactions on visualization and computer graphics*, PP, 2021. doi: 10.1109/TVCG.2021.3098240

[14] R Core Team. R: A language and environment for statistical computing. https://www.R-project.org/, 2021.

[15] K. Reda, A. A. Salvi, J. Gray, and M. E. Papka. Color nameability predicts inference accuracy in spatial visualizations. *Computer Graphics Forum*, 40(3):49–60, 2021. doi: 10.1111/cgf.14288

[16] K. Reda and D. A. Szafir. Rainbows revisited: Modeling effective colormap design for graphical inference. *IEEE transactions on visualization and computer graphics*, 27(2):1032–1042, 2021. doi: 10.1109/TVCG.2020.3030439

[17] B. E. Rogowitz and L. A. Treinish. Data visualization: the end of the rainbow. *IEEE Spectrum*, 35(12):52–59, 1998. doi: 10.1109/6.736450

[18] A. Sarikaya and M. Gleicher. Scatterplots: Tasks, data, and designs. *IEEE transactions on visualization and computer graphics*, 24(1):402–412, 2018. doi: 10.1109/TVCG.2017.2744184

[19] P. Schulze-Wollgast, C. Tominski, and H. Schumann. Enhancing visual exploration by appropriate color coding. *Proceedings of International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG)*, pp. 203–210, 2005.

[20] S. Smart, K. Wu, and D. A. Szafir. Color crafting: Automating the construction of designer quality color ramps. *IEEE transactions on visualization and computer graphics*, 26(1):1215–1225, 2020. doi: 10.1109/TVCG.2019.2934284

[21] C. Tominski, G. Fuchs, and H. Schumann. Task-driven color coding. *12th International Conference Information Visualisation (IV)*, pp. 373–380, 2008. doi: 10.1109/IV.2008.24

[22] E. R. A. Valiati, M. S. Pimenta, and C. M. D. S. Freitas. A taxonomy of tasks for guiding the evaluation of multidimensional visualizations. *BELIV ’06: Proceedings of the 2006 AVI workshop on BEyond time