Q0

(a) DIF（Data Interchange Format）files . A DIF file is a spreadsheet supported by spreadsheet software such as Microsoft Excel. **Advantage:** DIF file can be viewed and edited with a text editor ,and user can export data from a spreadsheet and this spreadsheet can be imported by a variety of programs. **Disadvantage:** it does not support multiple spreadsheets. This means that the DIF file can only store one spreadsheet. Therefore, if you need to export multiple spreadsheets in the DIF format you will need to create multiple DIF files.

（b）Advantage: ① give each data an unique index so it is easy to screen throughout them and call them by their index.② type list is mutable thus we can change the original data type into the type we need easily.**Disadvantage:** The row index in the list of list is actually the vertical coordinates of map, and the column index is the horizontal coordinates, thus to represent a point's value whose coordinate is (x ,y), we need to write code like data[y][x], which may be confusing when the problem gets complicated.

(c) we can also represent the files in a dictionaries structure. The keys are map's coordinates and the value is the information we need.

Advantage: the values can be modified of change, and it is also iterable. **Disadvantage:** the keys are immutable and slicing over a dictionary type is not easy.

(d) I choose the dataset called ACT Water Features Line, because water lines can affect the risk and spread of bushfire .

Q1

a) every function in my answer to Q1 returns a data of type **list**

b) ① for function **load_vegetation_type,** I did nothing with blank account because we cannot create a vegetation type for an area if the data is missing. **So it just return '' for blank value.**

② for function **load_vegetation_density**, I presume **the blank values to be -1.0** because it can be distinguished from the original data where the cell value is 0 (the original dataset doesn't have value -1) and it can also be compared with other numerical data.

③ for function **load_wind_speed**, I presume the **blank values to be -1.0** because it can be distinguished from the original data where the cell value is 0 (the original dataset doesn't have value -1) and it can also be compared with other numerical data.

④ for function **load_bushfire,** I did nothing with blank account because it is reasonable to regard those areas as "NOT KNOW" instead of "DO HAPPEN" ,which is True, of "DO NOT HAPPEN" ,which is False, when considering the spread of bushfire or the existence of bushfire. **So it just return '' for blank value.**

c) There are 15 files in total. Thus three tables are built.

For **act dictionary,** the file and the corresponding blank value are:

vegetation_type.csv	95313
vegetation_density.csv	95313

wind.csv	86037
act_2003_bushfire.csv	95313
initial_2003_bushfire.csv	95313

For **south dictionary**, the file and the corresponding blank value are:

vegetation_type.csv	8348
vegetation_density.csv	8348
wind.csv	295
act_2003_bushfire.csv	95313
initial_2003_bushfire.csv	95313

For **anu dictionary**, the file and the corresponding blank value are:

vegetation_type.csv	1181
vegetation_density.csv	1181
wind.csv	0
act_2003_bushfire.csv	1181
initial_2003_bushfire.csv	1181

Q2

a) for act, the maximum wind speed is 8.60718
for south, the maximum wind speed is 7.0872
for anu, the maximum wind speed is 7.29959

b) The time complexity is T(n) = O(n^2)
and in this case, the n equals to the size of each side of the object map
Thus for anu map, the time complexity T(n) = 50^2 = 2500
Thus for south map, the time complexity T(n) = 150^2 = 22500
And for act map, the time complexity T(n) = 441*380 = 167580

Q3

a)

	anu	south	act
Open Woodland	50	5475	23077
Woodland	125	3182	11934

Grassland	65	2599	17291
Forest	370	44	3710
Open Forest	368	252	3732
Urban Vegetation	315	1866	4023
Golf Course	0	326	712
Pine Forest	0	325	6728
Shrubland	0	30	705
Arboretum	26	43	260

