# Colorscale Identification Study

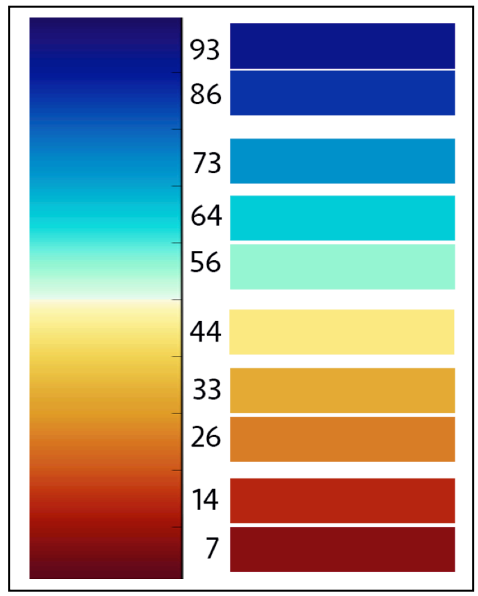
Terece Turton – Center for Agile Technology, University of Texas at Austin
Francesca Samsel – Center for Agile Technology, University of Texas at Austin
Joanne Wendelberger – Statistical Sciences Team, Los Alamos National Laboratory

## Hypothesis

Our hypothesis is that we can use statistical evaluation of the accuracy and precision with which people can identify a data value associated with a specific color on a colorscale to impact and improve colorscale development through an iterative process.

## Experimental Setup

The colorscale user study methodology consisted of ten questions, each question consisting of the same colorscale and 10 randomized color panels. The colorscale data range was defined as running from 0 to 100 and the ten color panels in the study were chosen at specific data values: 7, 14, 26, 33, 44, 56, 64, 73, 86, and 93. Participants were asked to estimate the data value associated with the color panel shown by using a Qualtrics slider ranging from 0 to 100. Each colorscale study collected between 50 and 100 participants.




As an example, the complete picture of the blue/orange divergent colorscale study (reworked version), BOD2, is shown to the right. This includes the colorscale and the ten color panels corresponding to the ten data values used in the study.

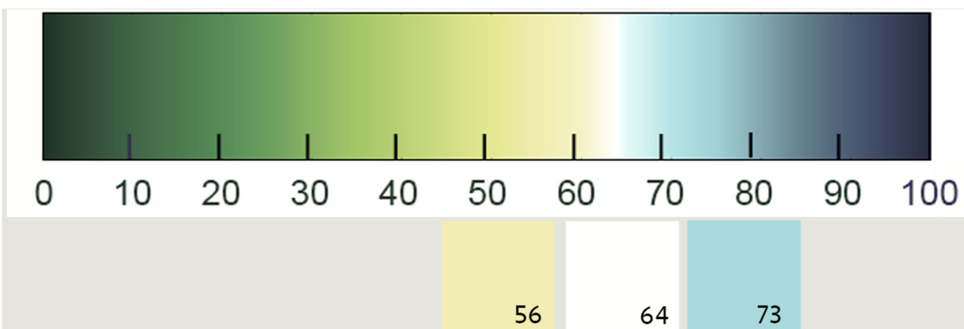
This paper looks at three sets of colormaps. The first set are some of the commonly used standard colormaps: rainbow (RA), cool/warm (CW), heat map (HM), and ParaView grayscale (GPV). The second set are colormaps designed by our team: blue/orange divergent (BOD), blue/green asymmetric divergent (BGAD), extended cool/warm (ECW), blue/orange asymmetric divergent (BOAD), linear yellow/green/blue (YGB), and a set of grayscales, G3, G4, G5 and G8. The last set of colormaps are iterated versions of some of the above: BOD2, BOAD2, BGAD2, ECW2, and YGB2. The colorscales associated with the colormaps of interest are shown at the end of this summary.

