6th European Conference on Computational Mechanics (ECCM 6)
7th European Conference on Computational Fluid Dynamics (ECFD 7)
11-15 June 2018, Glasgow, UK

COMPUTER VISION AIDED DATA EXTRACTION FROM CM AND CFD RESULTS

JARNO KIESILAINEN[1] AND TERO TUOVINEN[2]

[1] Faculty of Information Technology, University of Jyvaskyla, Finland
Mattilanniemi 2, 40014 Jyvaskyla
jarno.kiesilainen@student.jyu.fi

[2] Faculty of Information Technology, University of Jyvaskyla, Finland
Mattilanniemi 2, 40014 Jyvaskyla
tero.tuovinen@jyu.fi

Key words: Colormap, Visulization, Tool, Data extraction, Result comparison

Abstract. Meta-analysis and comparison of the research results in the field of computational mechanics (CM) and fluid dynamics (CFD) has never been straightforward. In a usual situation, result plots are generated through complicated numerical processes and only a final image or graph is produced in the journal pages. Depending on the author's taste, colormaps, scales etc. could be chosen in various different manners [1]. Because re-producing the images is a tedious (often impossible) task, the usual way is to compare results in a qualitative manner. However, this approach creates uncertain aspects and even false conclusions for the meta-analysis and does not support wide reliable (quantitative) comparison of multiple methods. Our solution is to use computer vision methodology and tools to overcome these challenges. In this article, we will present a simple web based tool for data extraction.

1 INTRODUCTION

As is widely known, computations in the field of mechanics and fluid dynamics are extremely time consuming, and if one would like to compare the results in a rigorous manner, computed solutions are often very difficult to attain. Without enormous work of implementation of complex numerical methods and solving, many times by extremely heavy computations, comparison and validation of the results are usually done in qualitative manner by checking that the result figures look fine. However, if quantitative data would be more readily available, reliable meta-analysis and comparison with other previously computed computations would pay attention to possible inconsistencies and flaws in the solutions. Or, at least, validate in some level that studies really are comparable.

The traditional way of representing 3D data in computational mechanics and computational fluid mechanics is XY-graphs with the color as the third dimension. In this manner, representing data is very flexible as many variables are left to the discretion of the

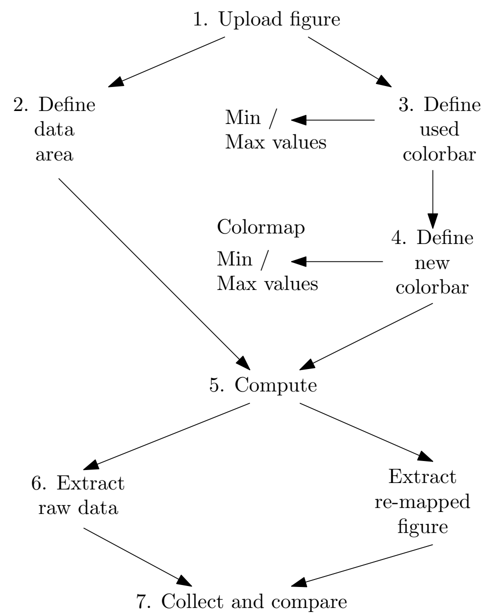


Figure 1: Flow chart of the image extraction process.

researcher. Especially choices regarding the color dimensions can produce visually very different representations. It is not widely known among researchers that poorly selected choices in the colormap can lead to some artifacts in the representation, which are not from the data. Usually this is unintentional selection, but sometimes this might be done with the aim of emphasizing results that are mediocre or minuscule, and behaviour of this kind should be easily captured in the review process. The other way of influencing the attractiveness of the results is the range of the color dimensions. The common way is to normalize the colormap so that the minimum value of the data is in the other end of the colormap and the maximum in the other. However, this can create situations even inside one publication, and figures are not easy to compare if the range of the colormap is not chosen equally. Generally comparing data with a different colormap or colormap range is difficult, and often it is not very useful or might lead to false conclusions.