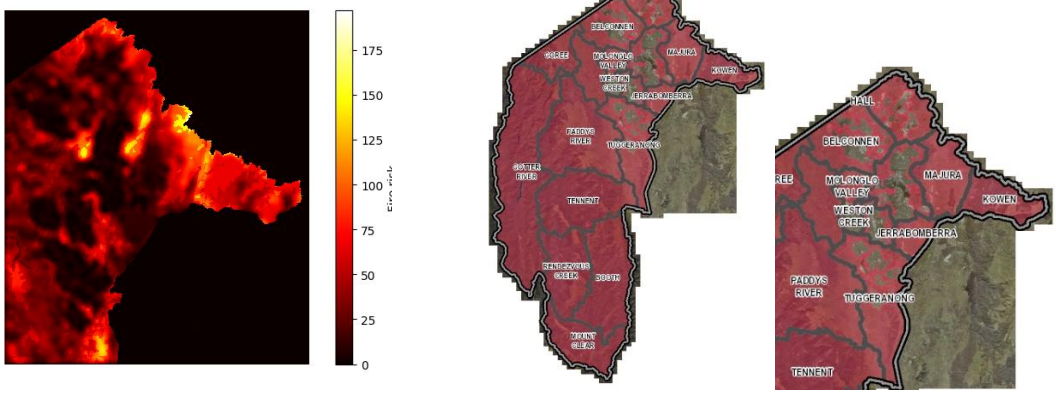
b)

	anu	south	act
Open Woodland	100000.0	10950000.0	46154000.0
Woodland	625000.0	15910000.0	59670000.0
Grassland	13000.0	519800.0	3458200.0
Forest	3700000.0	440000.0	37100000.0
Open Forest	2944000.0	2016000.0	29856000.0
Urban Vegetation	0.00	0.00	0.00
Golf Course	0.00	0.00	0.00
Pine Forest	0.00	0.00	0.00
Shrubland	0.00	0.00	0.00
Arboretum	0.00	0.00	0.00

c)In both function, I index the list for n^2 times.

Q4

a) My firerisk report of act is as below



When it is compared with BPA:
Similarities: the basic shape are similar.
Differences: there are more area inside the high-risk area where the fire risk is almost 0 in my simulation compared to BPA. Besides, the total area of BPA is larger than the act map in the data, The generated risk report map is like the upper right part of BPA

b) In this question, the edge of the map is not easy to handle. I handle this by dividing the whole map into 9 area, and write conditional code to tell which area contains the target cell.

1	2	3
4	5	6
7	8	9

n

n

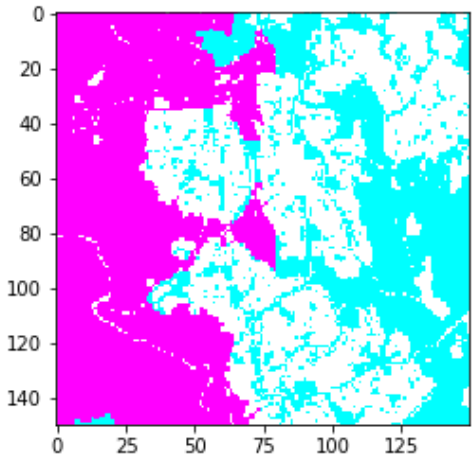
n equals to the integer part of wind speed. For example, if the target cell is in area 6, then we only consider all the other cells that have row index [y-8,y+8] and column index [x-8, len(column)] instead of considering all the cells in this map.

By doing this, we can eliminate most redundant and unnecessary calculation and comparison thus the efficiency will be greatly enhanced.

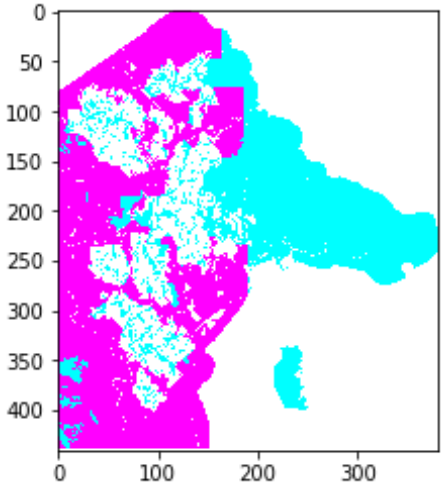
Q5

a) for south map, I simulate 78 steps
and for act map, I simulate 188 steps

b)