## Data Analysis and Results

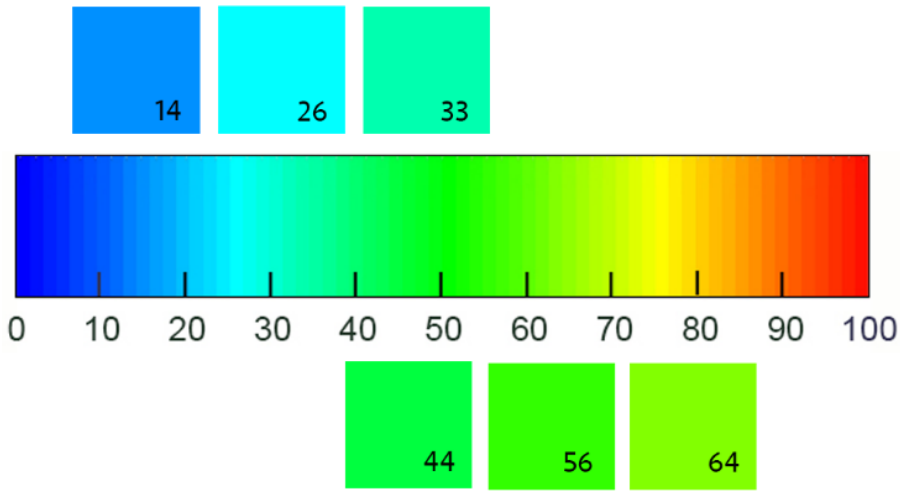
Each response was screened to insure that participants understood and faithfully attempted the task. When using a Qualtrics slider, it is relatively easy to make two different types of errors as a participant. One is to auto-click on a value that ends with five or zero (e.g. 5, 10, 15, 20, etc.). Participants that had five or more responses that ended with a five and/or a zero were removed as it indicated that the participant may not have been very careful about accuracy in their response. Another possible error is to accidentally click along the slider and continue with an unintentional selection. In order to cut out those responses, responses with one or more extreme outliers were removed (6 or more points away from the nearest other response for the respective color panel). To determine accuracy and precision, we calculated the mean and standard deviation for each of the 10 color panels (estimated data value) for the colorscales tested. This information is summarized in the table at the end of this summary. The top half of the table details the means and standard deviations for the first five data values/color panels (7, 14, 26, 33, 44) and the bottom half of the tables details the means and standard deviations for the second five data values/color panels (56, 64, 73, 86, 93).

## Identifying Weaker Areas of a Colorscale

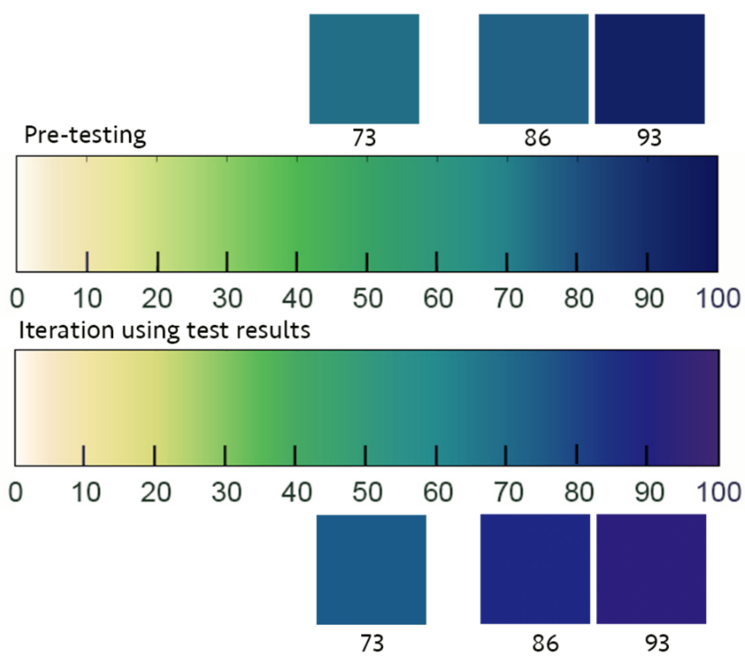
As can be seen in the table, the majority of the means are within one standard deviation of the expected (actual) data value. This is consistent with a null hypothesis that people can accurately identify the data value associated with a specific color. The standard deviations, a measure of how precisely participants can identify the data value, range from a low of 1.20 for the blue/green asymmetric divergent at a data value of 64 to a high of 8.89 for the ParaView grayscale at the data value of 33. Looking closely at the means and standard deviations and comparing to the associated colorscales, it is apparent that the means tend to be close to the expected value where there is a divergence in the colorscale or where there is a transition from one hue to another. This makes intuitive sense -- when there is a significant shift in hue, saturation and/or value, it is easier to compare the color panel to the colorscale and select a data value close to the correct data value. The blue/green asymmetric divergent colorscale (BGAD) is shown below, along with the color panels at the data values of 56, 64 and 73. At a data value of 64, there is a strong divergence, very different colors from the next testing points and little chance of choosing a value far from the actual value.




The rainbow colorscale, shown below, has a color transition around the data value of 26, leading to a low standard deviation of 1.98 at that data value. The color panels at 14, 26, and 33 are quite different colors. Conversely, the lack of perceptual differences along the green range in the rainbow leads to a less precise estimation and higher standard deviations of 5.68, 5.51, and 7.01, respectively for the data values of 44, 56 and 64. The higher standard deviations indicate a lower precision for that data value and indicate an area where data may suffer from distortion.



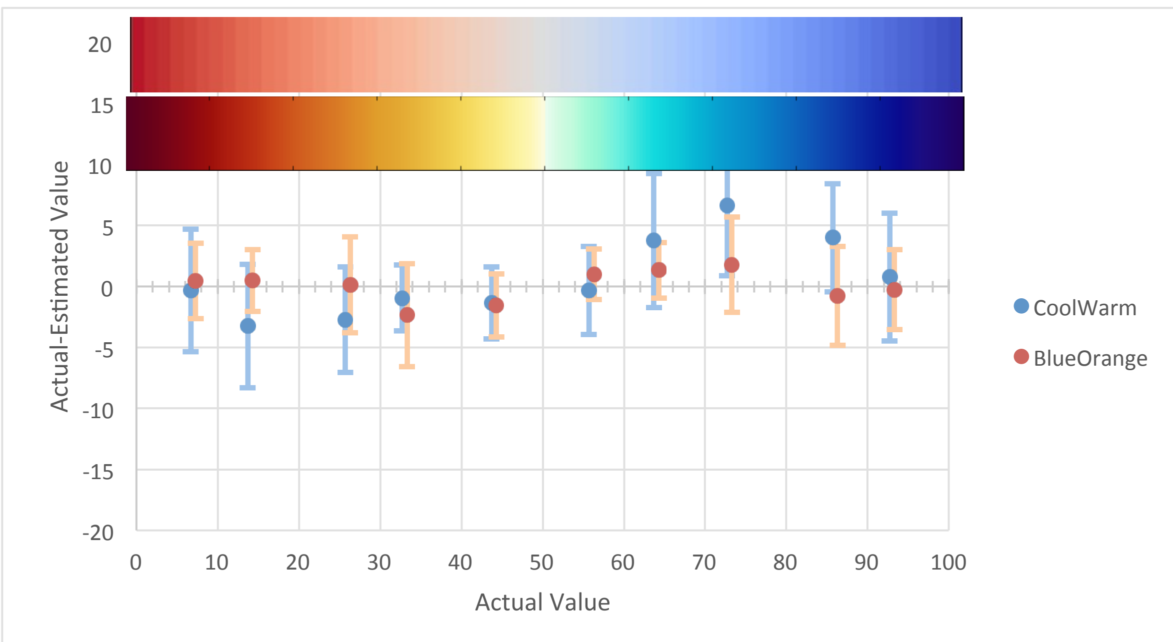

In addition to the information on precision we have using the standard deviations, there are specific areas with poor accuracy, where the mean is more than one or two standard deviations away from the actual data value. This indicates where a colorscale could be better designed. To test if this information can impact colorscale design, we looked at the YGB colorscale around the data value of 86. At this point, the mean was 76.32, with a standard deviation of 4.35, thus the measured value was more than two standard deviations away from the actual data value of 86. The original yellow/green/blue is shown in the upper half of this figure with the color panels for the data values of 73, 86, and 93.




This colorscale was reworked by adjusting the hue, value, and saturation values with the goal of creating more perceptual steps in the upper data ranges. In particular, the blues were adjusted to create more steps in hue in that region of the colorscale. The new colorscale, YGB2, was then tested. There is significant improvement in the mean and standard deviation at the data value of 86, with a new mean of 87.19 and a new standard deviation of 4.00. The new YGB2 colorscale is shown in the same figure with the color panels for data values of 73, 86 and 93 again. While gaining accuracy and improving the mean and standard deviation at the data values of 86 and 93, there was an associated increase in the standard deviation at the data values of 64 and 73, an indication of how a shift in one area in a colorscale can impact along the full color sequence. However, as a case study in using user studies to impact and improve colorscale design, this was an important first step.

Other colorscales were then reworked to impact specific locations, resulting in ECW2, BGAD2, BOAD2, and BOD2.