Previous studies related to colormaps have been published e.g. by Ware [2]. In their study they performed experiments on how colormaps are percieved and they propose rules for selecting optimal colormap depending on the data. The experiments also reveal that with achromatic colormaps variations and extrema in the data are easier to detect, but it is easier to pick certain values from chromatic colormaps. In work by Borland and Ii [3] and also Rogowitz and Treinish [4] they explain the problems with the rainbow colormap and propose ways to pick alternatives. In context of machine vision Rohatgi [5] devoloped a tool that can extract data from two-dimensional graphs, but lacks the capability to extract three-dimensional surface data. In the article by Jeronen (see, [6]), there has been a long discussion about reliable scaling and production of the resulting figures in the case of uncertain data.

In this article, we have presented one publicly available solution, based on computer vision techniques, to tackle these thematics. A web based tool that uses simple computer

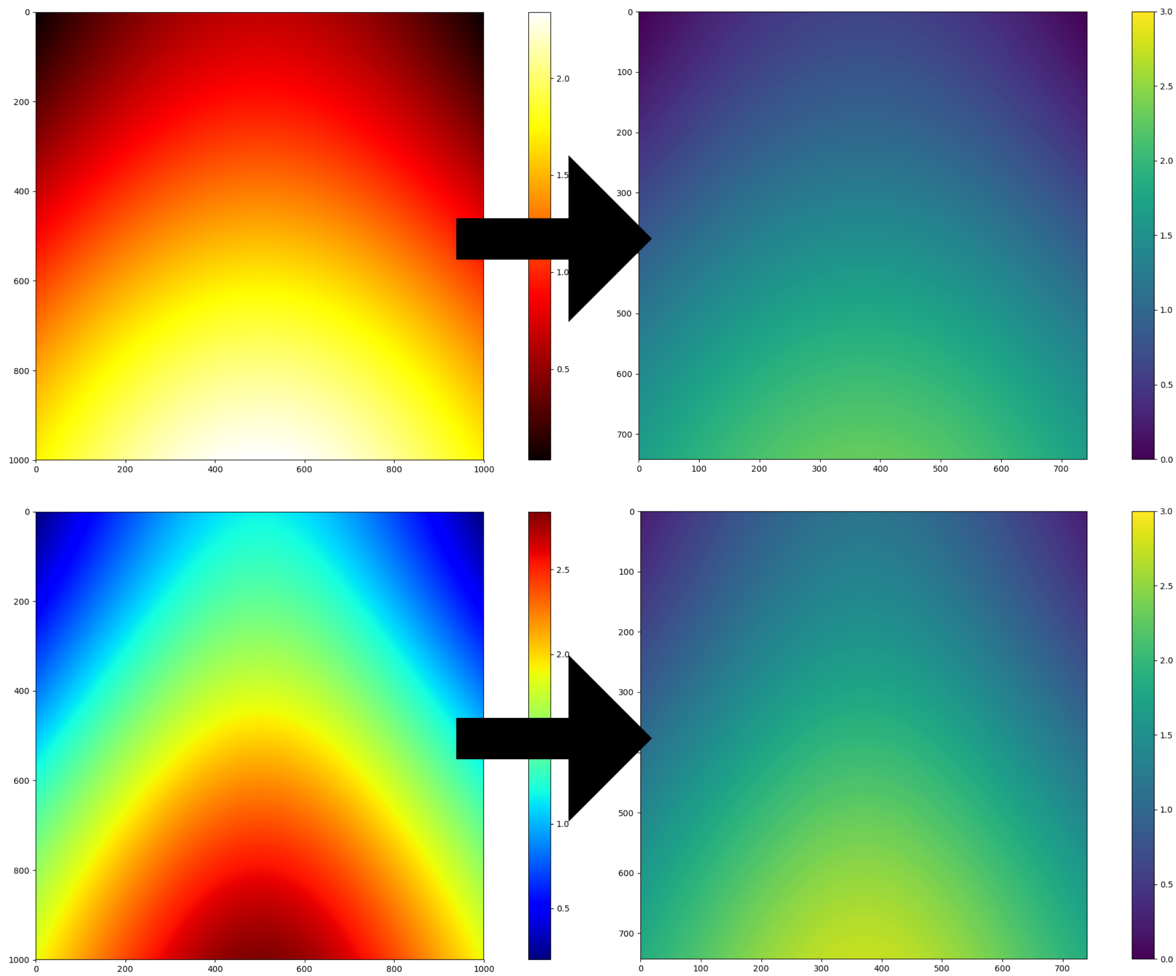


Figure 2: Two graphs that have slightly different data but completely different scales as well as colormaps converted to the same scale and a colormap with our tool. This makes the comparison of the data vastly easier.

vision methods to extract data from user-provided graphs and allows the user to save the extracted data in CVS, Matlab and NumPy formats. The idea is that this tool can be used for data extraction as well as for more advanced purposes like re-mapping (or re-painting) and comparison of the data. Using our approach, it is possible to conduct meta-analysis based on historical results published where original implementations are not available. In the Section 1 we have clarified the aim and importance of the topic. In the Section 2 we will discuss about chosen methodology. Various colormap and their selection criterias are in consideration in the Section 3. We will present our web-based application in the Section 4 and finally, in the Section 5, we will complete our study in conclusions and recommendations.

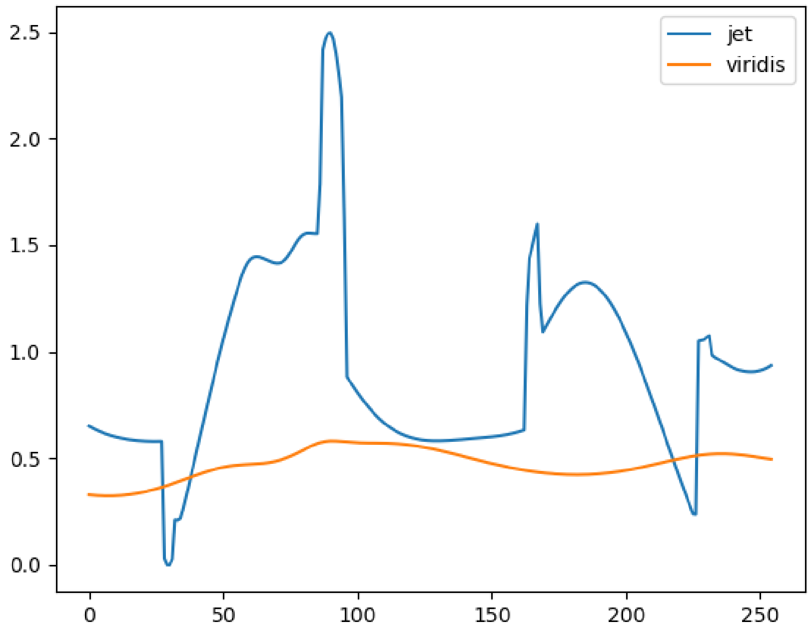


Figure 3: Perceptual deltas of viridis and jet colormaps according to CIEDE2000 metric.

2 Methodology

We often face the situation where the original data of the research result is not available, or it is produced by complicated numerical processes and reproducing the data might be hard or even impossible. In that case the easiest method might be to extract data directly from a graphical representation or even from a figure of a printed article. Multiple tools exist to extract data from graphs (see e.g. [5, 7]) but many of them work only with two-dimensional XY-data or only in limited capability with three-dimensional colormapped data, which is the most common one in the field of CM and CFD. The need for extensions in the tools is therefore evident.

In our approach, the program asks the user to upload an image file containing the graph and then to select the data area and colormap information from the image. After these defining steps, the program extracts the data based on colouring information, shows a preview to the user and allows him or her to download the data in multiple formats. This approach is simple but efficient, since using raw numerical data, the program can re-paint the uploaded image file, or the researcher can collect raw data from several sources and combine the information with a software of their choice. In Figure 1, we have presented a schematical flowchart of the process.

3 Selection of the colormaps

As all numerical researchers know, colormaps are widely used, and often they are the most practical way to visualize and represent data. Colormaps compress a lot of data in a convenient rectangular figure, from where from which the analysis can be accomplished qualitatively and even quantitatively, at least in principle. Because there are naturally limitations of on the length of the articles, numerical tables that contain the same information have vanished nearly completely, even though their information is more precise and more easily re-usable (although usually also more coarce).

There are multiple types of colormaps available, but the most common ones are sequential colormaps where in which the colors are arranged on in some order. The most used order is to have the color vary from dark to light or inverse vice versa. This is useful when representing numerical data, since the ordering of the data is preserved in some sense. Another common type of a colormap is a rainbow or jet colormap in which, instead of the color brightness, the color hue is varied. The rainbow colormap is very prevalent in the data visualization probably because it is the default colormap in many tools, even though research has shown that it is rarely the optimal choice [3, 4].

A color can be defined by three properties: hue, chroma and brightness. Changing any one of these changes the perception of the color. Difficulties with colormaps usually arise from the fact that the perception of property is not linear and there are relations between the properties, for example blue with the same brightess as yellow appears darker (hue affects the perception of brightness) [4]. This means that the linear change in the data doesn't necessarily appear linear in the graph. This non-linear behaviour can create artifacts and contrast in the graph, which do not exist in the actual data [8].

In Figure 2, the left side shows some chosen data plotted using *hot* and *jet* colormaps. It is easy to notice that bands can be seen in the graphs, especially in the yellow region. On the right side the same data is mapped with a *viridis* colormap that is perceptually uniform and it is possible to notice, how smooth the data appears with the *viridis* colormap and how features that do not exists in the data appear in the graphs on the left side like they do in the yellow bands earlier. Additionally, the range of the colormaps is set to be the same in the graphs on the right side, which makes comparing different data sets much easier. On the contrary, comparing the graphs on the left side is much more difficult.

One way to measure how the perceptual uniformity of a colormap is to have some system of measurement between the color values. The obvious initial choice is the euclidian distance (norm) of color's RGB or LAB values. This metric does not take into account the nonlinearity and relationships between color properties. More complicated measures have been developed that take into account the more complex properties of perception like the CIEDE2000 formulation [9].

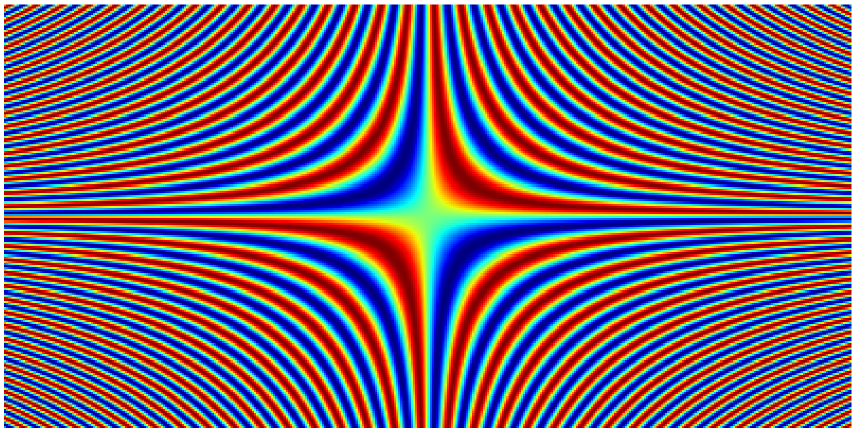


Figure 4: Function $f(x, y) = \sin(xy)$ plotted with the *jet* colormap.

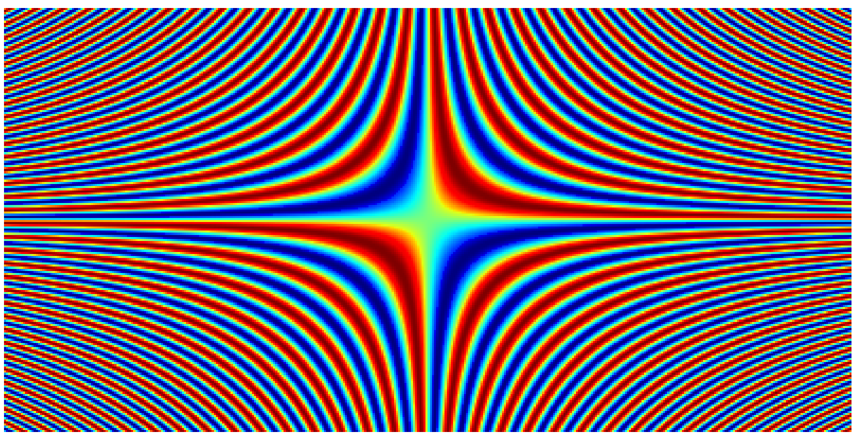


Figure 5: Data from figure 4 extracted and replotted with our tool.

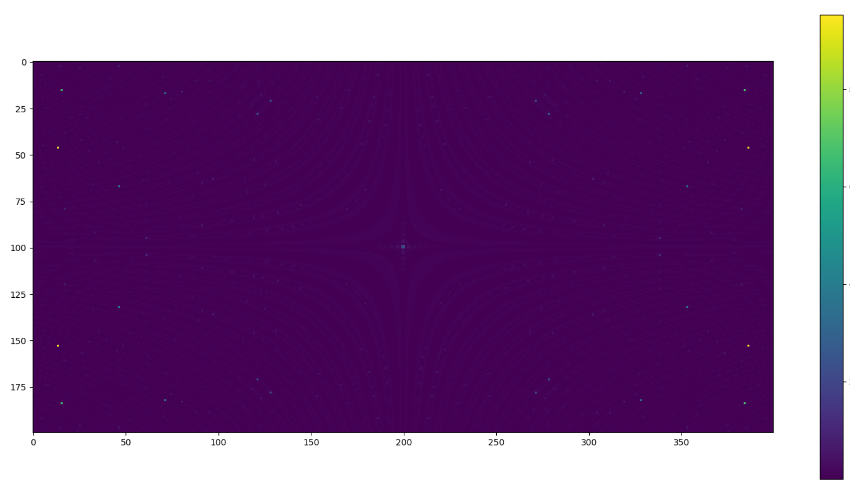


Figure 6: Relative error of the original data and the extracted data. An error differs from zero only in a few points, where the original data is close to zero and calculating the relative error is not numerically stable.

4 Web application for data extraction

We have developed and published a computer vision based web application that allows the user to extract approximation of the original data from a colormapped graph. The application performs a so-called 'inverse colormapping' on the graph image. This means that it compares each of the selected pixels to the values of the colormap to find the approximation of the original data value. The application is implemented with Python programming language. For numerical operations Numpy, Scipy and Matplotlib libraries are used and Flask is used as web framework. The application and it's source code is freely available at http://unplot.it.jyu.fi.

In next, we will present some benchmark cases of the capabilities of our tool. In Figure 4, we have generated test data from the following function:

$$f(x, y) = \sin(xy) \tag{1}$$

In Figure 5, the same data is extracted from the image file with our tool and then replotted using the same colormap. Qualitatively it is easy to see that the figures look alike, and in a more precise analysis, the comparison of the original and extracted data is preseted in Figure 6. The error appears to be very close to zero except in few points where the original data is close to zero and calculating the relative error is not numerically stable. The table of the measured values can be exported in multiple formats like CSV, Matlab and NumPy.

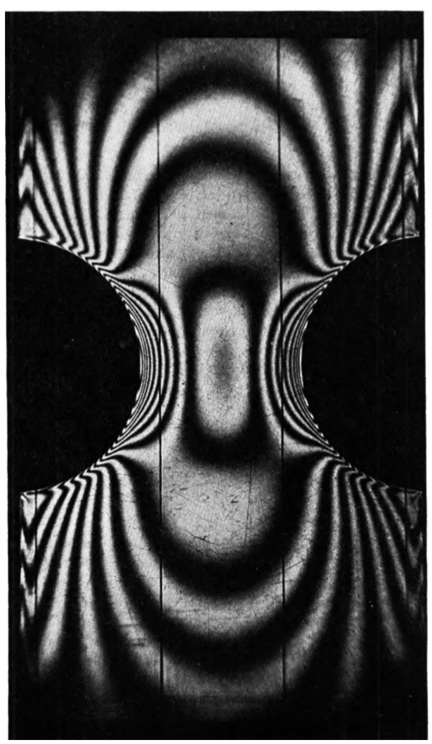
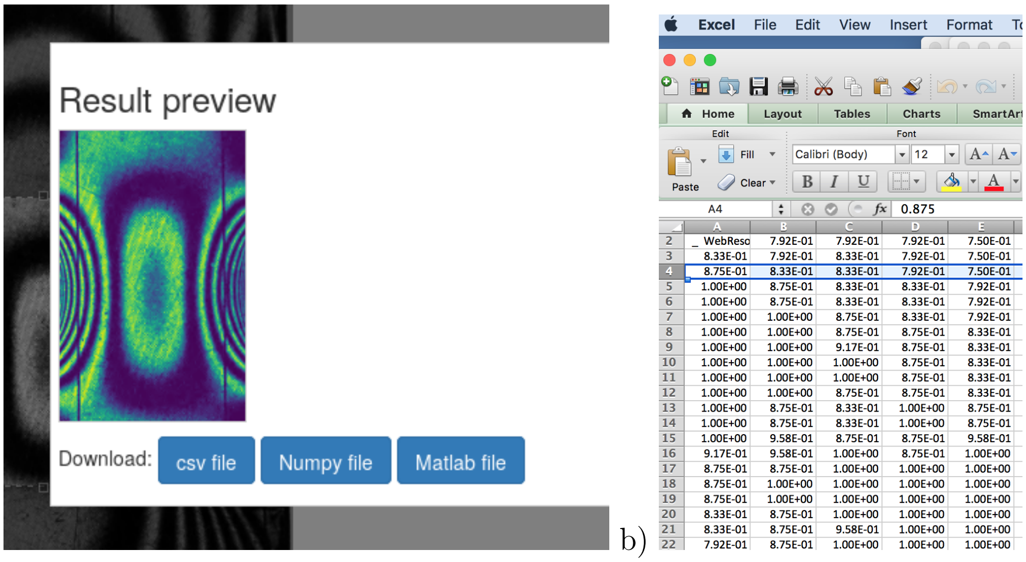


Figure 7: One randomly chosen result from the article [10].

a) b)



	A	B	C	D	E
2	_ WebReso	7.92E-01	7.92E-01	7.92E-01	7.50E-01
3	8.33E-01	7.92E-01	8.33E-01	7.92E-01	7.50E-01
4	8.75E-01	8.33E-01	8.33E-01	7.92E-01	7.50E-01
5	1.00E+00	8.75E-01	8.33E-01	8.33E-01	7.92E-01
6	1.00E+00	8.75E-01	8.33E-01	8.33E-01	7.92E-01
7	1.00E+00	1.00E+00	8.75E-01	8.33E-01	7.92E-01
8	1.00E+00	1.00E+00	8.75E-01	8.75E-01	8.33E-01
9	1.00E+00	1.00E+00	9.17E-01	8.75E-01	8.33E-01
10	1.00E+00	1.00E+00	1.00E+00	8.75E-01	8.33E-01
11	1.00E+00	1.00E+00	1.00E+00	8.75E-01	8.33E-01
12	1.00E+00	1.00E+00	8.75E-01	8.75E-01	8.33E-01
13	1.00E+00	8.75E-01	8.33E-01	1.00E+00	8.75E-01
14	1.00E+00	8.75E-01	8.33E-01	1.00E+00	8.75E-01
15	1.00E+00	9.58E-01	8.75E-01	8.75E-01	9.58E-01
16	9.17E-01	9.58E-01	1.00E+00	8.75E-01	1.00E+00
17	8.75E-01	8.75E-01	1.00E+00	1.00E+00	1.00E+00
18	8.75E-01	1.00E+00	1.00E+00	1.00E+00	1.00E+00
19	8.75E-01	1.00E+00	1.00E+00	1.00E+00	1.00E+00
20	8.33E-01	8.75E-01	1.00E+00	1.00E+00	1.00E+00
21	8.33E-01	8.75E-01	9.58E-01	1.00E+00	1.00E+00
22	7.92E-01	8.75E-01	1.00E+00	1.00E+00	1.00E+00

Figure 8: a) Preview of the data from the application. b) All data is exported to Excel to wait for further analysis.

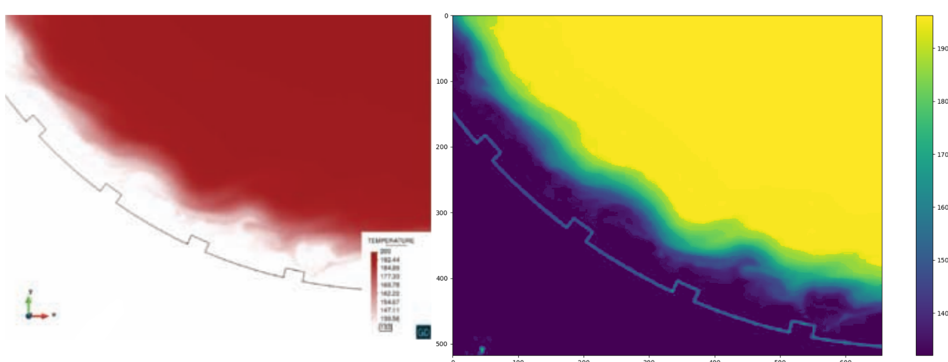


Figure 9: Plot from a another reasearch article [11] transformed to a better colormap with our tool.

In a one case study, we have chosen one (random) historical research from the 1930's (see [10]). Our aim could be to compare the results from historical measurements to more recent studies or, if the old measurements would contradict newer ones, we could use for example parametric identifation methods and simulations for tracking the conditions of measurement and other parameters from 1930's. The aim would be to know or guess where the discrepency arises. In general, we aim that the data would be easier to analyse numerically with various methods.

In Figure 7, we have randomly chosen one of the result figure from the article. We cannot use any color-bar from the article, because there is none, so we have to use a pre-defined binary color-bar which gives us values in the range [0,1]. In Figure 8, we have computed some area of the original study an the application asks how to export results. Later on, we could scale these values easily to correct order, focus areas for consideration on top of each other, compute e.g. difference values and complement the analysis (see Figure 8).

In the final case study, we have chosen CFD simulation from the paper industry and inside of drying cylinder. In Figure 9 a result is replotted with the help of our tool to a different colormap that has better perceptual uniformity.

5 Conclusion

In this article, we have presented our freely available web-based application, which uses computer vision methods for extracting results from scientific articles and measurements. We have demonstrated that this tool can provide easy access to the implementation of meta-level analysis between historical and present days studies by enabling semi-automatic numerical information conversion and collection. Moreover, we have considered reliable color mapping and proposed one conversion technique that produces similar mapping between the chosen results. The comparison of results would be more easier qualitatively and quantitatively. The tool will help reviewers and researchers to spot situations such as possible over-emphasing of research results and confusing by mapping scale.

REFERENCES

[1] R. Bujack, T. L. Turton, F. Samsel, C. Ware, D. H. Rogers, and J. Ahrens. The good, the bad, and the ugly: A theoretical framework for the assessment of continuous colormaps. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):923–933, Jan 2018.

[2] C. Ware. Color sequences for univariate maps: theory, experiments and principles. *IEEE Computer Graphics and Applications*, 8(5):41–49, Sept 1988.

[3] D. Borland and R. M. Taylor Ii. Rainbow color map (still) considered harmful. *IEEE Computer Graphics and Applications*, 27(2):14–17, March 2007.

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

[5] A. Rohatgi. Webplotdigitizer, October 2017.

[6] J. Jeronen. SAVU: Sample-based Analysis and Visualization of Uncertainty. In *Proceedings of CAO2011 – ECCOMAS Thematic Conference on Computational Analysis and Optimization*, pages 165–168. University of Jyväskylä, 2011. ISBN 978-951-39-4331-8.

[7] B. Muftakhidinov M. Mitchell and T. Winchen et al. Engauge digitizer software. Accessed: January 30, 2018.

[8] L. D. Bergman, B. E. Rogowitz, and L. A. Treinish. A rule-based tool for assisting colormap selection. In *Proceedings of the 6th Conference on Visualization '95*, VIS '95, pages 118–, Washington, DC, USA, 1995. IEEE Computer Society.

[9] G. Sharma, W. Wu, and E. N. Dalal. The ciede2000 color difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Research & Application*, 30(1):21–30, 2004.

[10] AM Wahl and R Beeuwkes Jr. Stress concentration produced by holes and notches. *Stress*, page 141, 1930.

[11] J. Jeronen T. Tuovinen and P. Becker. Simulation of sloshing of condensate water inside a drying cylinder using particle finite element method. *Advances in Pulp and Paper Research, Oxford 2017*, 1:383–407, September 2017.