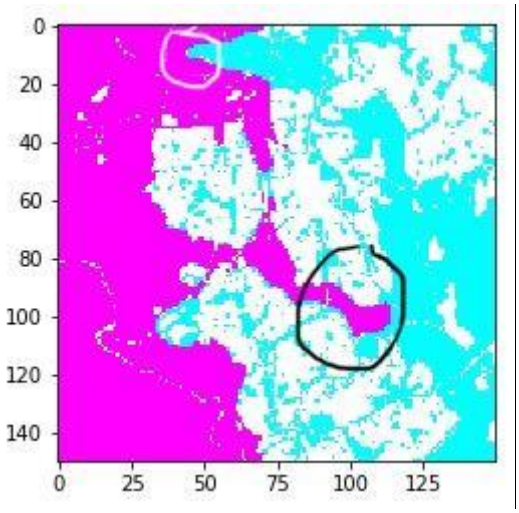


South_simulate



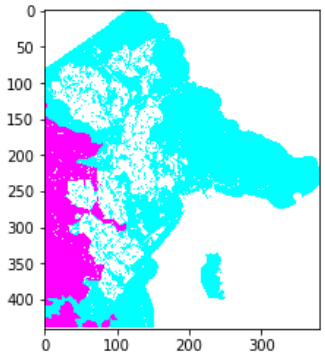
act_simulate

c)

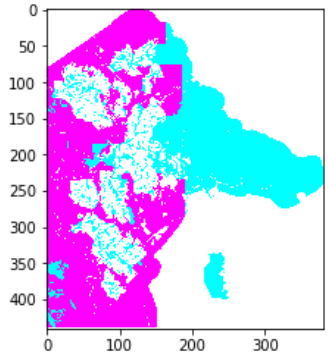


The basic shape are almost the same for two maps. But the black marked part of south_real stretches out while the white marked part of south_real was not on fire in real situation.

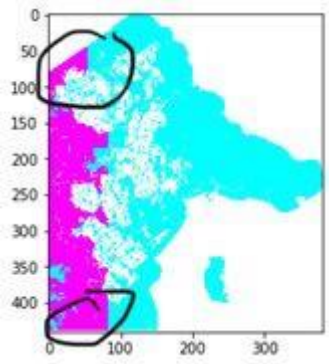
south_real



act_real



act_simu_step188



act_simu_step100

In act map, the real situation and my simulation is greatly different considering the total area of bushfire. But if we cut down the simulation step into 100, we can see that the simulation is better. Thus the main reason for this simulation failure is that in real situation, the bushfire won't spread to half of the map for act. But anyway, the direction of bushfire spread is similar between simulation and real situation except for the north and bottom south of the map.

Q6

a) the accuracy of south simulation map of Q5 is 95.15%

the accuracy of act simulation map of Q5 is 62.34%

b) for south map, we can see that it is a fairly good simulation.

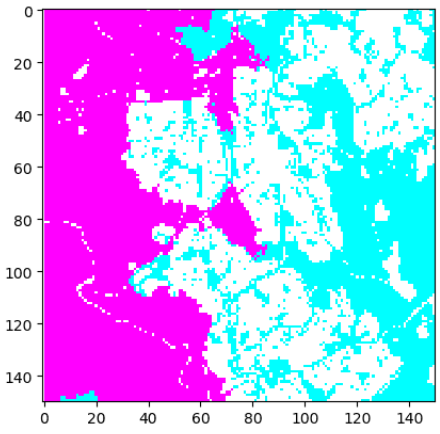
For act map, we can say that this simulation is not very good, because this simulation did not take real situation into consideration, which is for a large area like act, the probability of getting half of the area on fire is so rare. The bushfire will be under control before it spread out.

Q7

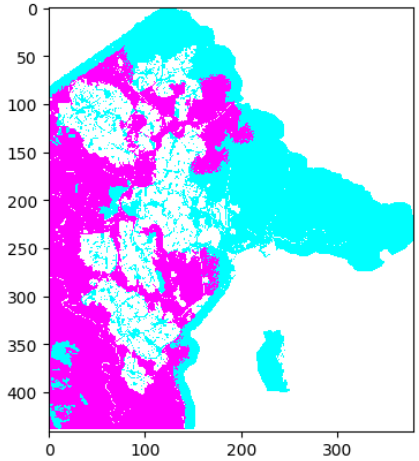
a) for south map, I simulate 105 steps.

For act map, I simulate 430 steps.

b)



South_simu_stochastic_105

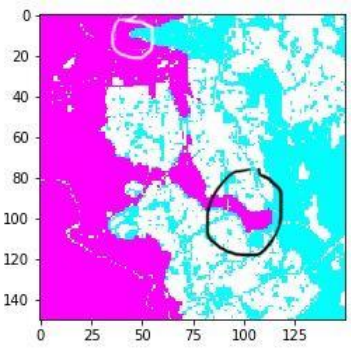


Act_simu_stochastic_430

c) Comparison with the real situation:

south map:

qualitatively:

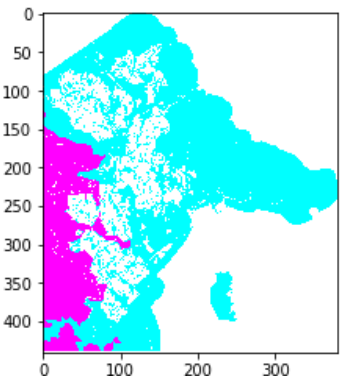


The basic shape are almost the same for two maps. But the black marked part of south_real stretches out while the white marked part of south_real was not on fire in real situation.