Plotting the difference between actual and estimated values is a useful way to identify areas of the colorscale that might be improved and can also be used to compare two colorscales. The standard cool/warm and the final blue/orange divergent are compare in the following plot:



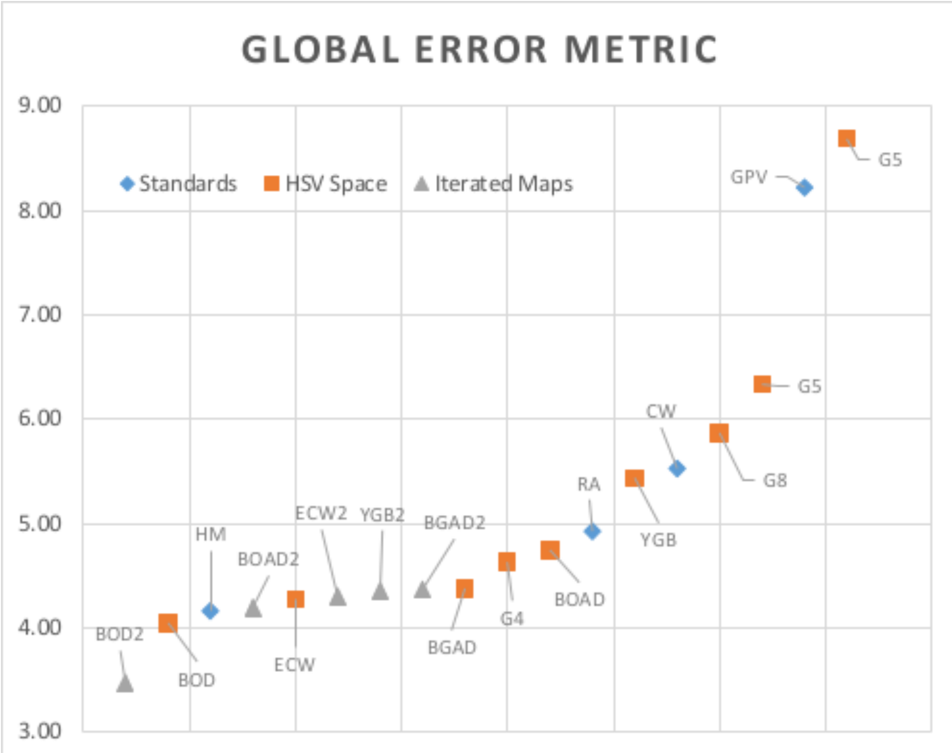

While looking at the means and standard deviations for each of the 10 data values gives us information about specific points, we need a more global approach. A useful metric would be an overall measure of how accurately study participants can identify the data values associated with a specific color across the full colormap. We draw on the statistical concept of the root mean squared error. We start by using the difference between the actual data value and the estimated data value. By squaring and summing those differences, we add up the squared sum of all the errors across the full colorscale. We then normalize that error by dividing by the number of data points, N=10*n, and taking the square root. We therefore define the global error metric or GEM as:

$$GEM = \sqrt{\frac{\sum_i^{10} \sum_j^n (x_{ij} - X_i)^2}{10 * n}}$$

where X_i is the ith actual data value and n is the number of responses.

We can use this variable to compare the global accuracy of the tested colorscales. The following table shows each of the colorscales tested in increasing order of GEM. The GEM values are plotted in the following plot.

| ColorScale | | GEM |
|---|---|---|
| **Blue/Orange Div 2** | **BOD2** | **3.47** |
| **Blue/Orange Div** | **BOD** | **4.05** |
| **Heat Map** | **HM** | **4.16** |
| **Blue/Orange AsyDiv 2** | **BOAD2** | **4.19** |
| **Ext Cool/Warm** | **ECW** | **4.28** |
| **Ext Cool/Warm 2** | **ECW2** | **4.30** |
| **Yellow/Green/Blue 2** | **YGB2** | **4.35** |
| **Blue/Green AsyDiv 2** | **BGAD2** | **4.36** |
| **Blue/Green AsyDiv** | **BGAD** | **4.38** |
| **Gray4** | **G4** | **4.63** |
| **Blue/Orange AsyDiv** | **BOAD** | **4.74** |
| **Rainbow** | **RA** | **4.92** |
| **Yellow/Green/Blue** | **YGB** | **5.44** |
| **Cool/Warm** | **CW** | **5.52** |
| **Gray8 -BlGrBr** | **G8** | **5.87** |
| **Gray3** | **G3** | **6.33** |
| **Gray ParaView** | **GPV** | **8.22** |
| **Gray5** | **G5** | **8.69** |




Note that the colorscales with the lowest global metric are those that have frequent color transitions or divergences. Redoing the yellow/green/blue (YGB versus YGB2) colorscale dropped its global error metric from 5.496 to 4.377, reflecting the overall improvement in the colorscale when one area was reworked to improve a local accuracy problem. We also saw a drop in GEM for the reworked blue/orange asymmetric divergent and blue/orange divergent. However, modifying the ECW and BGAD colorscales, while sometimes impacting individual areas, did not decrease the overall GEM. As a step towards globally assessing accuracy across a full colorscale, the global metric, as defined above, gives us a useful measure to compare different colorscales.

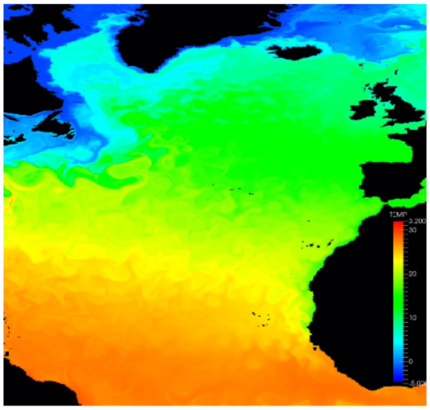
While a low GEM is associated with better perceptual accuracy, a higher GEM does not necessarily imply a colorscale that should not be used. The grayscales naturally tend towards higher GEM but are often appropriate choices depending on the type of data and scientist needs.

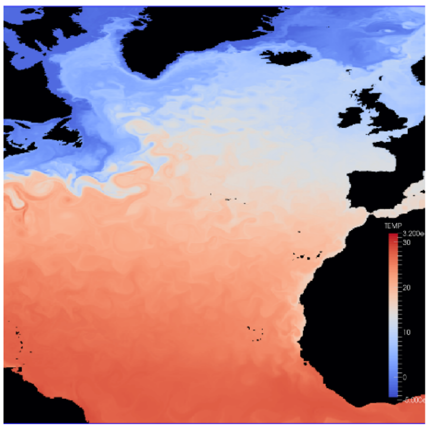
The above considerations: mean, standard deviations and GEM, allow us to develop some rules of thumb when identifying areas within a colormap that might be improved. Generally, if the estimated mean is within $\pm 2$ of the actual mean and the standard deviation is less than 5.0, the contrast in that area has sufficient accuracy and precision such that it is difficult to further improve the colorscale.

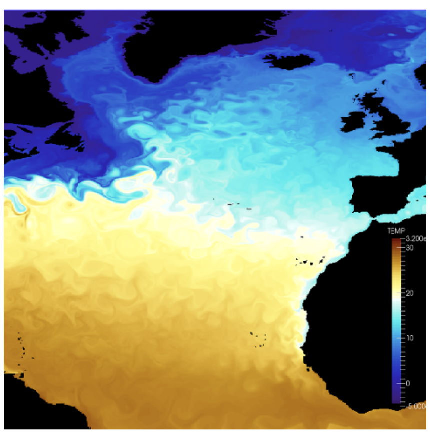
## Conclusions

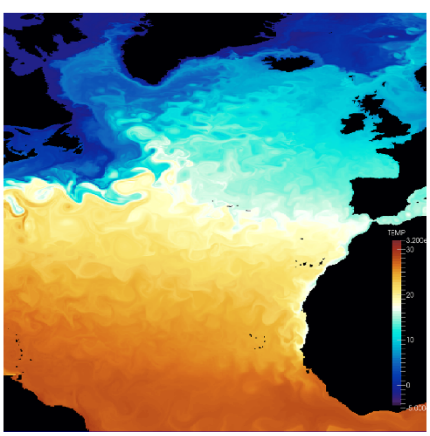
Well-designed user studies can provide the basis for evaluating and improving colormaps. The methodology presented can be employed to make significant improvements to the colormaps developed, resulting in colormaps that address the needs and task goals of the scientist. The following figure compares a section of the POP ocean data rendered in the rainbow, cool/warm, blue/orange asymmetric divergent and the reworked blue/orange asymmetric divergent colormaps (RA, CW, BOAD,

BOAD2). While we certainly expect the ability to see greater detail in the data with the artist-designed colormap, BOAD, versus the standards, RA and CW, we also see that the reworked BOAD2 provides still more detail than the original BOAD.


TEMP
30
20
10
0


TEMP
30
20
10
0


TEMP
30
20
10
0

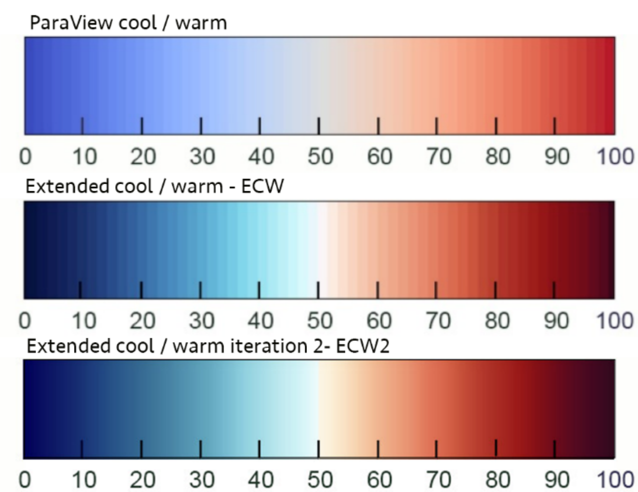

TEMP
30
20
10
0

| Data Value: | | 7 | | 14 | | 26 | | 33 | | 44 | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **ColorScale** | **n** | **Mean** | **StDev** | **Mean** | **StDev** | **Mean** | **StDev** | **Mean** | **StDev** | **Mean** | **StDev** |
| **HM** | 40 | 6.85 | 3.28 | 16.03 | 3.46 | 25.65 | 4.79 | 35.85 | 3.92 | 43.98 | 3.46 |
| **RA** | 47 | 4.40 | 3.77 | 17.13 | 3.88 | 26.49 | 1.98 | 32.11 | 3.50 | 49.43 | 5.68 |
| **YGB** | 47 | 6.47 | 2.30 | 13.91 | 2.77 | 27.94 | 4.45 | 35.36 | 4.43 | 43.40 | 5.33 |
| **YGB 2** | 94 | 6.67 | 3.24 | 14.17 | 3.38 | 27.68 | 3.67 | 34.66 | 3.91 | 44.02 | 4.35 |
| | | | | | | | | | | | |
| **CW** | 45 | 7.78 | 5.24 | 18.00 | 4.45 | 32.64 | 5.77 | 36.76 | 5.49 | 43.67 | 3.60 |
| **ECW** | 42 | 5.79 | 2.94 | 16.76 | 4.55 | 29.86 | 4.47 | 36.86 | 4.10 | 44.57 | 2.62 |
| **ECW 2** | 66 | 4.88 | 3.33 | 13.97 | 4.82 | 28.39 | 5.13 | 34.79 | 4.78 | 43.92 | 3.76 |
| | | | | | | | | | | | |
| **BGAD** | 99 | 8.16 | 3.87 | 16.51 | 4.33 | 26.22 | 4.12 | 37.76 | 3.70 | 47.44 | 5.16 |
| **BGAD 2** | 65 | 7.62 | 3.81 | 15.75 | 4.93 | 27.89 | 5.49 | 34.54 | 5.05 | 46.68 | 4.73 |
| | | | | | | | | | | | |
| **BOAD** | 66 | 6.77 | 2.51 | 15.29 | 3.85 | 27.52 | 2.94 | 33.03 | 1.33 | 40.42 | 3.07 |
| **BOAD 2** | 67 | 6.25 | 2.90 | 14.60 | 2.52 | 28.04 | 3.20 | 34.04 | 2.46 | 44.63 | 2.95 |
| **BOD** | 42 | 6.57 | 3.45 | 14.88 | 2.96 | 27.71 | 4.19 | 36.33 | 3.85 | 44.24 | 3.01 |
| **BOD 2** | 67 | 6.61 | 3.07 | 13.51 | 2.55 | 25.99 | 3.85 | 35.54 | 4.00 | 45.52 | 2.58 |
| | | | | | | | | | | | |
| **G3** | 90 | 4.96 | 3.39 | 11.18 | 4.97 | 25.34 | 6.66 | 34.47 | 6.60 | 46.66 | 6.15 |
| **G4** | 94 | 6.21 | 2.95 | 14.13 | 3.27 | 28.56 | 3.99 | 34.87 | 4.80 | 44.12 | 3.87 |
| **G5** | 82 | 6.71 | 3.87 | 13.09 | 5.54 | 25.32 | 7.18 | 32.77 | 7.68 | 46.88 | 9.99 |
| **G8** | 91 | 6.91 | 3.35 | 12.64 | 4.14 | 25.30 | 5.69 | 32.43 | 4.66 | 45.14 | 6.49 |
| **GPV** | 91 | 6.64 | 3.93 | 13.38 | 5.91 | 24.77 | 8.27 | 33.36 | 8.89 | 44.15 | 8.71 |
| | | | | | | | | | | | |
| **Data Value:** | | **56** | | **64** | | **73** | | **86** | | **93** | |
| **ColorScale** | **n** | **Mean** | **StDev** | **Mean** | **StDev** | **Mean** | **StDev** | **Mean** | **StDev** | **Mean** | **StDev** |
| **HM** | 40 | 57.43 | 3.21 | 61.65 | 3.35 | 72.00 | 4.38 | 83.85 | 3.64 | 96.08 | 3.33 |
| **RA** | 47 | 53.60 | 5.51 | 61.60 | 7.01 | 74.17 | 2.49 | 84.36 | 2.72 | 92.68 | 3.00 |
| **YGB** | 47 | 55.68 | 5.29 | 64.21 | 4.56 | 72.68 | 3.77 | 76.32 | 4.35 | 94.38 | 5.21 |
| **YGB 2** | 94 | 58.00 | 4.45 | 64.65 | 5.51 | 73.31 | 4.52 | 87.19 | 4.00 | 92.60 | 4.67 |
| | | | | | | | | | | | |
| **CW** | 45 | 54.64 | 2.95 | 63.04 | 2.70 | 70.27 | 4.34 | 82.76 | 5.07 | 92.67 | 5.02 |
| **ECW** | 42 | 53.98 | 2.03 | 61.60 | 3.96 | 71.00 | 3.39 | 87.71 | 3.46 | 95.14 | 2.49 |
| **ECW 2** | 66 | 54.05 | 2.36 | 61.64 | 4.29 | 72.41 | 4.02 | 86.00 | 3.20 | 94.59 | 3.35 |
| | | | | | | | | | | | |
| **BGAD** | 99 | 58.51 | 3.39 | 64.80 | 1.20 | 71.67 | 2.72 | 86.17 | 3.72 | 94.75 | 3.53 |
| **BGAD 2** | 65 | 57.62 | 3.33 | 64.23 | 2.35 | 71.18 | 2.65 | 85.05 | 3.41 | 92.42 | 3.80 |
| | | | | | | | | | | | |
| **BOAD** | 66 | 52.56 | 5.51 | 61.44 | 6.10 | 72.71 | 5.76 | 85.45 | 5.46 | 92.74 | 4.25 |

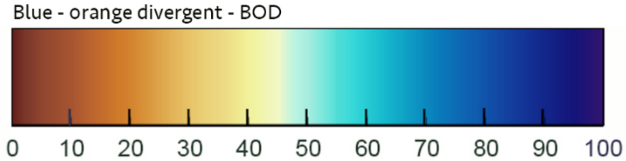
| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **BOAD 2** | 67 | 53.30 | 4.14 | 62.36 | 3.69 | 72.96 | 4.48 | 83.12 | 5.82 | 92.01 | 5.05 |
| **BOD** | 42 | 55.38 | 3.09 | 61.95 | 4.20 | 72.50 | 4.44 | 87.88 | 4.00 | 94.36 | 3.66 |
| **BOD 2** | 67 | 55.00 | 2.09 | 62.69 | 2.29 | 71.18 | 3.91 | 86.69 | 4.00 | 93.19 | 3.24 |
| | | | | | | | | | | | |
| **G3** | 90 | 55.42 | 7.32 | 63.32 | 7.63 | 74.04 | 5.68 | 88.66 | 6.05 | 93.49 | 5.25 |
| **G4** | 94 | 55.54 | 5.33 | 65.27 | 4.12 | 72.41 | 4.67 | 87.05 | 5.68 | 93.26 | 5.23 |
| **G5** | 82 | 60.48 | 10.96 | 67.46 | 10.04 | 79.17 | 9.49 | 91.94 | 6.33 | 97.05 | 4.01 |
| **G8** | 91 | 52.10 | 6.42 | 59.67 | 7.37 | 71.35 | 6.01 | 85.65 | 5.27 | 96.04 | 3.25 |
| **GPV** | 91 | 57.45 | 7.33 | 71.54 | 7.73 | 78.92 | 7.72 | 94.25 | 5.47 | 98.35 | 2.56 |

Table 1: The top half of Table 1 shows the mean and standard deviation for the first five data values: 7, 14, 26, 33, and 44 for each colorscale tested. The bottom half of Table 1 shows the mean and standard deviation for the rest of the data values: 56, 64, 73, 86, and 93 for each colorscale tested.

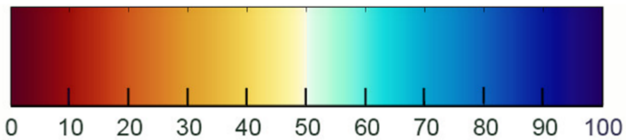
Below are the set of colormaps studied to date:



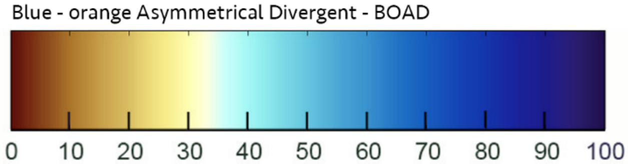
Blue - orange divergent - BOD



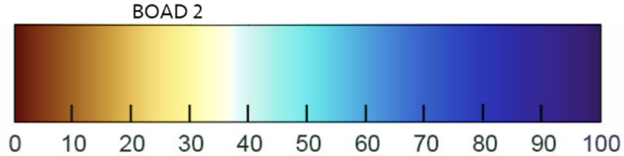

BOD2



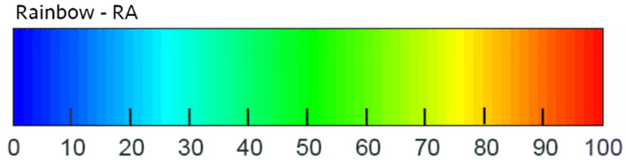

Blue - orange Asymmetrical Divergent - BOAD



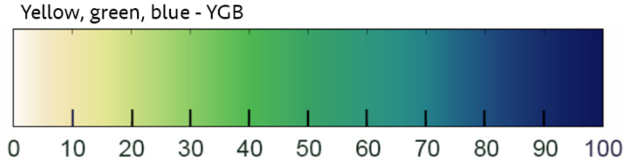

BOAD 2



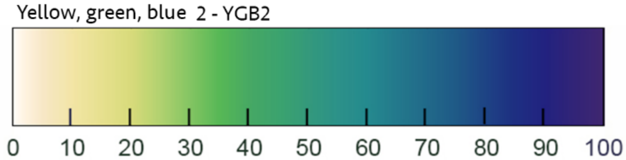

Rainbow - RA



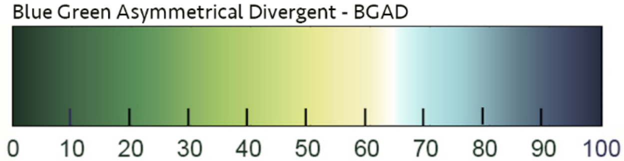

Yellow, green, blue - YGB



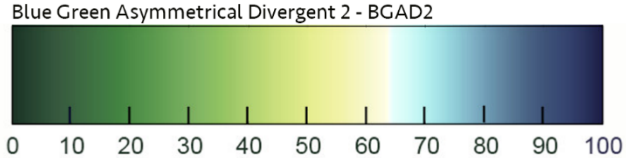

Yellow, green, blue 2 - YGB2



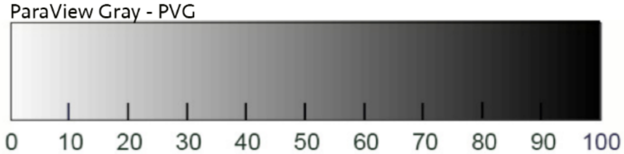

Blue Green Asymmetrical Divergent - BGAD



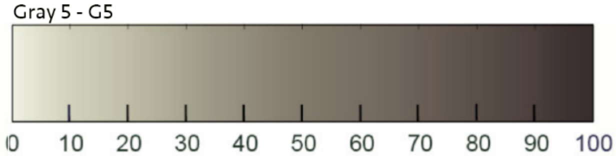

Blue Green Asymmetrical Divergent 2 - BGAD2



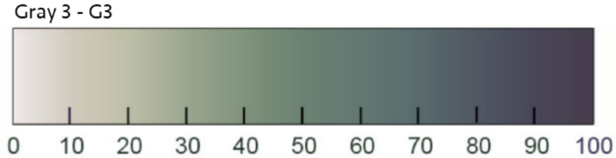
ParaView Gray - PVG



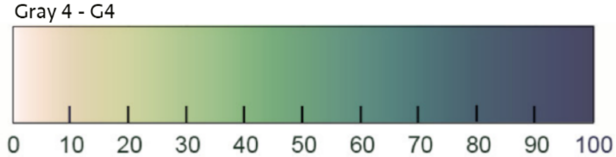

Gray 5 - G5




Gray 3 - G3




Gray 4 - G4




Gray 8 - G8