

# Data visualization HW2

Hongsup Oh

February 2020

## 1 Part1

### 1. Problem1

To generate image, readData and showData functions are used. Below image is image at resolution 17 and time step 0.

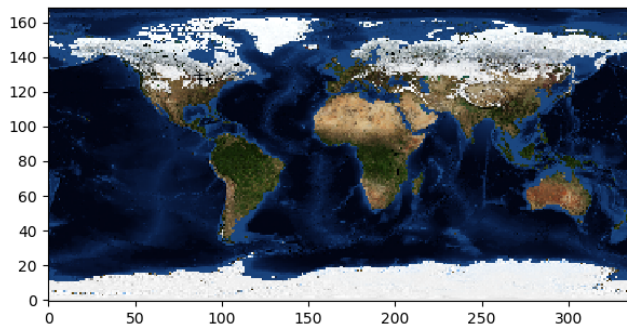


Figure 1: Image at time=0 and resolution=17

### 2. problem2

Resolution range is from 5 to 21 and step is 4. And time is constant as 0. For the detail, please check Part1-2 at code file.

### 3. problem3

Resolution range is from 5 to 21 and step is 4. And time range is from 0 to 11 and step is 1. For the detail, please check Part1-3 at code file.

4. problem4

Resolution range is from 5 to 21 and step is 4. And time range is from 0 to 11 and step is 1. In addition, threshold range is from 0 to 300 and step is 4. For the detail, please check Part1-4 at code file. Figure 2 shows one of the example of the code.

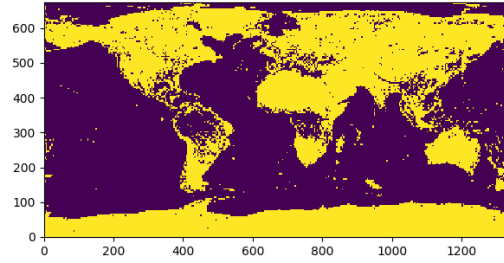


Figure 2: time=0, resolution=17, threshold=50

5. problem5-a

Below image is density of snow at resolution of 5,9,13,18 and 21. And density-sn and plot1-5 functions are used. As resolution increases, density of snow decreases in all time. Resolution 5's graph is largely different with that of rest.

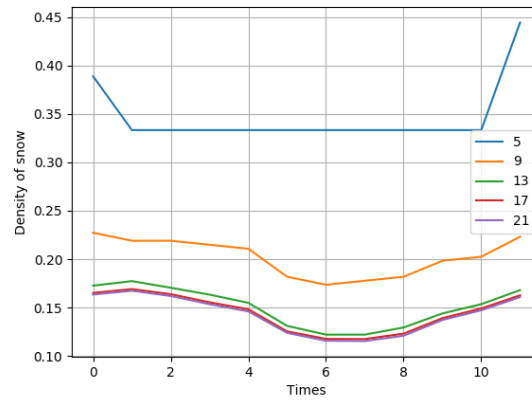


Figure 3: Density of snow at each resolution

6. problem5-b

Errors between adjacent two lines are measured. As resolution increases, error decreases. And standard deviation of 13-17 and 17-21 are less than 0.002. As resolution increases, errors decreases in all times. Thus, resolution 13,17 and 21 are selected as low error resolution for next step.

Answer is:

13-17 is less than 0.002 and std is 0.001310487734839918

17-21 is less than 0.002 and std is 0.0001843712523129319

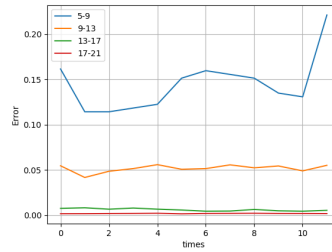


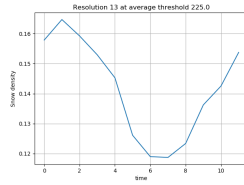
Figure 4: Density of snow at each resolution

7. problem6

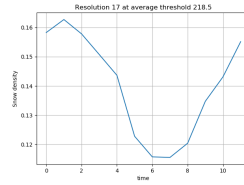
Northern hemisphere is defined by upper half of the data. If name of the data is dt, then northern hemisphere is  $dt[y//2: ]$ . And threshold values from 194 to 280 are checked for constraint condition. If all density values range from 0.05 to 0.15, then record the value and calculate average value of threshold. density-winter-NH and findRange functions are used to calculate optimal threshold. As a result 225 at threshold 13, 218.5 at threshold 17 and 218 at threshold 21.

8. problem7

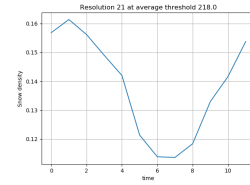
At resolution 13,17 and 21 shows same trend such as Feb is the highest snow density and Aug is the lowest snow density. But July's snow density is almost same with that of august and difference between two months are not meaningful. Even though all graphs show almost same value, density of snow decreases as resolution increases.



(a) Resolution at 13



(b) Resolution at 17



(c) Resolution at 21

9. problem8

From Dec to Jan Q2's snow density is higher than that of Q1. Rest of times, Q1's snow density is higher than that of Q2. Below plot is difference between density of snow at Q2 (d2) and difference between density of snow at Q1 (d1). That is, y-axis is  $d2-d1$  and x-axis is times. Label at 11 (Dec) and at 0 (Jan), difference value is positive. In the case of Jan, difference is small, so it is difficult to see difference at plot.

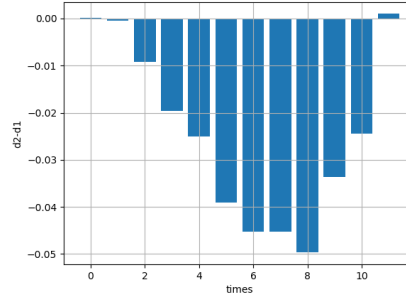


Figure 5: Density of snow at each resolution

10. problem8-extra

In order to know why Q1's density is higher than that of Q2 during most of time. I try to plot image from Jan to Dec (In the report, I just show part of it. Please check Interact code at 1-8(extra)). Figure 6 shows the trends of snowed place of Q1 and Q2 where Q1 is upper and Q2 is lower at each. The figure shows Q1's Greenland part's snow is not melted. Otherwise, most of snow at Q2 is melted. That's the reason, Q1's snow density is higher than that of Q2 during most of period.

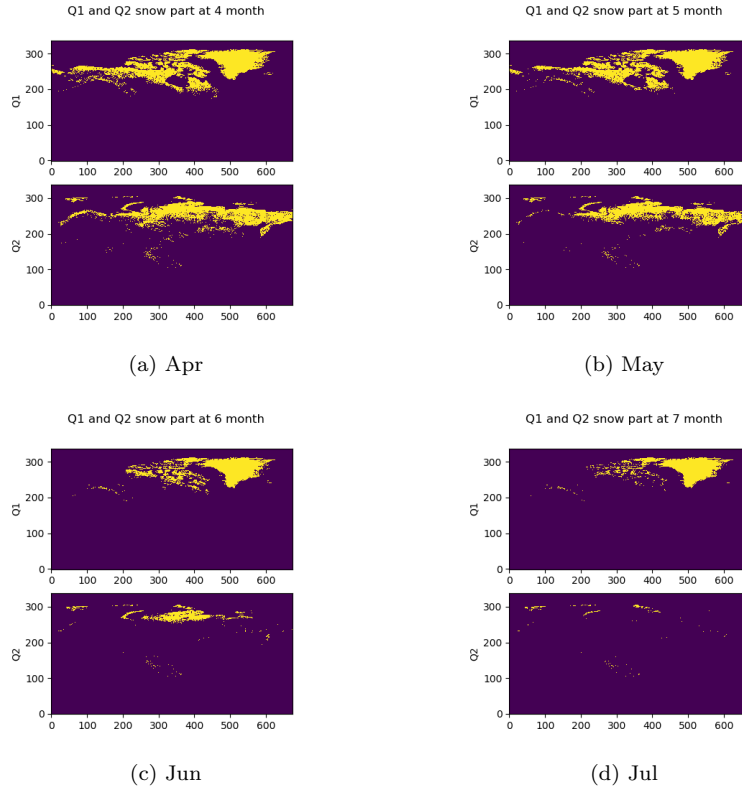


Figure 6: Snow part of Q1 and Q2 from Apr to Jul

## 2 Part2

1. problem1  
It is three dimensions which is x,y and time. And resolution is not related with the dimension.
2. problem2  
Embedding dimension is the dimension when it is visualized. Thus I think both situation's dimension is two. In the case of exercises 3, variables are time and resolution. But time is fixed at each image such as I have one image at one time. And resolution is not related with embedded dimension. So I think its dimension is two (2). And In the case of exercise 4, variables are time, resolution, and thresholds. I think threshold is not related with dimension. Thus its dimension also two (2).
3. problem3  
At resolution 5 memory usage is 54  
At resolution 9 memory usage is 726  
At resolution 13 memory usage is 10965  
At resolution 17 memory usage is 171366  
At resolution 21 memory usage is 2733750  
Please check code at Part2-3
4. problem4  
At resolution 5, it is just small number of grids so it is impossible to differentiate boarder. At resolution 9, it is possible to see the each continents but not exactly see where is the boarder. And at resolution 13, it is possible to see boarder, but I still can see the shape of pixel. From the resolution 17, everything is really clear.
5. problem5  
At resolution 5 memory usage is 648  
At resolution 9 memory usage is 8712  
At resolution 13 memory usage is 131580  
At resolution 17 memory usage is 2056392  
At resolution 21 memory usage is 32805000  
Each time step's byte of specific resolution values are same, for example from time=0 to at time=11, memory usage of resolution 5 for each time is 54. So result is just resolution of one time multiply 12. Please look at code Part2-5
6. problem6  
Resolution is important to measure the density of snow. And standard value (or density) to determine thresholds is important. In the homework, I know the standard density such as 5 to 15 to determine optimal threshold values, however if I don't know the values, it would be really difficult to find reasonable threshold values.