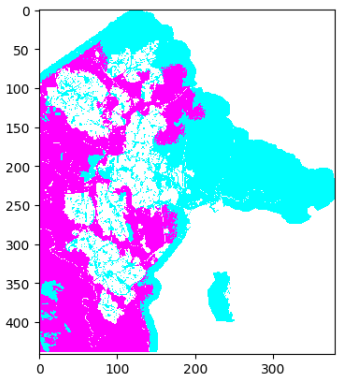
quantitatively: the accuracy of this 105 step stochastic simulation is 95.12%

act map:

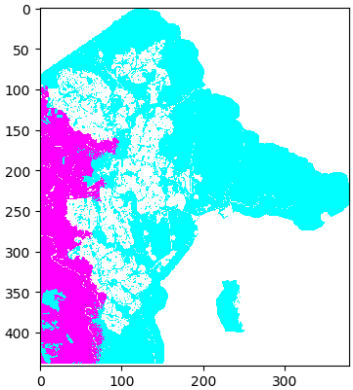
qualitatively:



Act_real



act_simu_stochastic_430



act_simu_stochastic_150

In act map, the real situation and my simulation is greatly different considering the total area of bushfire. But if we cut down the simulation step to 150 steps , we can see that the simulation is much better. Thus the main reason for this simulation failure is that the simulation step is far greater than it actually needs in real situation.

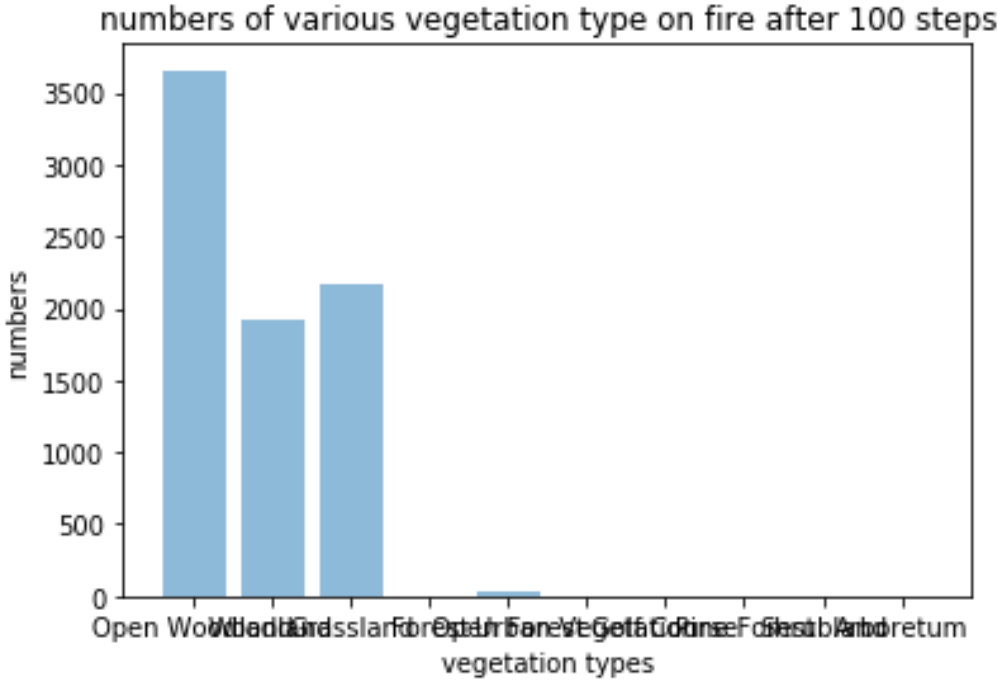
quantitatively: the accuracy of this 105 step stochastic simulation is 70.70%

d) For south map, I think it is a realistic simulation; and for act map, I think it's not a good realistic simulation. But if we can cut down the simulation steps, it could be a realistic simulation.

Q8

a)

for south map:



For act map: