SOM

2023-07-30

```
library(paletteer)
library(kohonen)
library(ggplot2)
library(datawizard)
```

```
data = read.csv('knn.csv') # Same as data located at https://raw.githubusercontent.com/Aser-Abdelfatah/
data <- data[-1]
head(data, 5)
```

```
##           x        y   aspect  curvature      d2m       dem ignition_points
## 1 -9.480614 38.75799 218.3233 -13973.861 282.9851  22.81842               0
## 2 -9.480614 38.76806 238.1447   4664.384 283.8849 212.68144               0
## 3 -9.470545 38.72779 285.1091  -2931.347 283.6486  22.93496               0
## 4 -9.470545 38.73786 258.7989  -5164.504 282.9717  18.10819               0
## 5 -9.470545 38.74792 234.1072  -6697.967 283.2570  34.61827               0
##         lai lc_agriculture lc_forest lc_grassland lc_settlement lc_shrubland
## 1 1.5604380     0.06186711 0.3587750   0.00000000    0.00000000   0.03465905
## 2 3.7000000     0.22247930 0.5011652   0.00000000    0.03712273   0.20172146
## 3 0.6288601     0.09432729 0.2896737   0.00000000    0.29954389   0.00000000
## 4 1.7882567     0.05526726 0.4675248   0.11053451    0.00000000   0.00000000
## 5 1.7178869     0.07792080 0.6105175   0.01284893    0.00000000   0.00000000
##   lc_sparse_vegetation lc_water_bodies lc_wetland  lst_day lst_night      ndvi
## 1           0.00000000       0.4418382 0.10286062 296.9296  288.0320 0.5803687
## 2           0.00000000       0.0000000 0.03751133 301.2200  291.5000 0.5433000
## 3           0.15037109       0.1660841 0.00000000 302.1800  289.7200 0.3559000
## 4           0.08725917       0.2794143 0.00000000 296.2039  289.9400 0.6409407
## 5           0.00000000       0.2987128 0.00000000 298.4094  287.9481 0.6247345
##   population        rh    slope     smi       sp    ssrd      t2m           tp
## 1   1.502658 0.2307284 1.558401 0.13765 96798.68 8743087 301.9096 2.902481e-04
## 2  54.910294 0.1980689 1.557888 0.13765 96990.47 8991473 304.0639 8.134970e-05
## 3  20.905184 0.2114293 1.537272 0.15544 96903.46 9025677 302.1347 2.370733e-05
## 4   4.693551 0.2471087 1.542442 0.15544 96842.64 8829953 301.0579 1.120156e-04
## 5   1.748590 0.2201719 1.553220 0.13765 96754.40 8826432 302.4637 1.780292e-04
##   wind_direction wind_speed year cos_day_of_the_year sin_day_of_the_year
## 1       335.4230   5.275411 2018          0.07309513           -0.997325
## 2       310.0422   4.687839 2018          0.07309513           -0.997325
## 3       345.9933   5.415241 2018          0.07309513           -0.997325
## 4       317.3863   4.934384 2018          0.07309513           -0.997325
## 5       316.8602   5.280971 2018          0.07309513           -0.997325
##   d2m_mean wind_speed_mean wind_direction_mean  sp_mean ssrd_mean t2m_mean
## 1 285.0277        4.736493            270.2619 96643.87   8244722 292.7158
## 2 285.1447        4.527946            268.0604 96855.20   8313999 292.6595
## 3 285.0955        4.831436            268.3753 96938.47   8340271 292.7061
## 4 285.0862        4.492348            259.7677 96719.48   8310035 292.7265
## 5 285.0046        4.727574            267.2218 96623.46   8251281 292.6960
```

```
##       tp_mean   rh_mean lst_day_mean lst_night_mean cos_month  sin_month
## 1 0.004570389 0.5598037     293.0544       284.8462       0.5 -0.8660254
## 2 0.004673695 0.5647932     292.7440       285.3649       0.5 -0.8660254
## 3 0.004115293 0.5581878     293.2422       286.1022       0.5 -0.8660254
## 4 0.004895319 0.5596613     293.3037       286.5644       0.5 -0.8660254
## 5 0.004376948 0.5575282     293.8945       284.3868       0.5 -0.8660254
##   repetitions roads_distance
## 1           1          0.000
## 2           1          0.172
## 3           1          0.000
## 4           1          0.000
## 5           1          0.000
```

```
# normalizes data
data <- normalize(data)
```

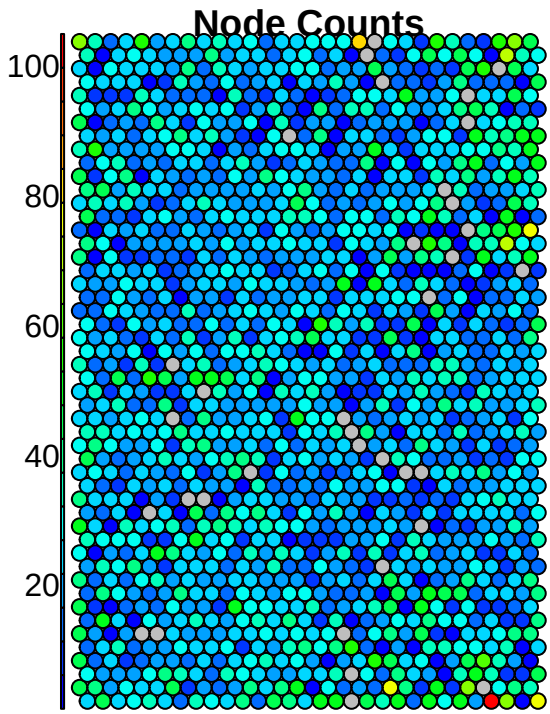
```
som_grid <- somgrid(xdim = 30, ydim=50, topo="hexagonal")
# The dataframe has to be transformed to matrix
data00<-as.matrix(data)
set.seed(73) # Set the seed of R‘s random number generator, which is useful for
#creating simulations or random objects that can be reproduced.
SOM_00 <- som(data00, grid=som_grid, rlen=200)
summary(SOM_00)
```

```
## SOM of size 30x50 with a hexagonal topology and a bubble neighbourhood function.
## The number of data layers is 1.
## Distance measure(s) used: sumofsquares.
## Training data included: 34569 objects.
## Mean distance to the closest unit in the map: 0.17.
```

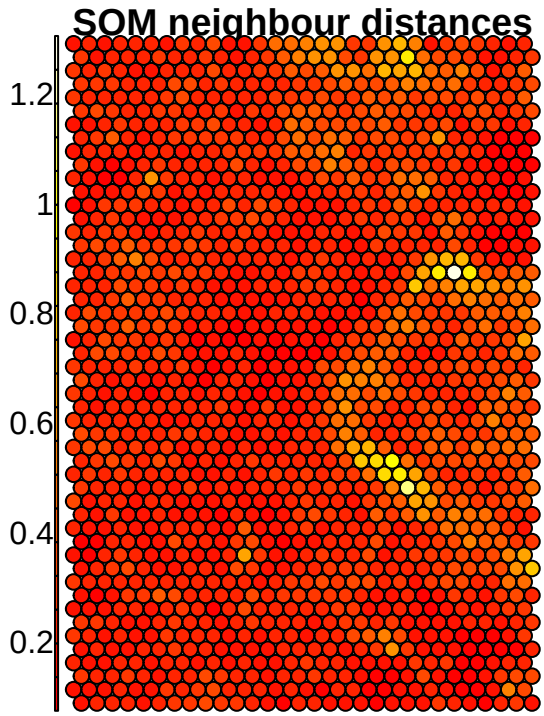
```
plot(SOM_00, type="changes")
```


```
coolBlueHotRed <- function(n, alpha = 1) {rainbow(n, end=4/6, alpha=alpha)[n:1]}
```

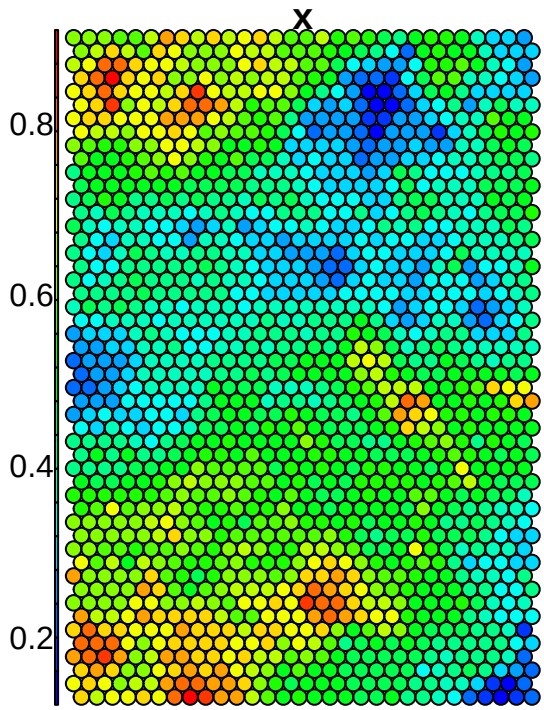
```
plot(SOM_00, type="count", main="Node Counts", palette.name=coolBlueHotRed)
```

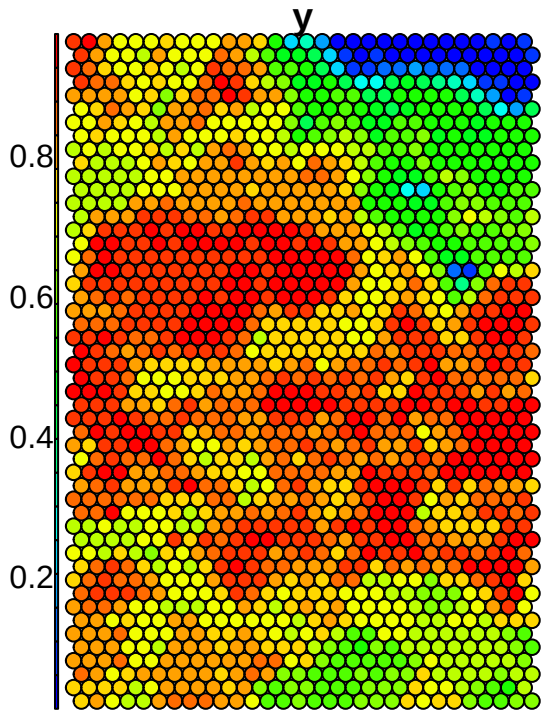


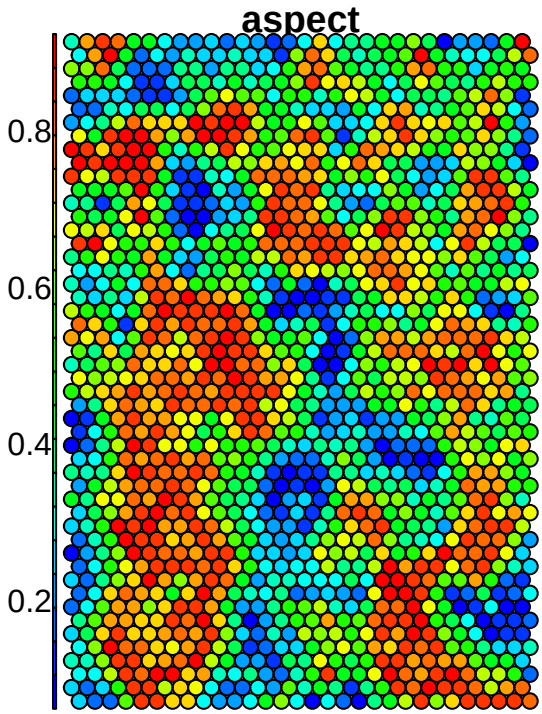

```
plot(SOM_00, type="dist.neighbours", main = "SOM neighbour distances")
```

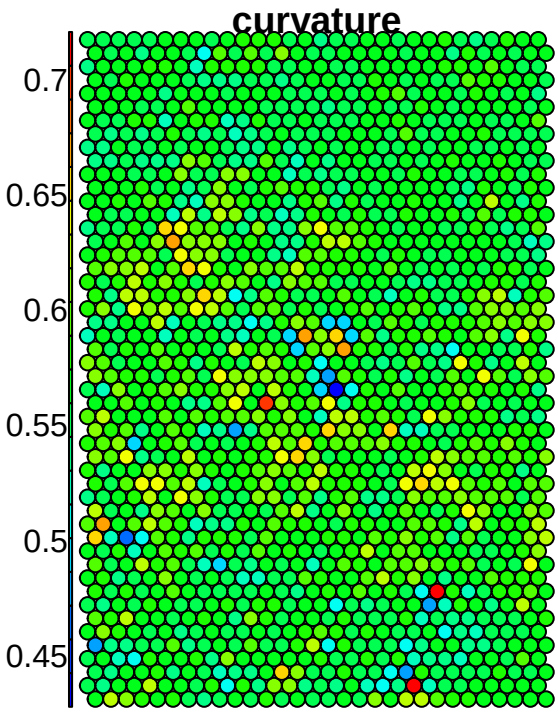


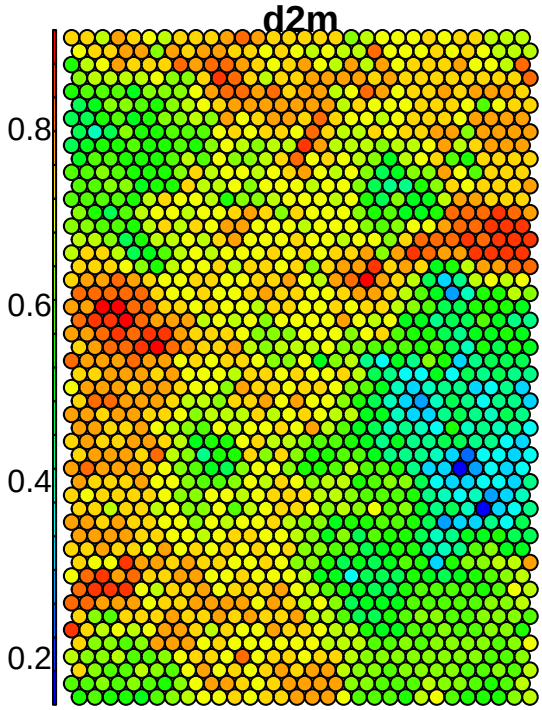

```
for (i in 1:46){
plot(SOM_00, type = "property", property = getCodes(SOM_00)[,i],
main=colnames(getCodes(SOM_00))[i], palette.name=coolBlueHotRed)}
```

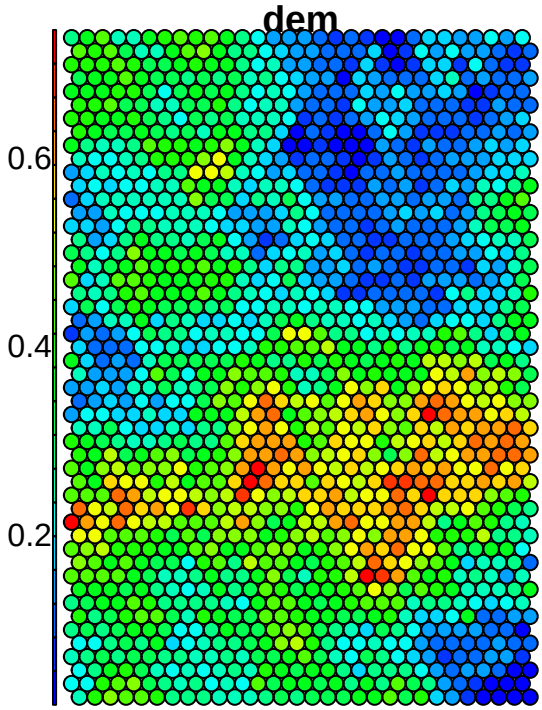

x
0.8
0.6
0.4
0.2

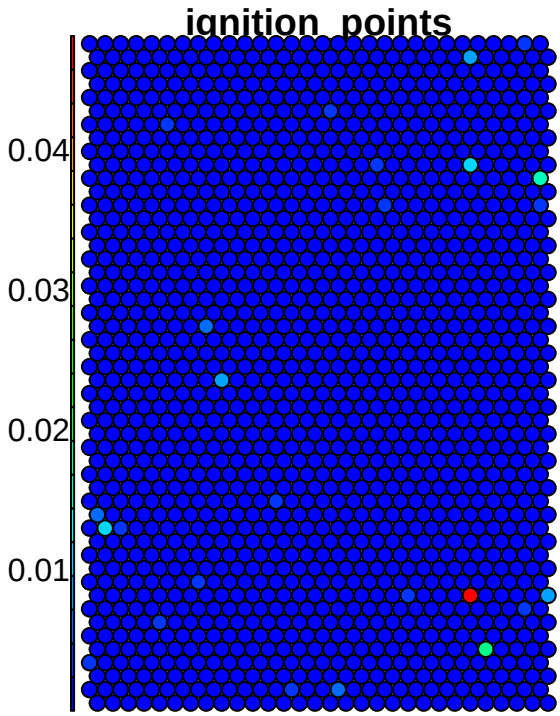

y
0.8
0.6
0.4
0.2

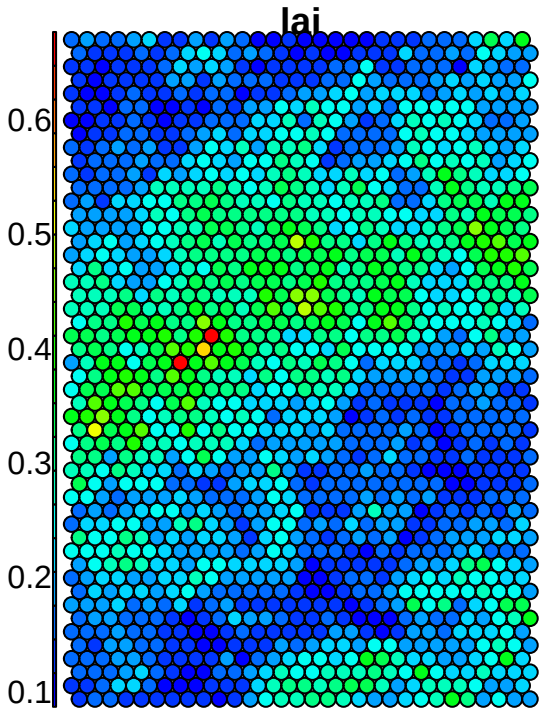

aspect
0.8
0.6
0.4
0.2

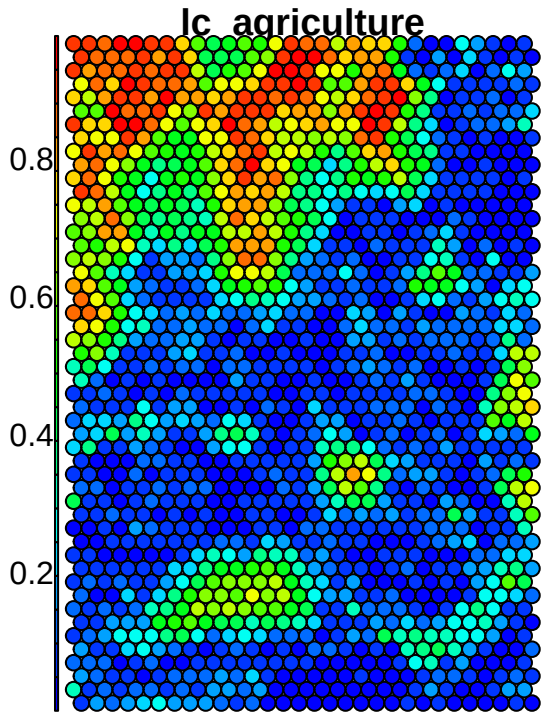

curvature
0.7
0.65
0.6
0.55
0.5
0.45

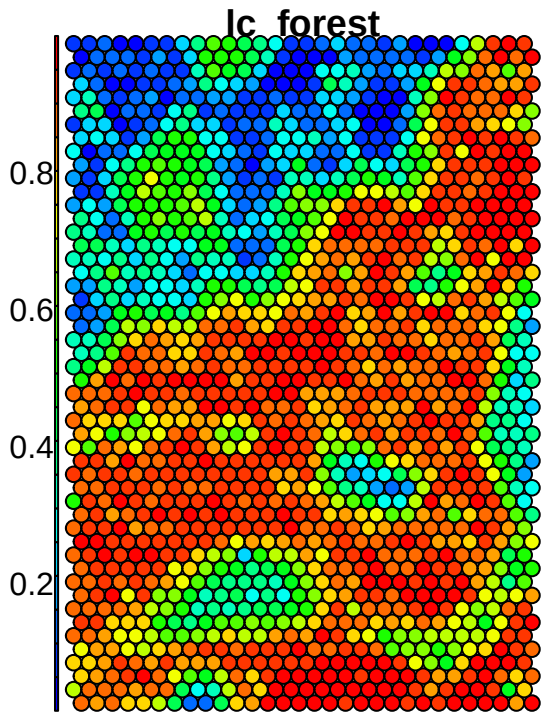

d2m
0.8
0.6
0.4
0.2

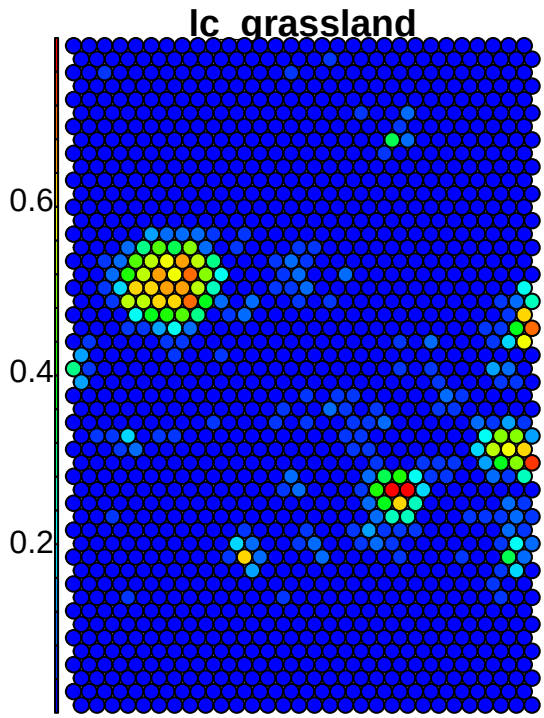

dem
0.6
0.4
0.2

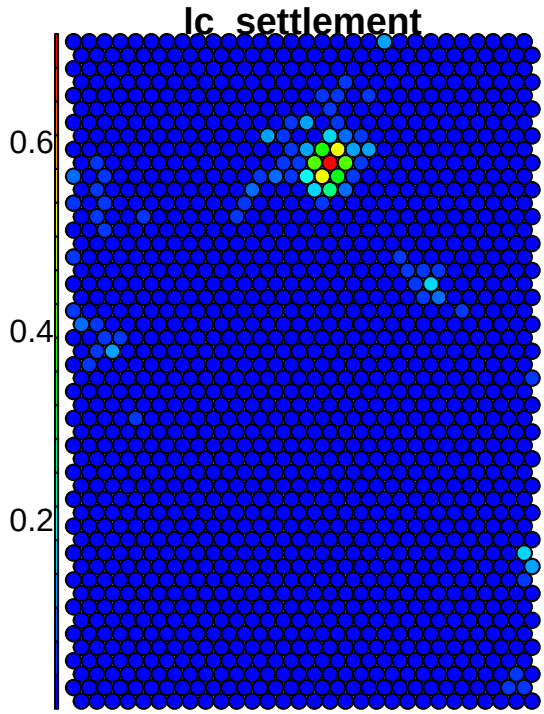

ignition_points
0.04
0.03
0.02
0.01

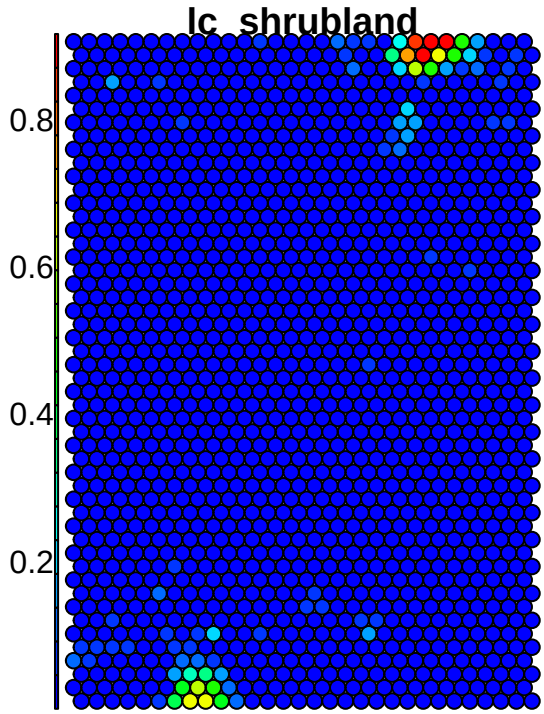

lai
0.6
0.5
0.4
0.3
0.2
0.1

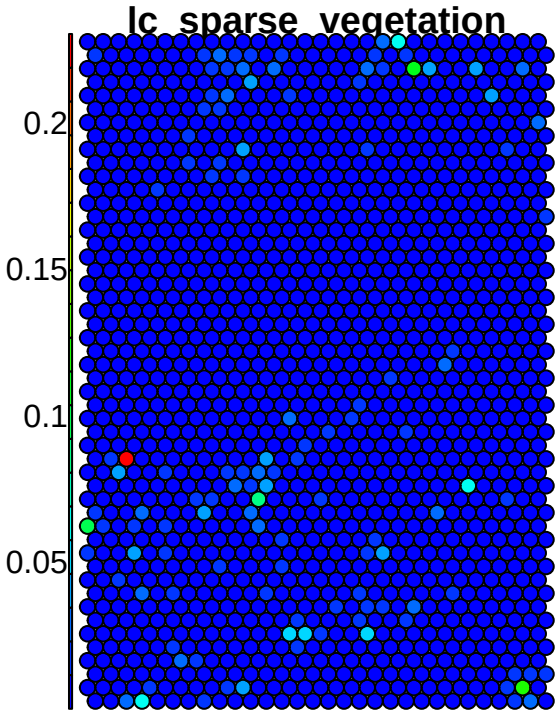

lc_agriculture
0.8
0.6
0.4
0.2

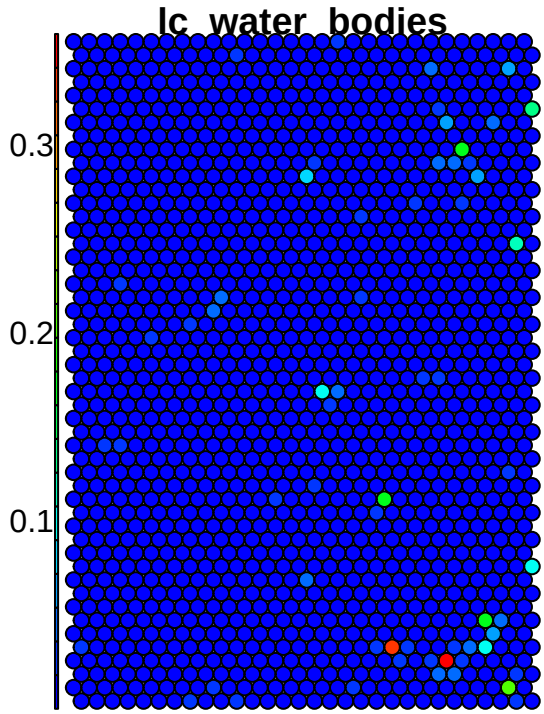

lc_forest
0.8
0.6
0.4
0.2

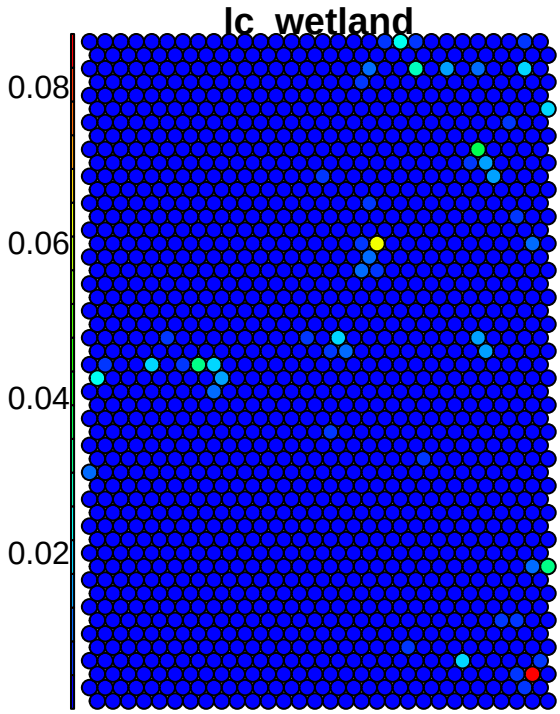

lc_grassland
0.6
0.4
0.2

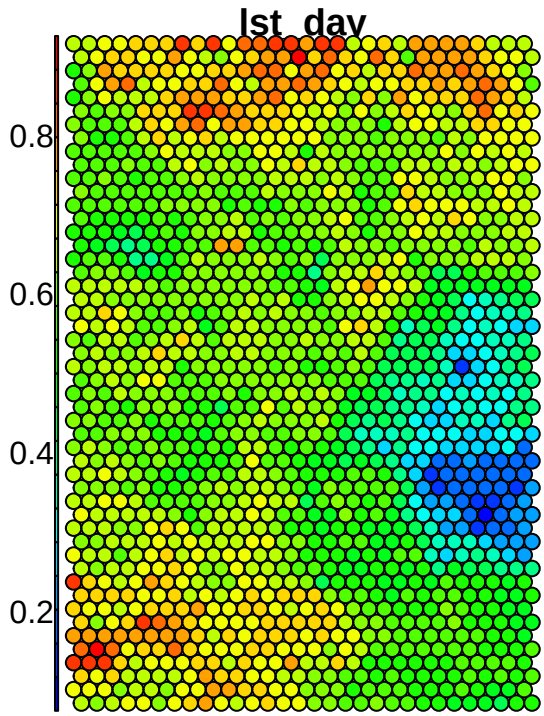

lc_settlement
0.6
0.4
0.2

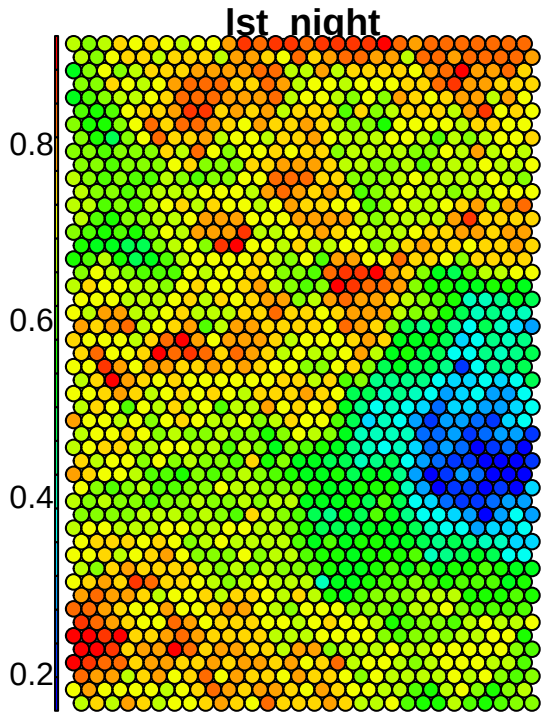

lc_shrubland
0.8
0.6
0.4
0.2

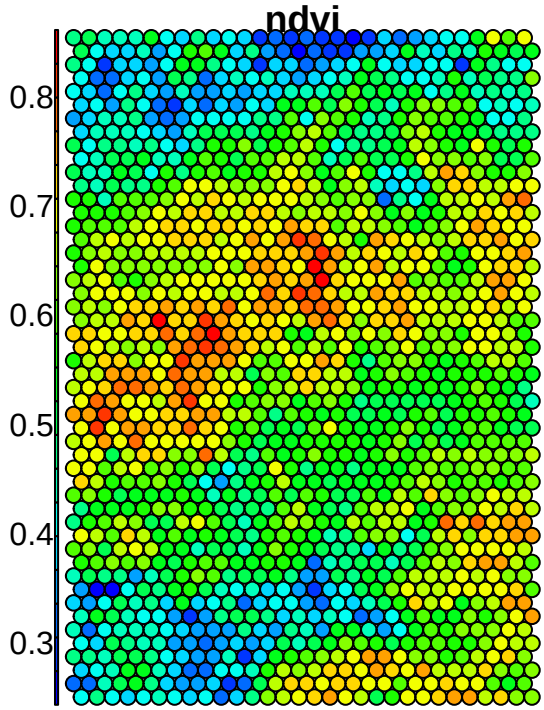

lc_sparse_vegetation
0.2
0.15
0.1
0.05

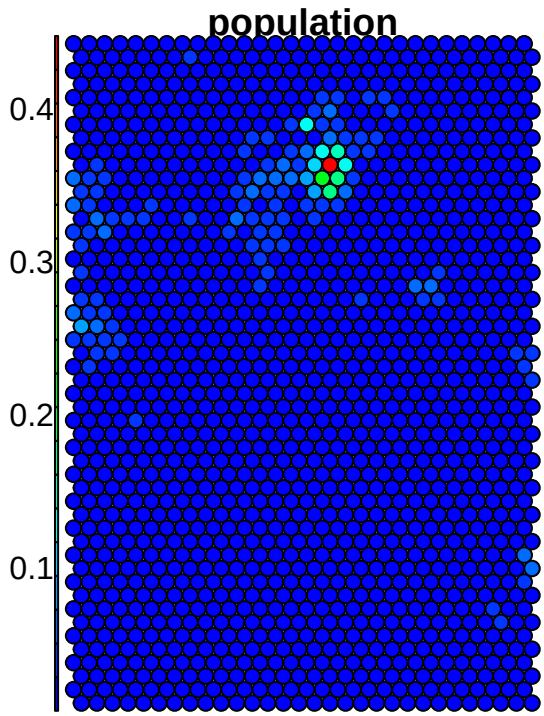

lc_water_bodies
0.3
0.2
0.1

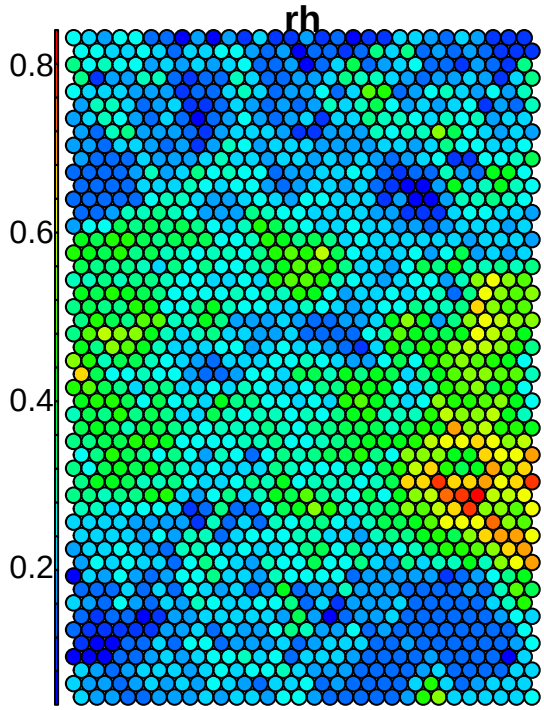

lc_wetland
0.08
0.06
0.04
0.02

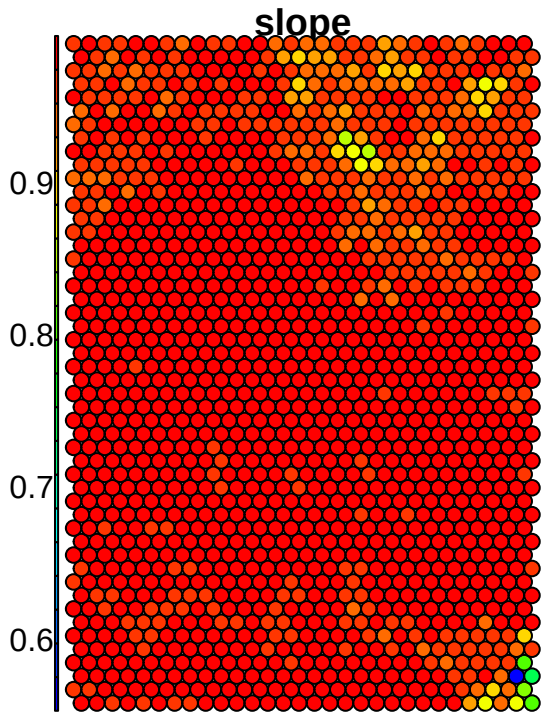

lst_day
0.8
0.6
0.4
0.2

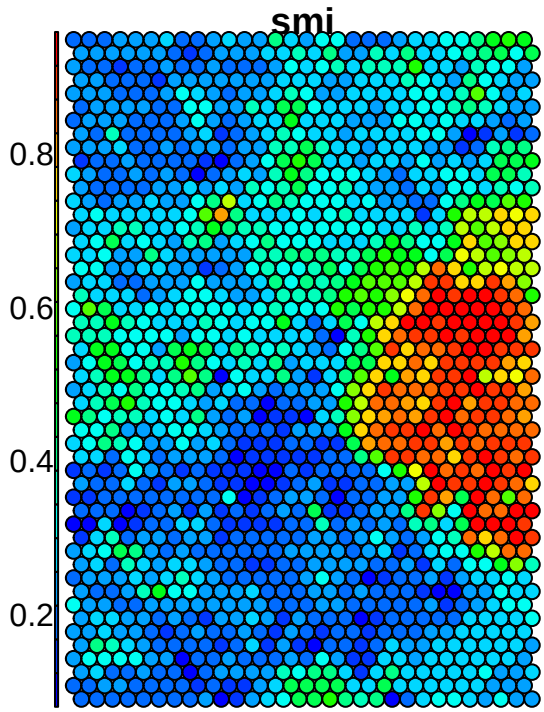

lst_night
0.8
0.6
0.4
0.2

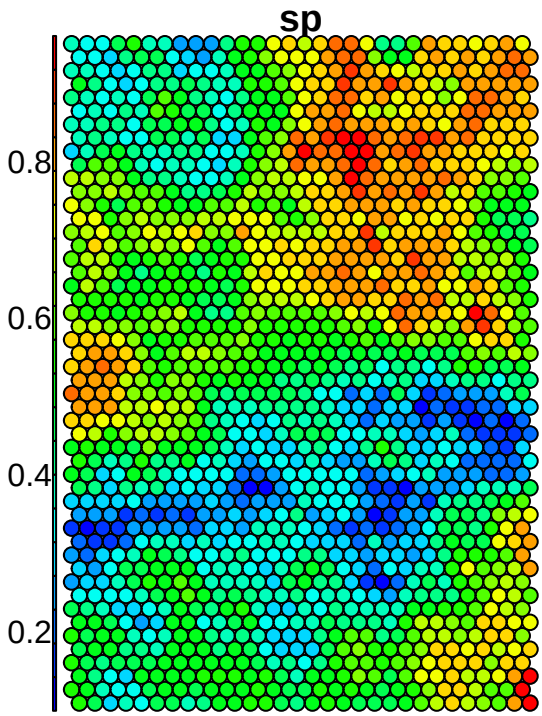

ndvi
0.8
0.7
0.6
0.5
0.4
0.3

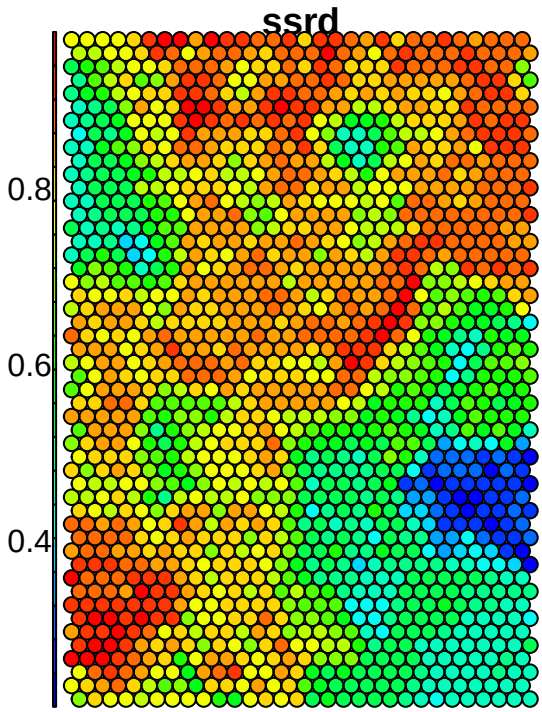

population
0.4
0.3
0.2
0.1

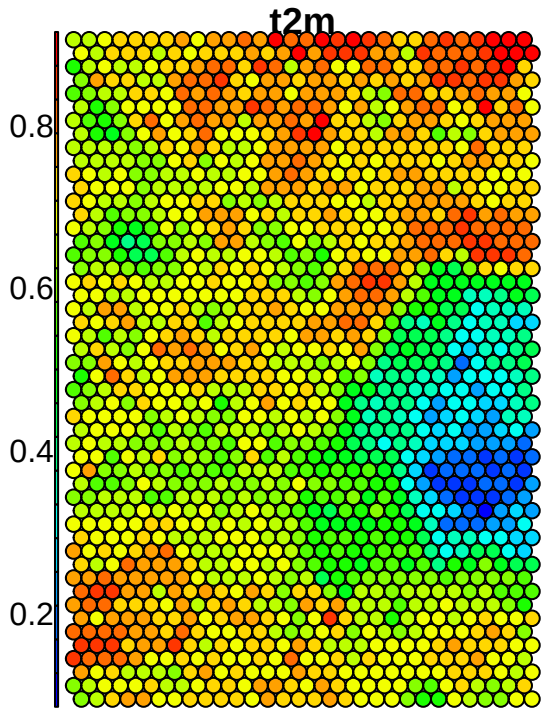

rh
0.8
0.6
0.4
0.2

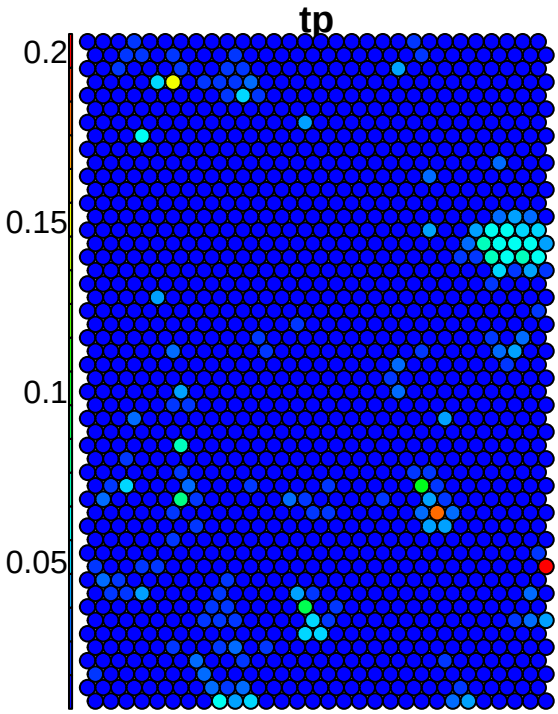

slope
0.9
0.8
0.7
0.6

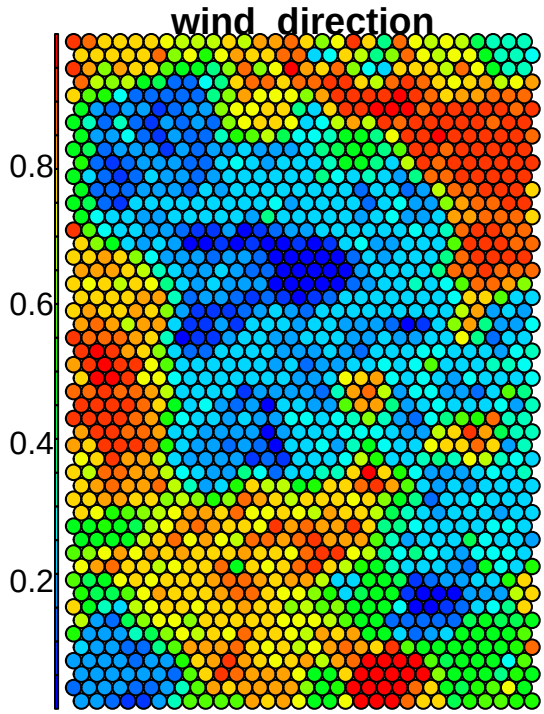

smi
0.8
0.6
0.4
0.2

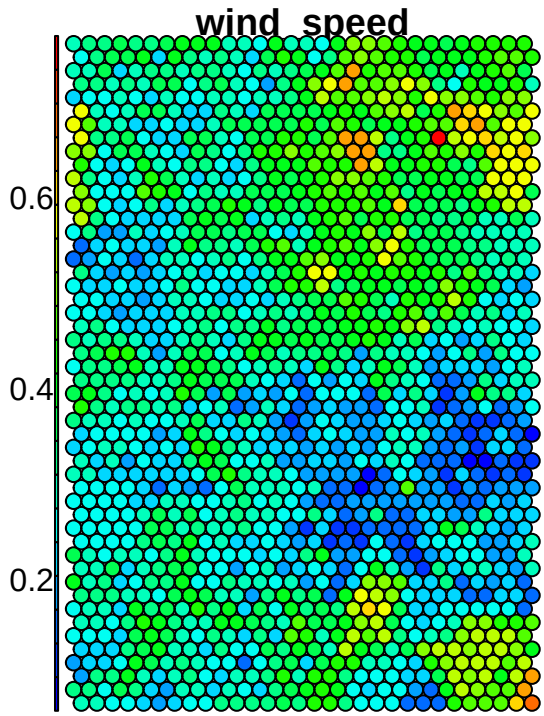

sp
0.8
0.6
0.4
0.2

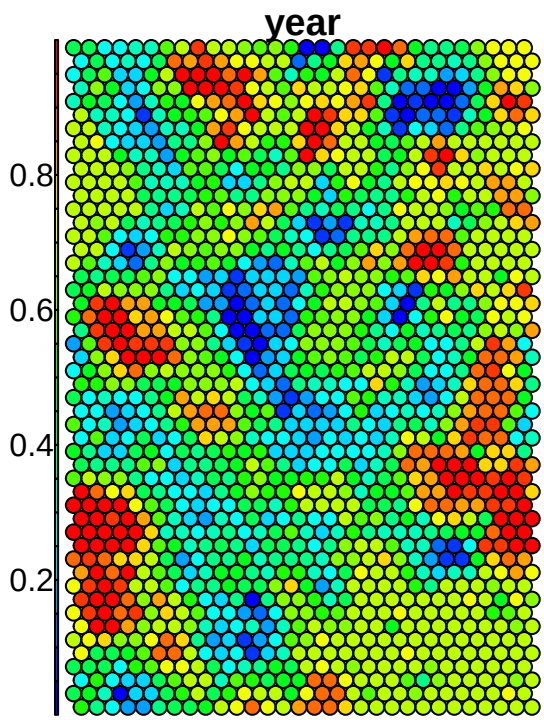

ssrd
0.8
0.6
0.4

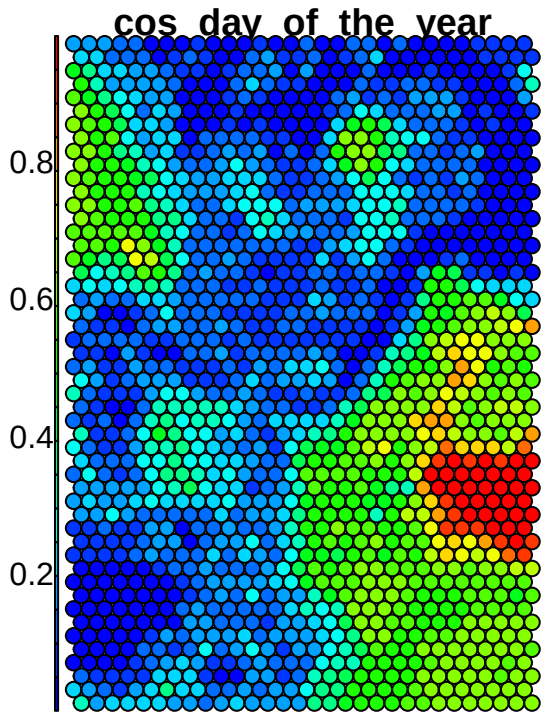

t2m
0.8
0.6
0.4
0.2

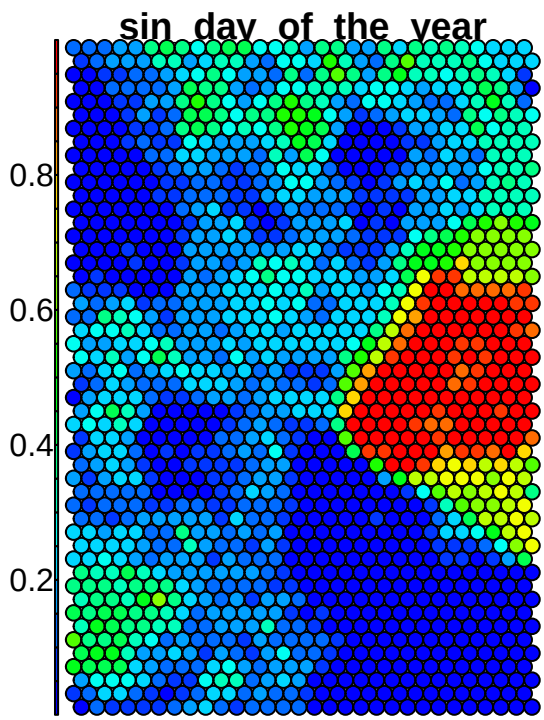

tp
0.2
0.15
0.1
0.05

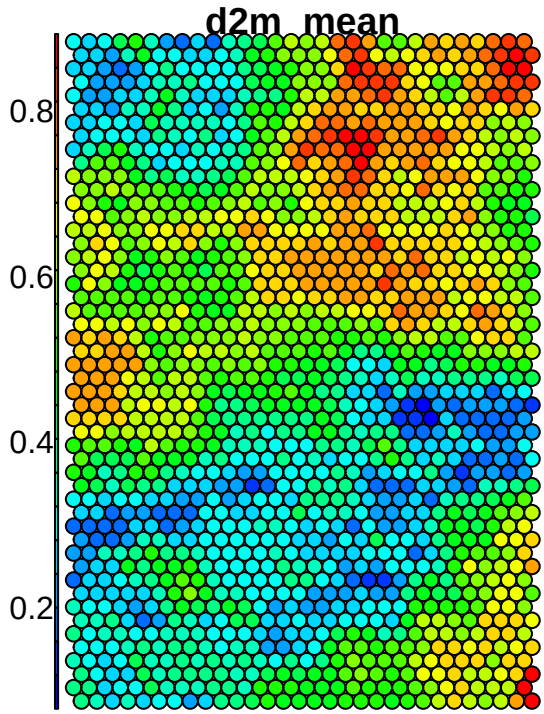

wind_direction
0.8
0.6
0.4
0.2

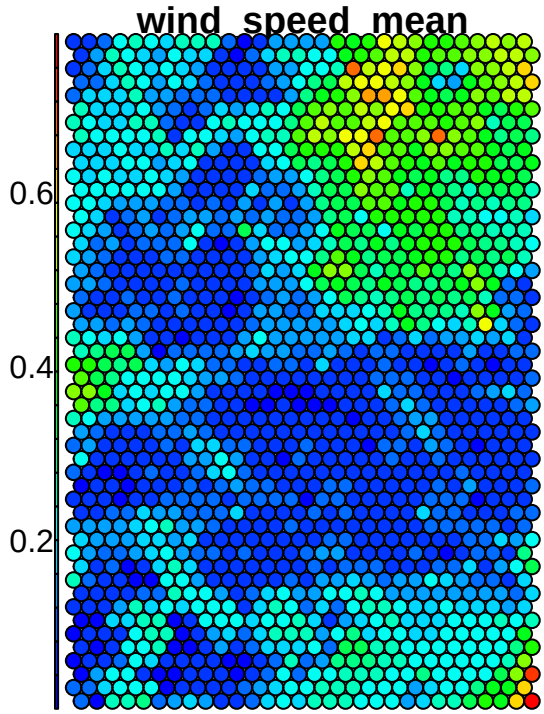

wind_speed
0.6
0.4
0.2

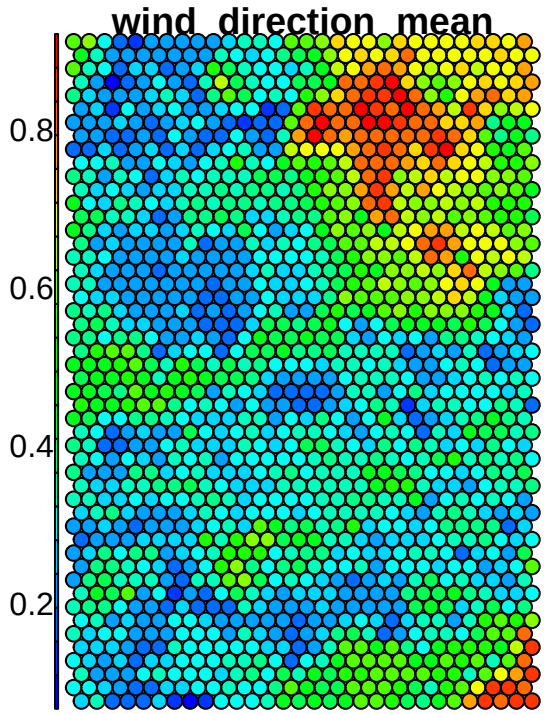

year
0.8
0.6
0.4
0.2

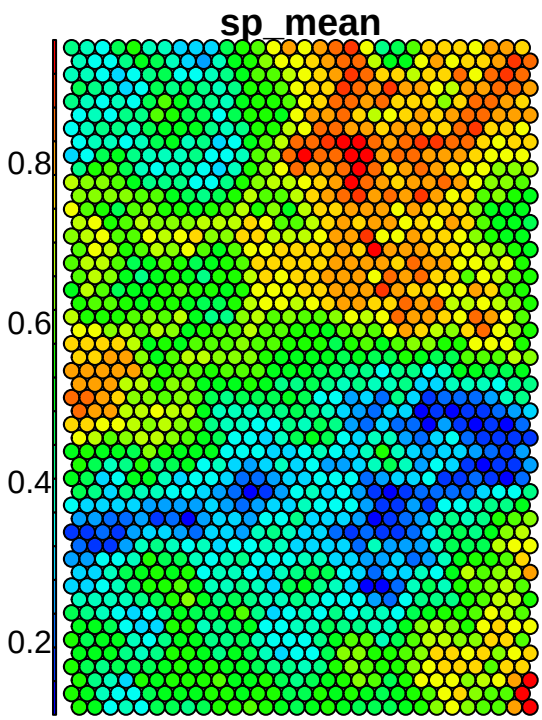

cos_day_of_the_year
0.8
0.6
0.4
0.2

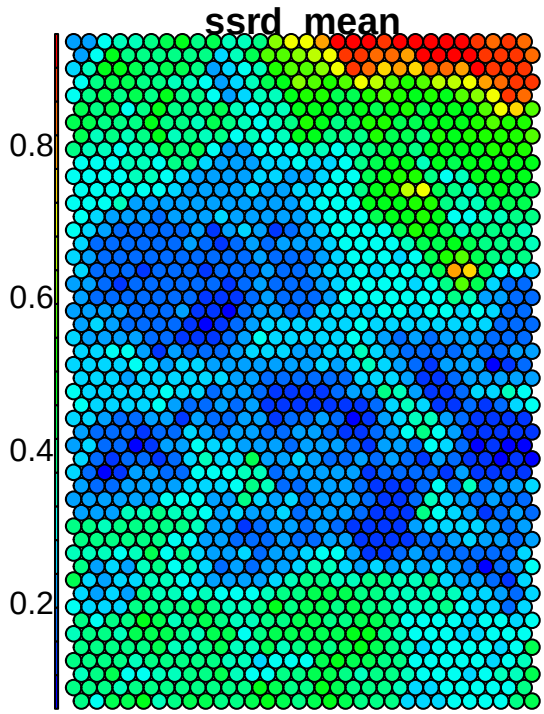

sin_day_of_the_year
0.8
0.6
0.4
0.2

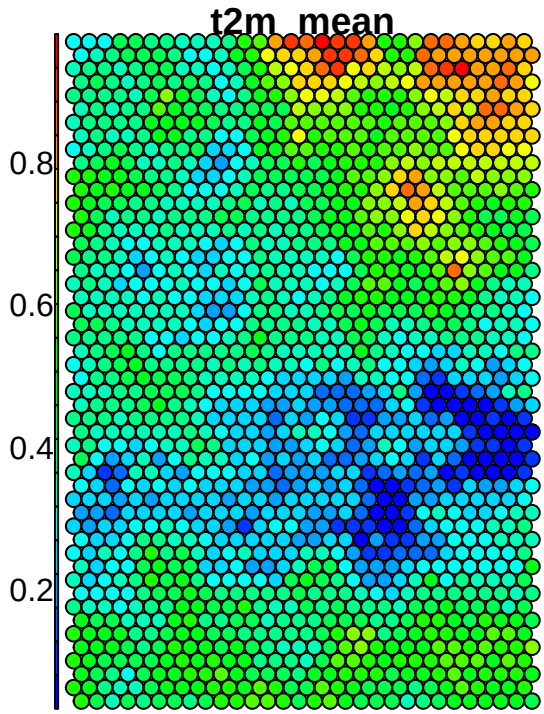

d2m_mean
0.8
0.6
0.4
0.2

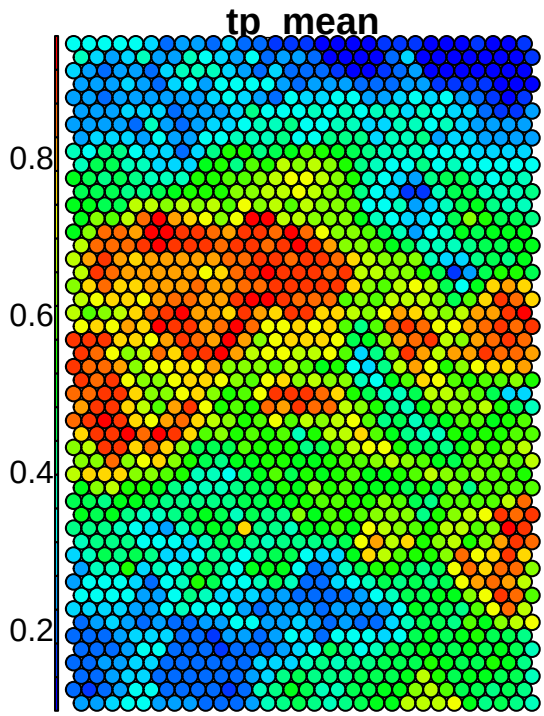

wind_speed_mean
0.6
0.4
0.2

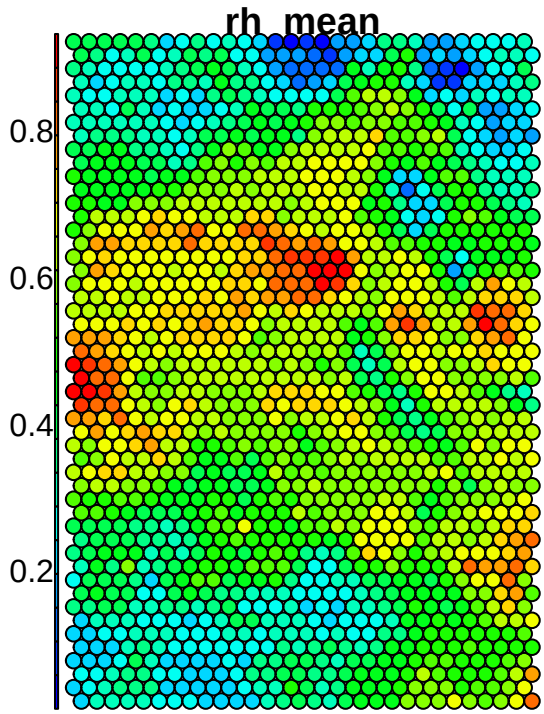

wind_direction_mean
0.8
0.6
0.4
0.2

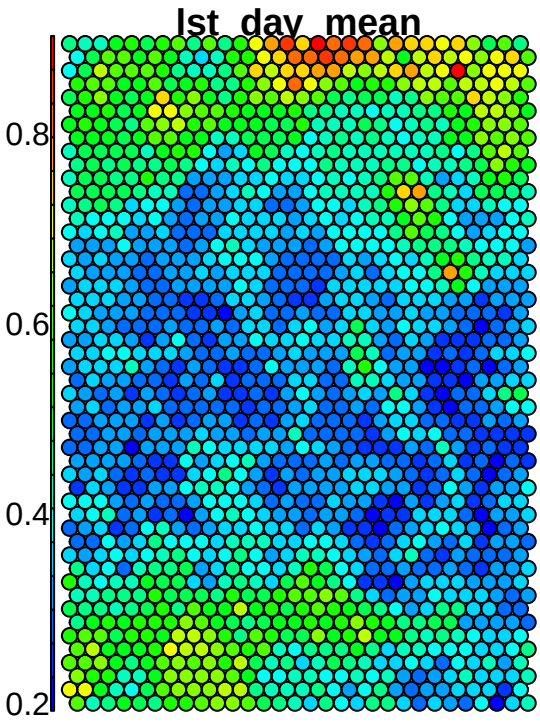

sp_mean
0.8
0.6
0.4
0.2

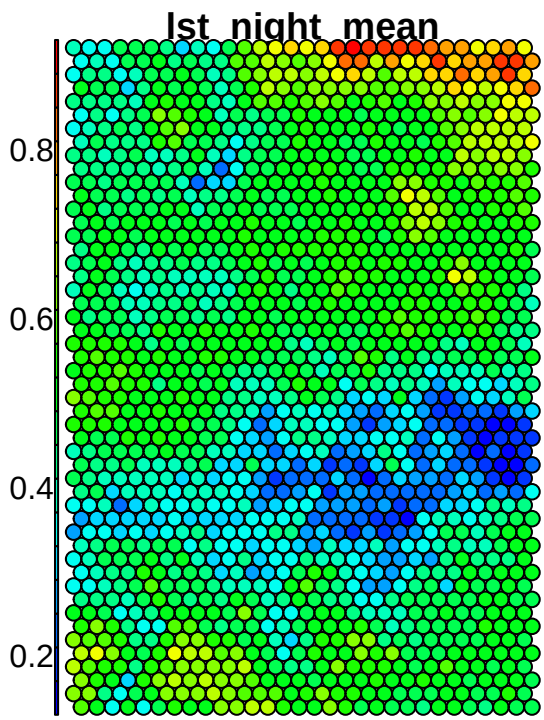

ssrd_mean
0.8
0.6
0.4
0.2

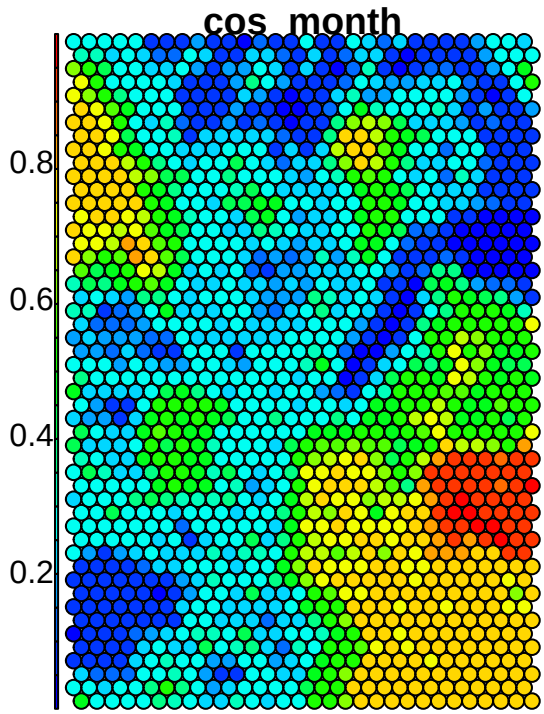

t2m_mean
0.8
0.6
0.4
0.2

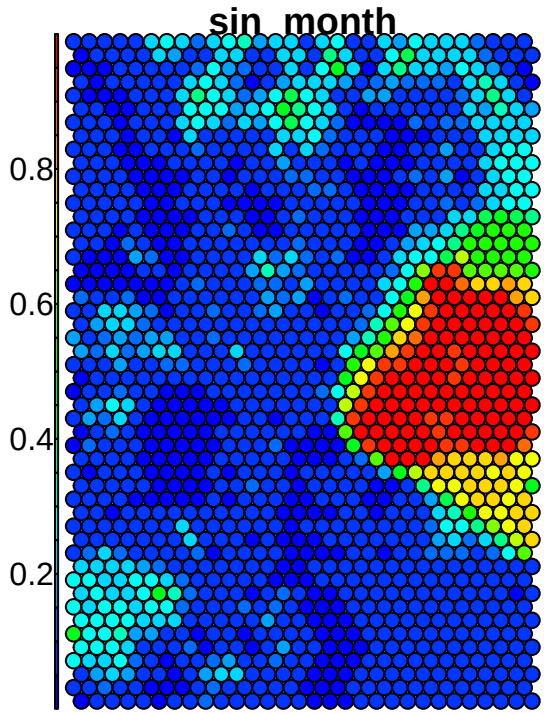

tp_mean
0.8
0.6
0.4
0.2

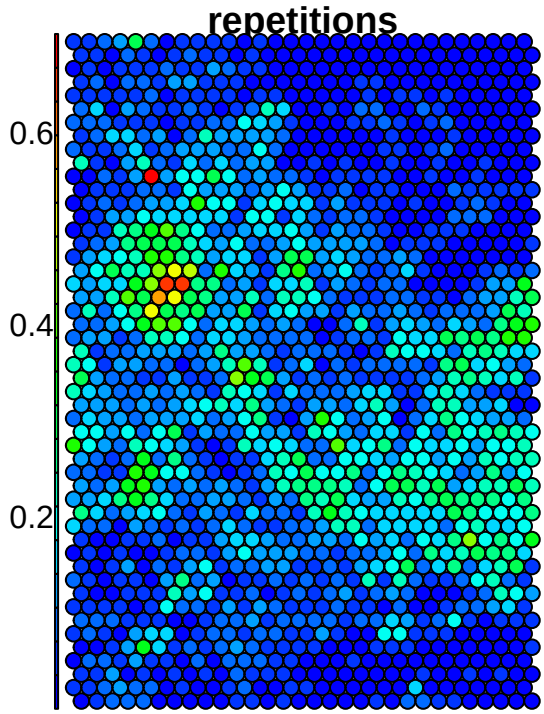

rh_mean
0.8
0.6
0.4
0.2

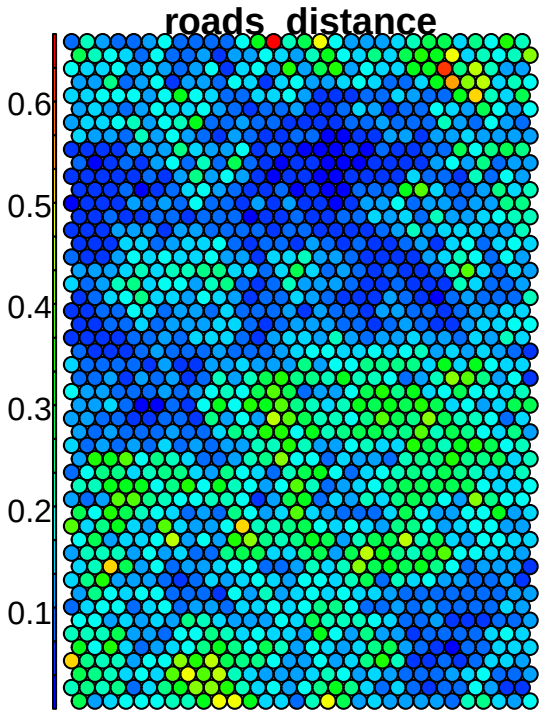

lst_day_mean
0.8
0.6
0.4
0.2


lst_night_mean
0.8
0.6
0.4
0.2


cos_month
0.8
0.6
0.4
0.2


sin_month
0.8
0.6
0.4
0.2


repetitions
0.6
0.4
0.2

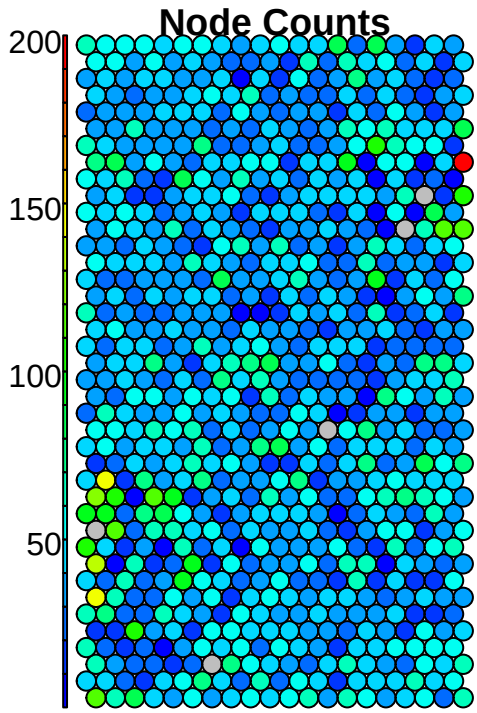



```
som_grid_small <- somgrid(xdim = 20, ydim=40, topo="hexagonal")
SOM_01 <- som(data00, grid=som_grid_small, rlen=200)
summary(SOM_01)
```

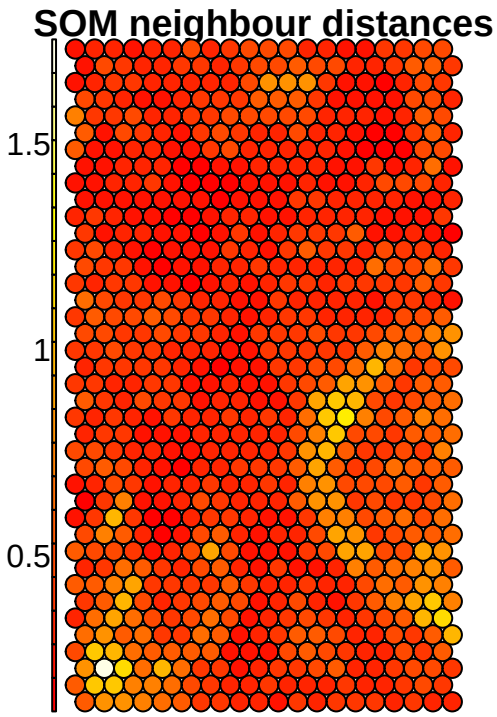
```
## SOM of size 20x40 with a hexagonal topology and a bubble neighbourhood function.
## The number of data layers is 1.
## Distance measure(s) used: sumofsquares.
## Training data included: 34569 objects.
## Mean distance to the closest unit in the map: 0.203.
```

```
plot(SOM_01, type="changes")
```

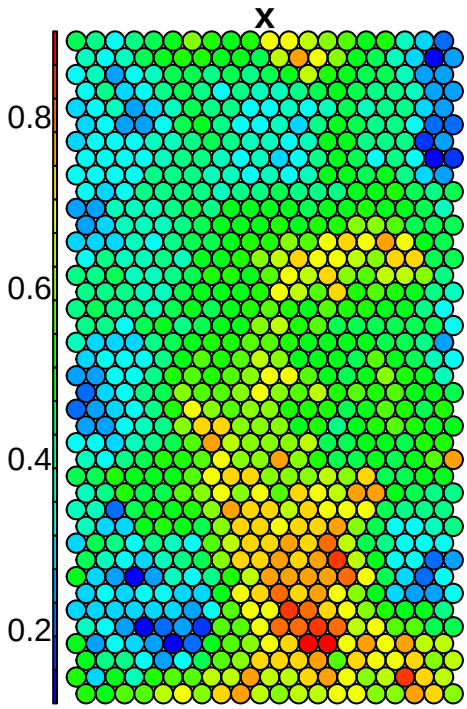

```
plot(SOM_01, type="count", main="Node Counts", palette.name=coolBlueHotRed)
```

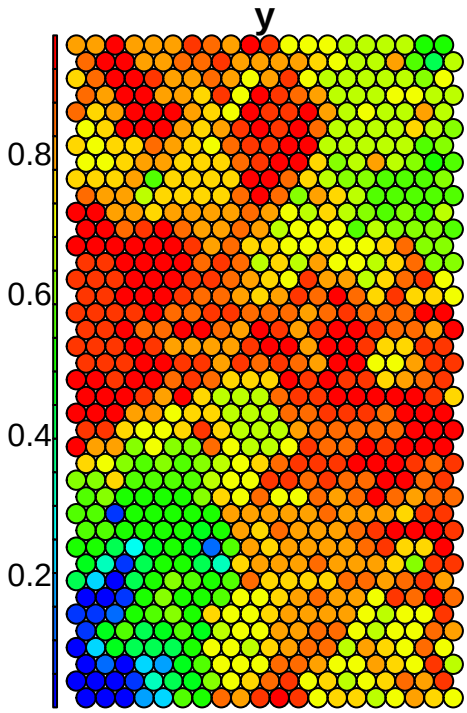


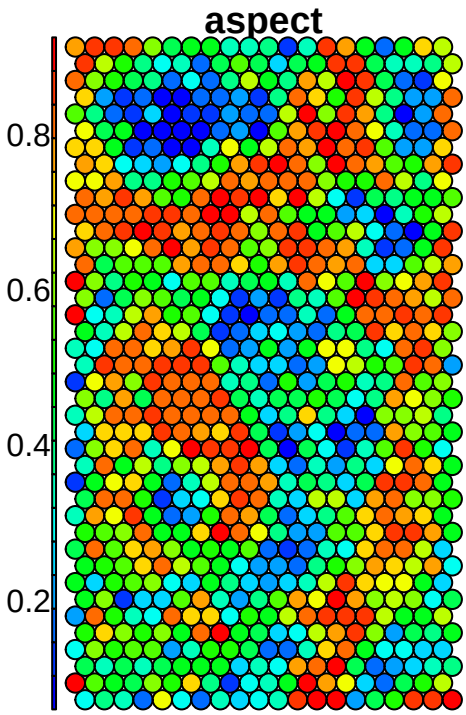

```
plot(SOM_01, type="dist.neighbours", main = "SOM neighbour distances")
```

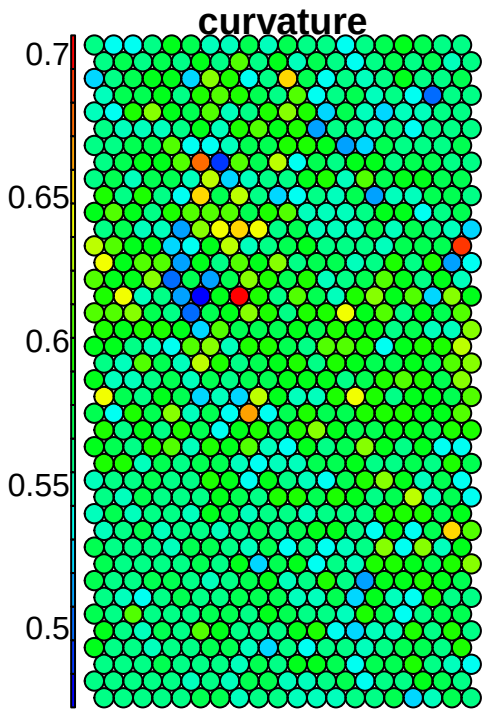


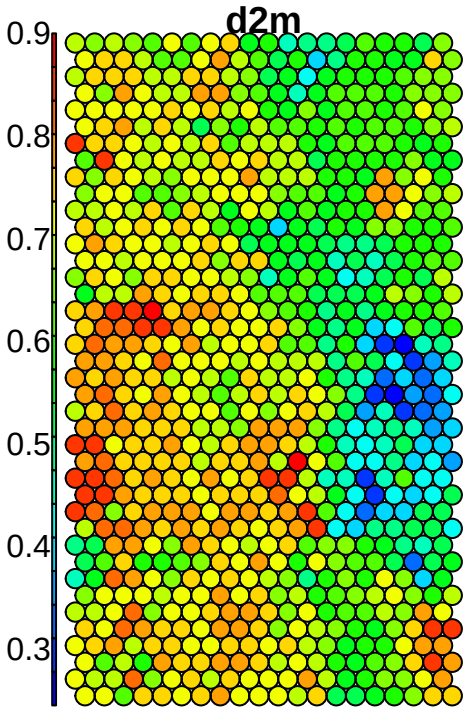

```
for (i in 1:46){
plot(SOM_01, type = "property", property = getCodes(SOM_01)[,i],
main=colnames(getCodes(SOM_01))[i], palette.name=coolBlueHotRed)}
```

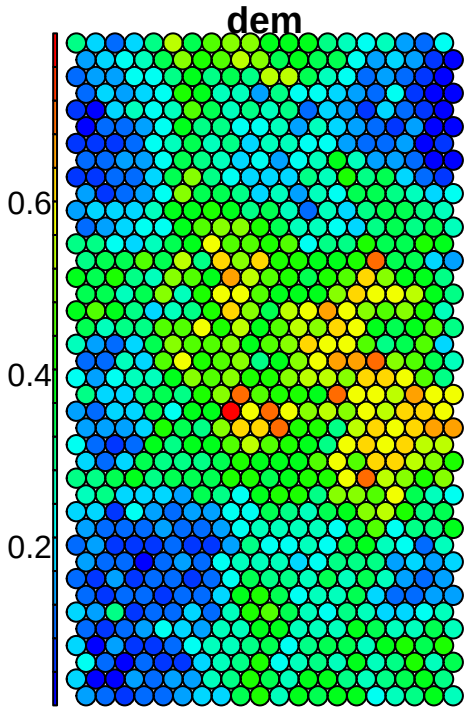

x
0.8
0.6
0.4
0.2

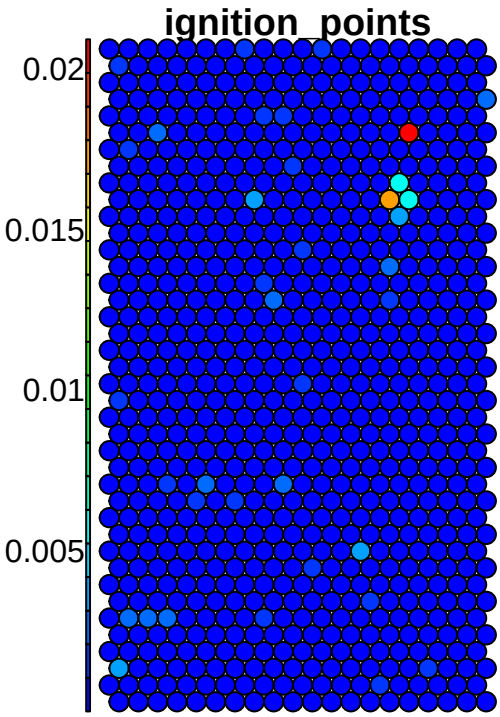

y
0.8
0.6
0.4
0.2

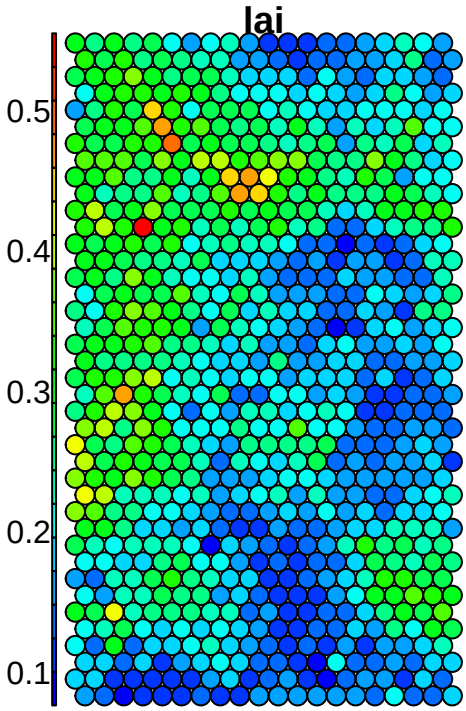

aspect
0.8
0.6
0.4
0.2

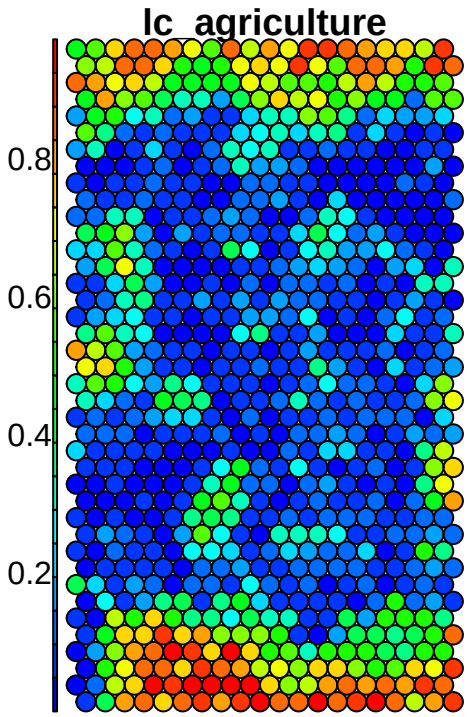

curvature
0.7
0.65
0.6
0.55
0.5

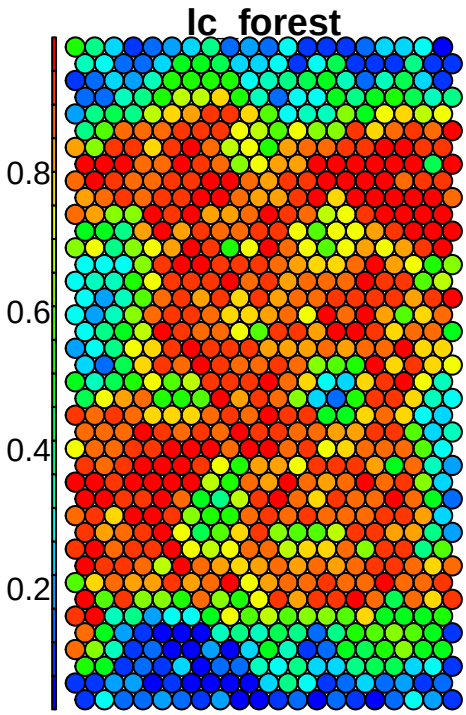

d2m
0.9
0.8
0.7
0.6
0.5
0.4
0.3

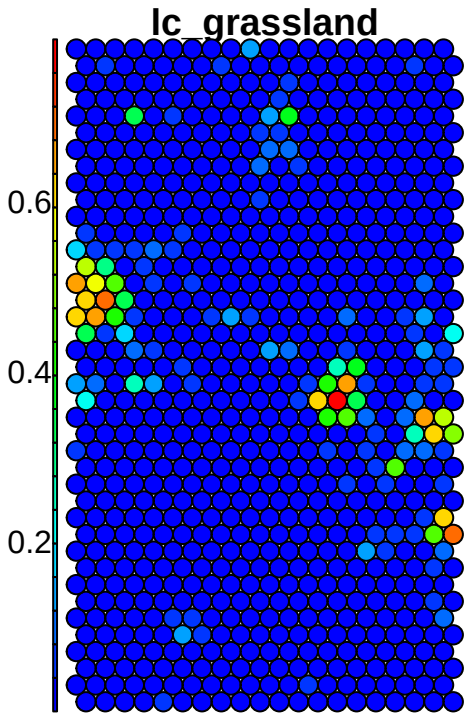

dem
0.6
0.4
0.2

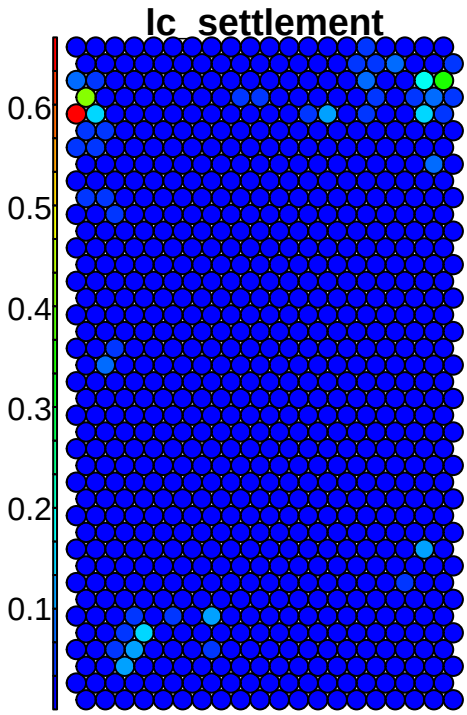

ignition_points
0.02
0.015
0.01
0.005

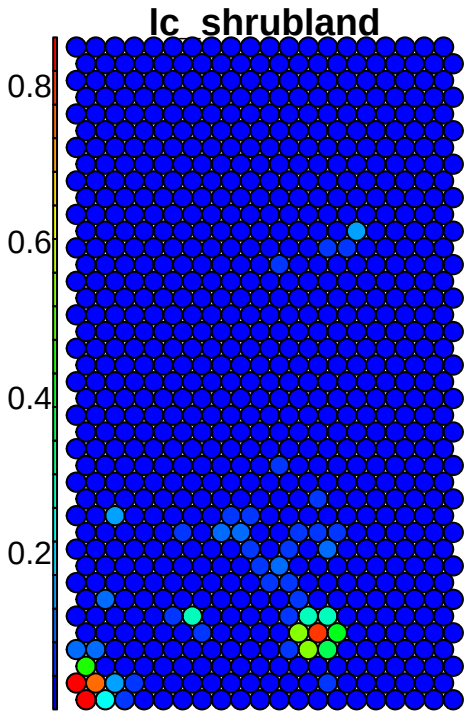

lai
0.5
0.4
0.3
0.2
0.1

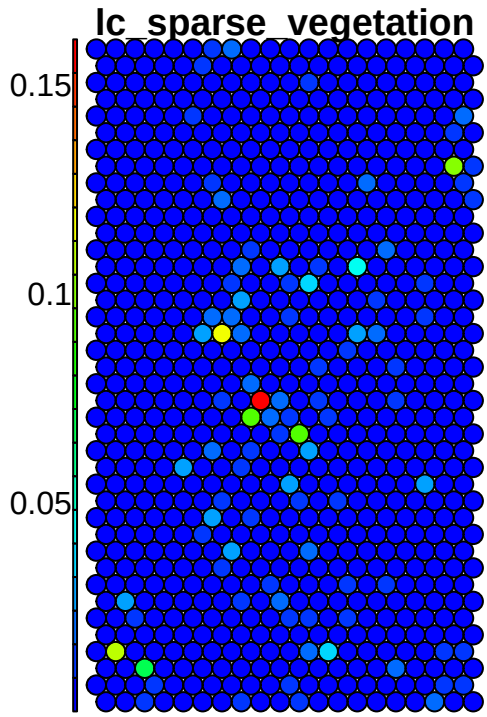

lc_agriculture
0.8
0.6
0.4
0.2

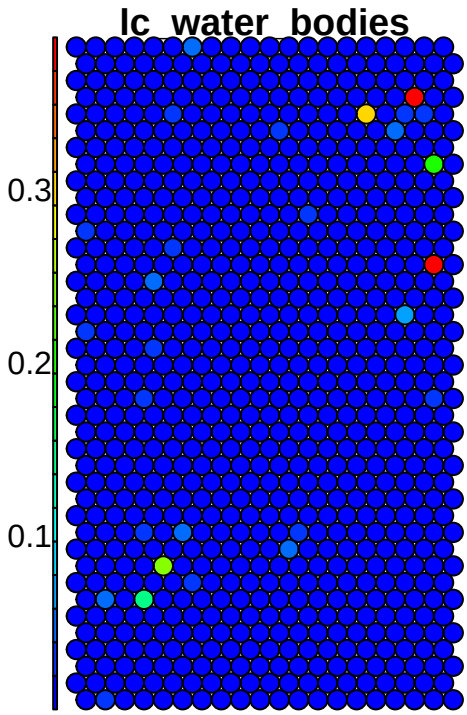

lc_forest
0.8
0.6
0.4
0.2

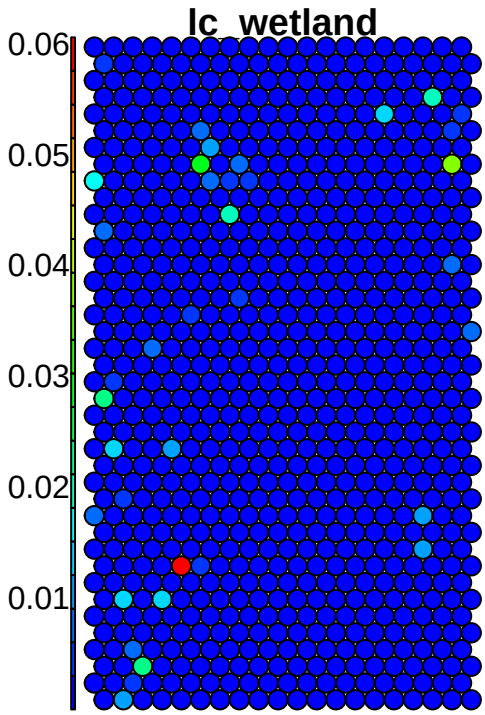

lc_grassland
0.6
0.4
0.2

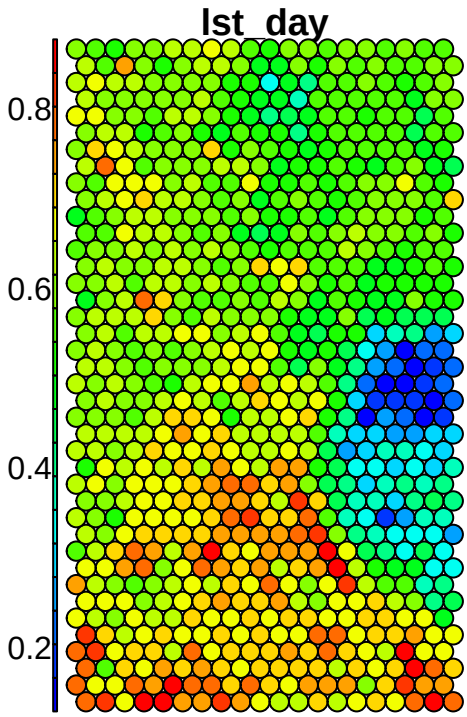

lc_settlement
0.6
0.5
0.4
0.3
0.2
0.1

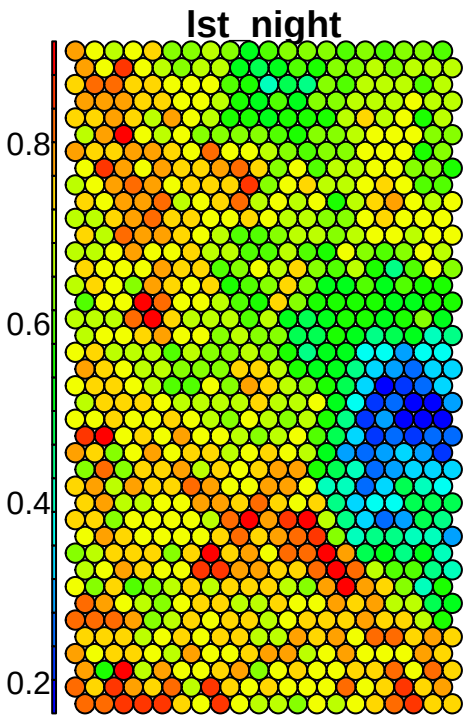

lc_shrubland
0.8
0.6
0.4
0.2

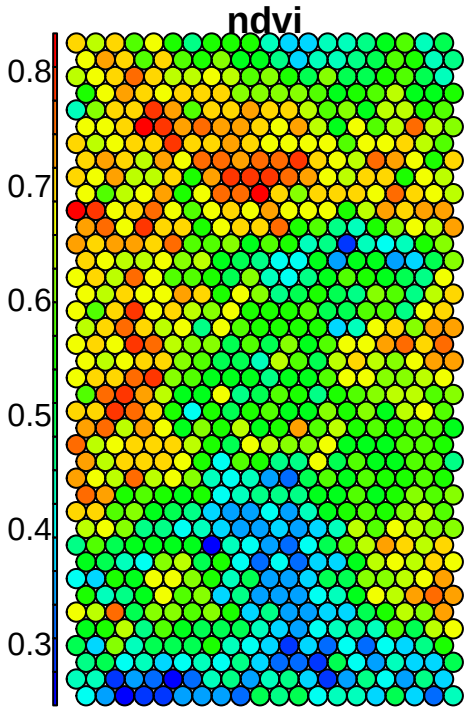

lc_sparse_vegetation
0.15
0.1
0.05

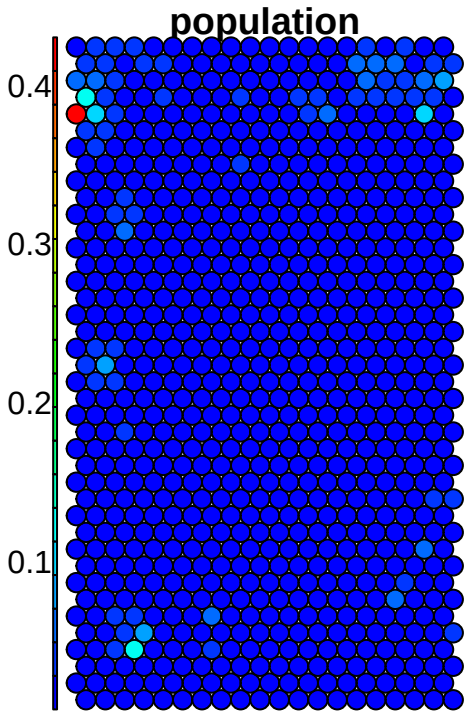

lc_water_bodies
0.3
0.2
0.1

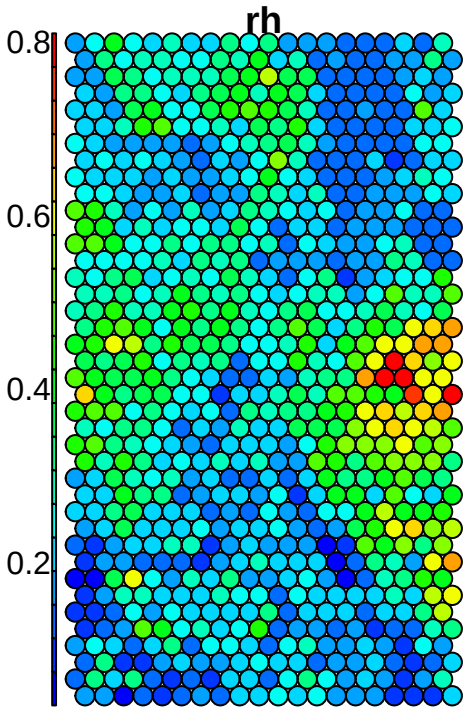

lc_wetland
0.06
0.05
0.04
0.03
0.02
0.01

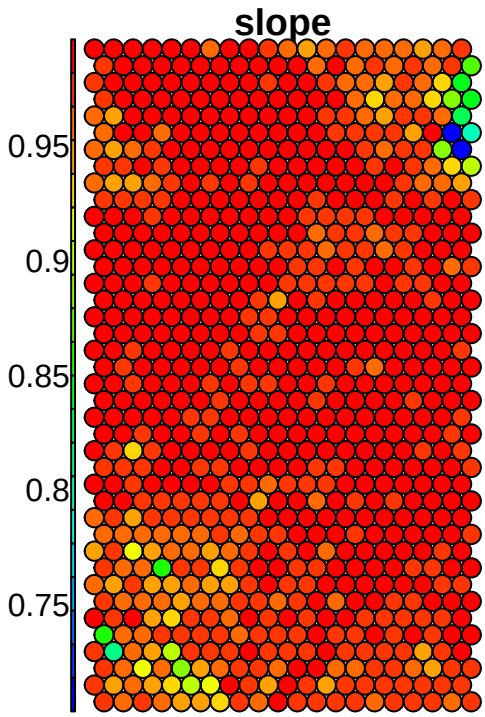

lst_day
0.8
0.6
0.4
0.2

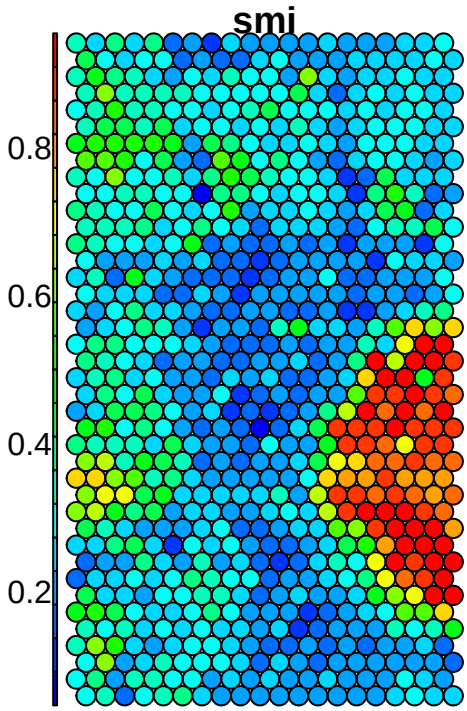

lst_night
0.8
0.6
0.4
0.2

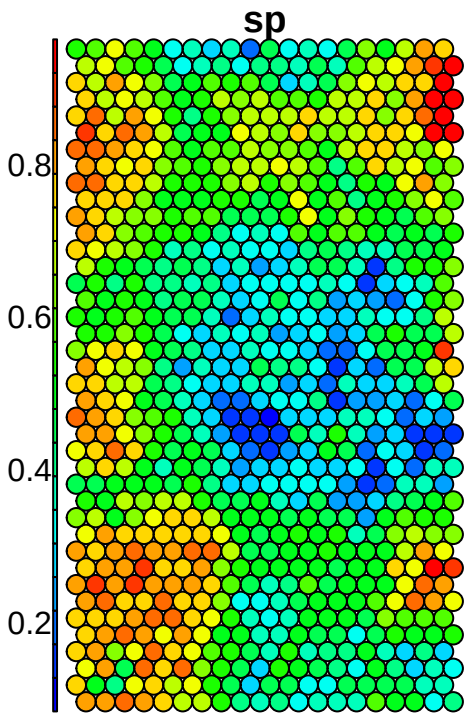

ndvi
0.8
0.7
0.6
0.5
0.4
0.3

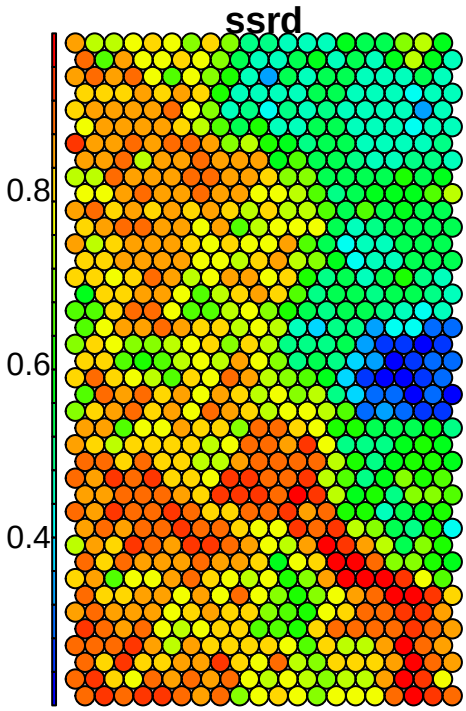

population
0.4
0.3
0.2
0.1

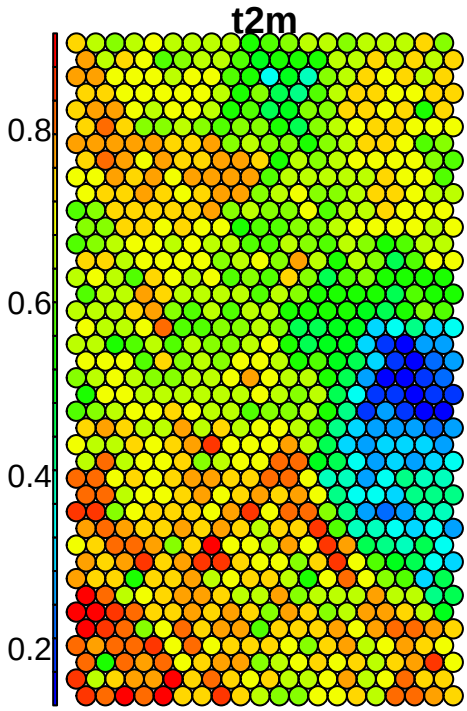

rh
0.8
0.6
0.4
0.2

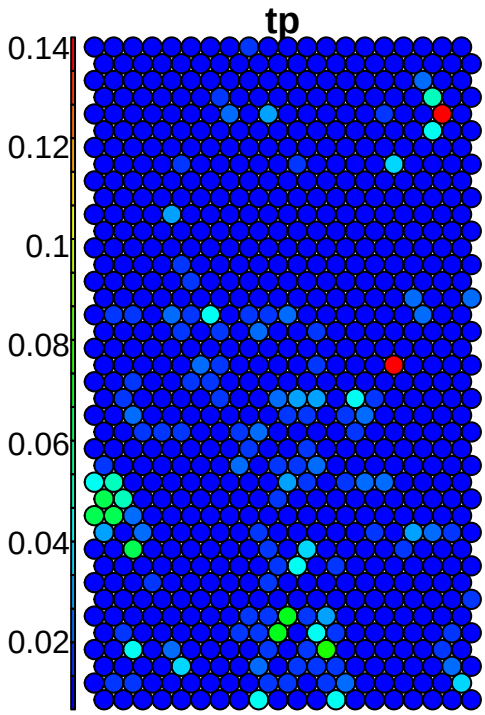

slope
0.95
0.9
0.85
0.8
0.75

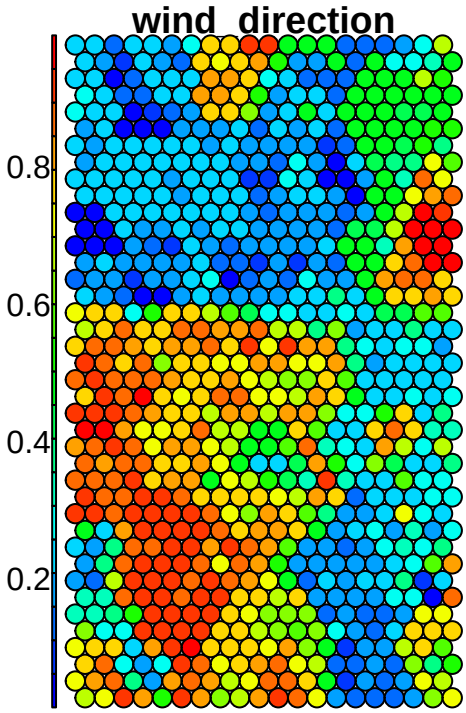

smi
0.8
0.6
0.4
0.2

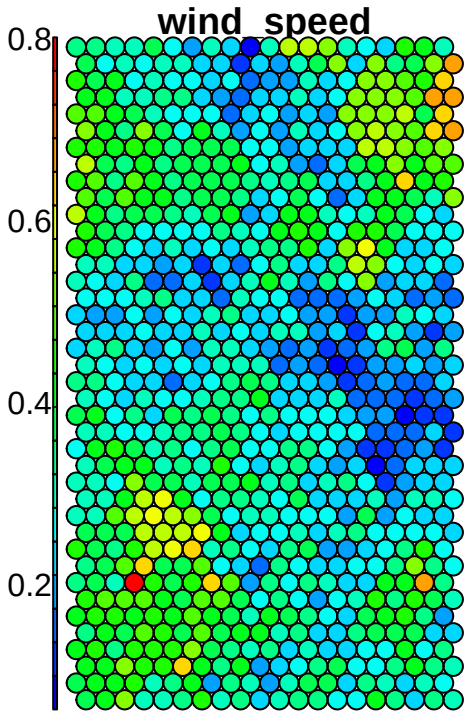

sp
0.8
0.6
0.4
0.2

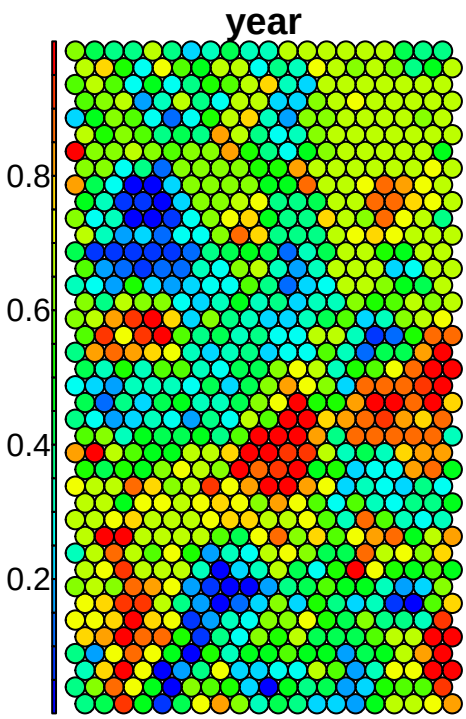

ssrd
0.8
0.6
0.4

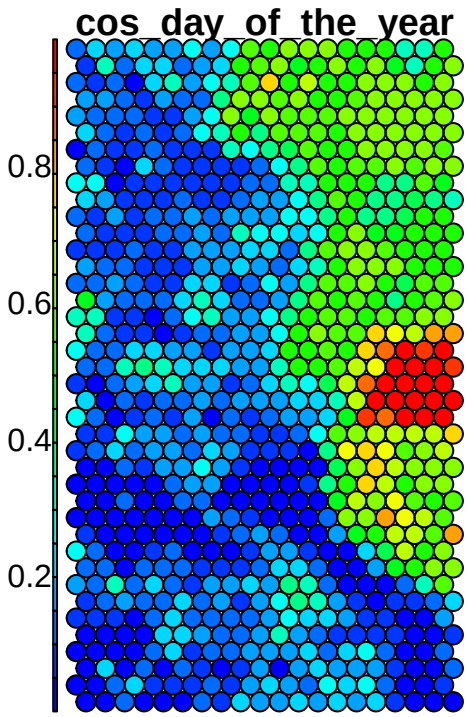

t2m
0.8
0.6
0.4
0.2

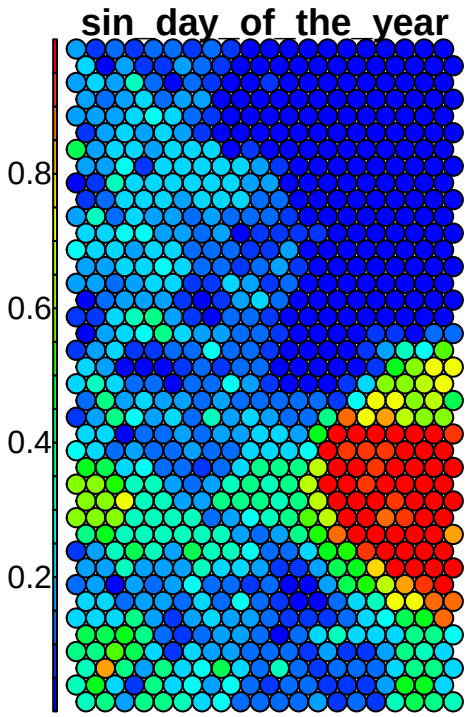

tp
0.14
0.12
0.1
0.08
0.06
0.04
0.02

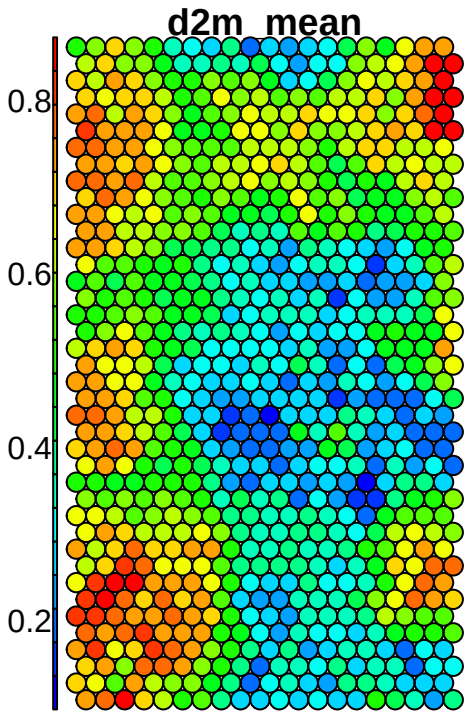

wind_direction
0.8
0.6
0.4
0.2

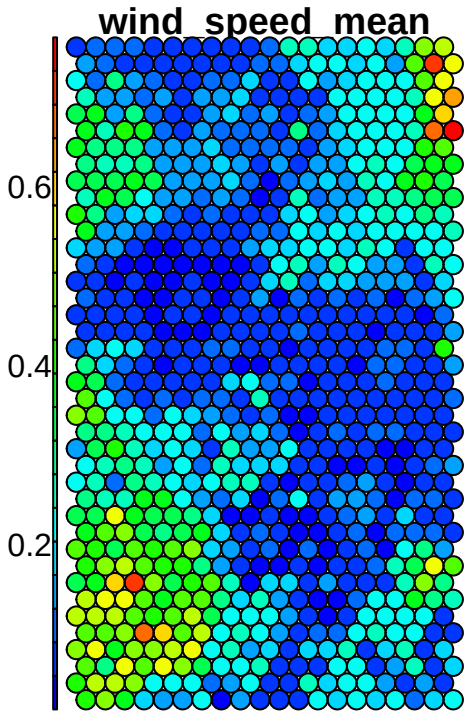

wind_speed
0.8
0.6
0.4
0.2

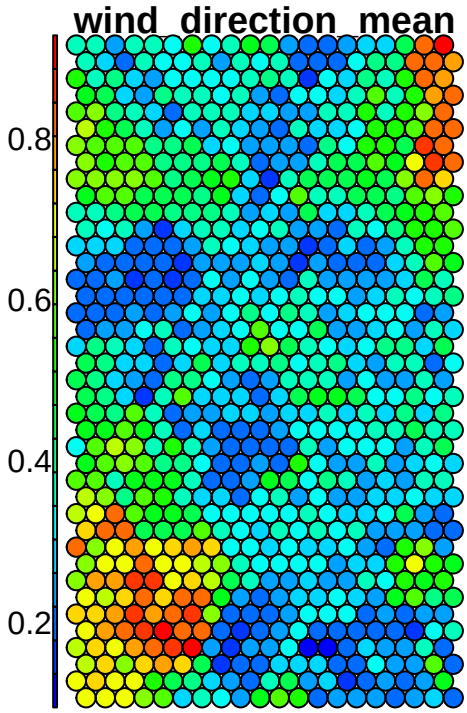

year
0.8
0.6
0.4
0.2

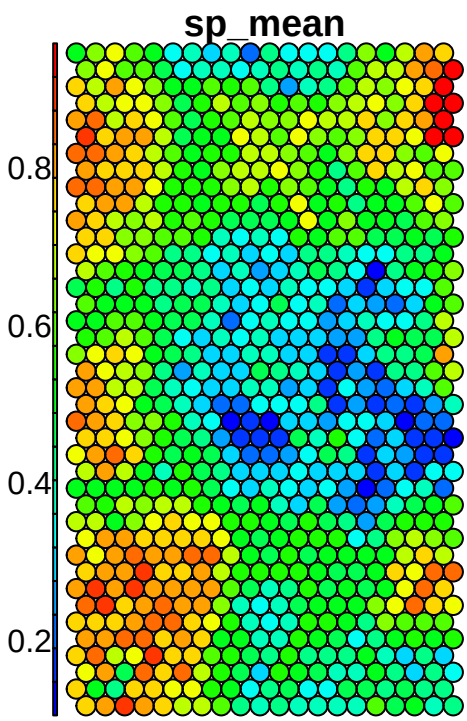

cos_day_of_the_year
0.8
0.6
0.4
0.2

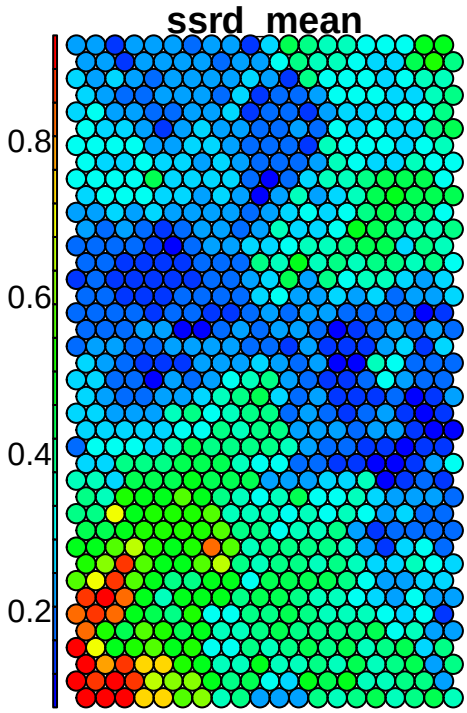

sin_day_of_the_year
0.8
0.6
0.4
0.2

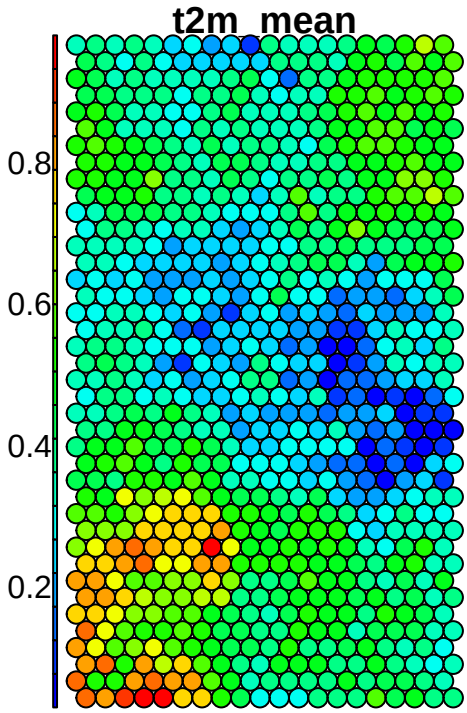

d2m_mean
0.8
0.6
0.4
0.2

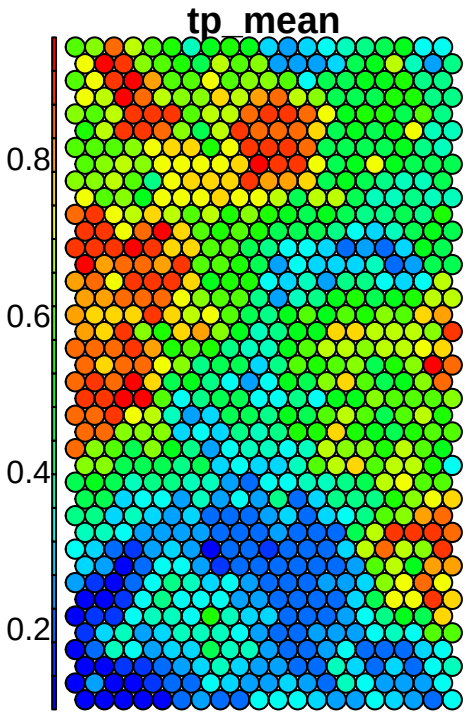

wind_speed_mean
0.6
0.4
0.2

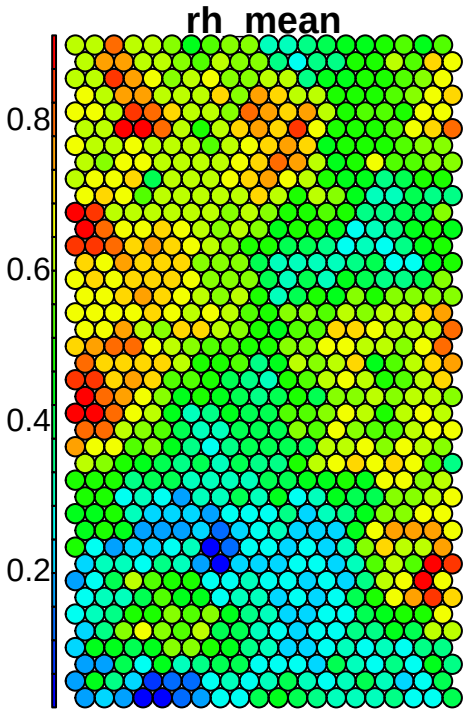

wind_direction_mean
0.8
0.6
0.4
0.2

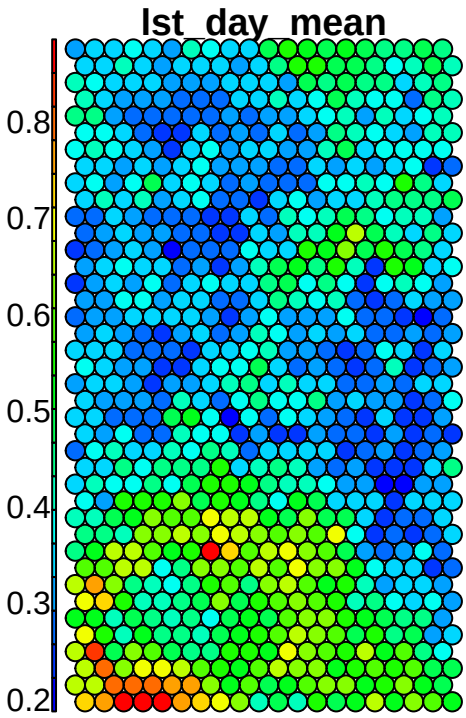

sp_mean
0.8
0.6
0.4
0.2

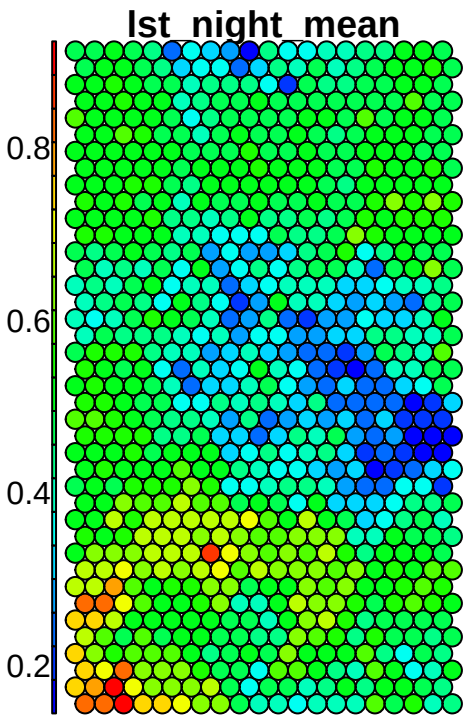

ssrd_mean
0.8
0.6
0.4
0.2

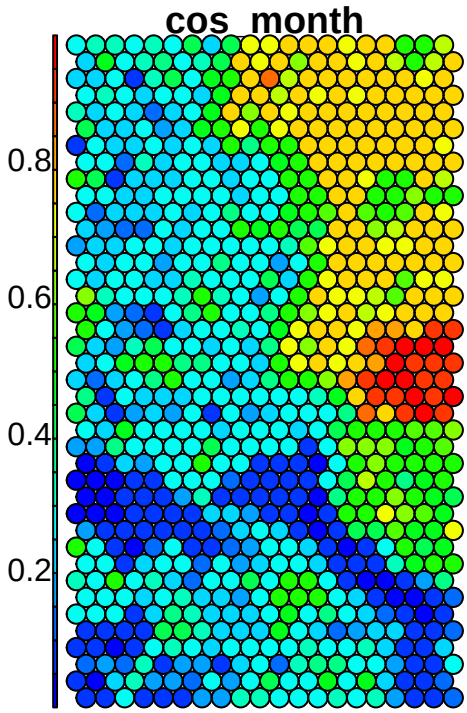

t2m_mean
0.8
0.6
0.4
0.2

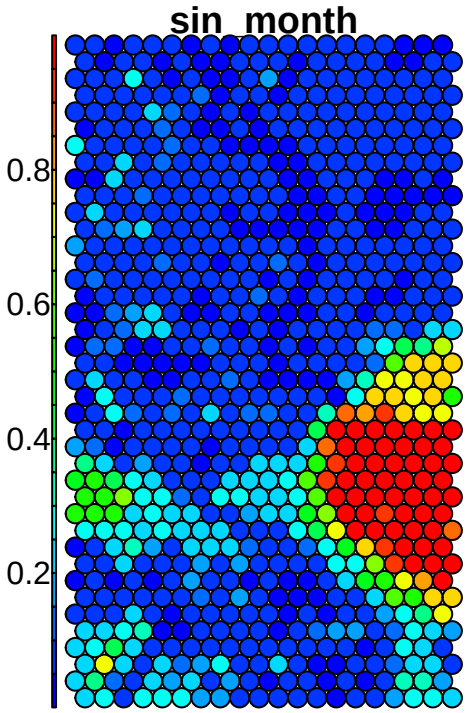

tp_mean
0.8
0.6
0.4
0.2

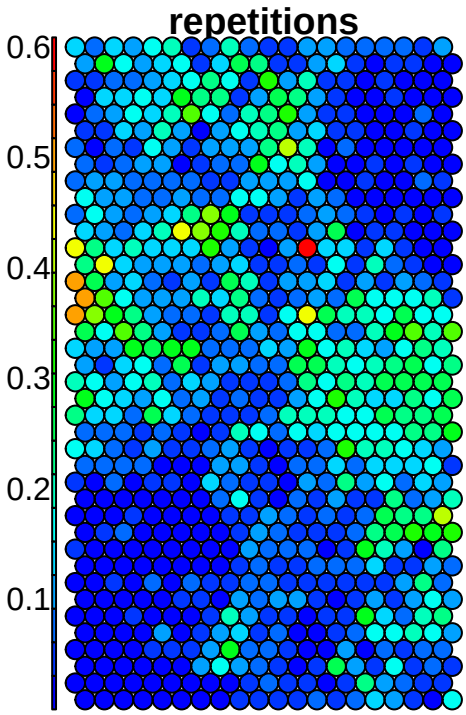

rh_mean
0.8
0.6
0.4
0.2

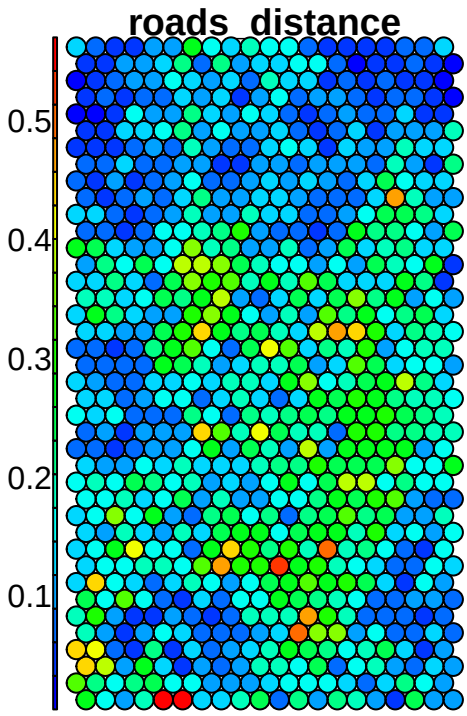

lst_day_mean
0.8
0.7
0.6
0.5
0.4
0.3
0.2


lst_night_mean
0.8
0.6
0.4
0.2


cos_month
0.8
0.6
0.4
0.2


sin_month
0.8
0.6
0.4
0.2


repetitions
0.6
0.5
0.4
0.3
0.2
0.1




```
mydata<-getCodes(SOM_00) # Extract codebook vectors from each unit in the SOM-grid
head (mydata)
```

```
##            x         y    aspect curvature       d2m       dem ignition_points
## V1 0.7483756 0.7781487 0.2483271 0.5633944 0.5451274 0.2651265    1.078782e-04
## V2 0.6669967 0.8193267 0.1491347 0.6231608 0.5970068 0.3912039    1.357306e-04
## V3 0.6871966 0.6876054 0.1876967 0.6133416 0.5483296 0.4817351    2.103173e-04
## V4 0.6886362 0.6659743 0.5534654 0.5888459 0.5488176 0.4194208    2.508795e-04
## V5 0.7140481 0.6597115 0.8804051 0.5621256 0.5141308 0.4293424    6.218073e-06
## V6 0.7071322 0.7974539 0.8402912 0.5761180 0.5752477 0.4026822    5.890037e-05
##          lai lc_agriculture lc_forest lc_grassland lc_settlement lc_shrubland
## V1 0.1426782      0.0865720 0.8741040 1.834527e-10  2.633332e-16 0.0383417353
## V2 0.1720474      0.1656981 0.8274523 1.035314e-10  1.412368e-07 0.0063463647
## V3 0.1473114      0.1071073 0.8612393 5.807975e-10  4.390680e-05 0.0083031555
## V4 0.1180967      0.1874855 0.7652092 1.953901e-14  3.460090e-05 0.0007594938
## V5 0.1378614      0.2005293 0.7936882 2.288132e-12  1.615994e-06 0.0014922470
## V6 0.1452837      0.2512732 0.6961457 4.125716e-11  3.129612e-10 0.0475190927
##    lc_sparse_vegetation lc_water_bodies    lc_wetland   lst_day lst_night
## V1         1.267281e-03    1.269676e-04 8.399116e-323 0.6321201 0.6632268
## V2         6.923147e-05    4.564030e-04 8.399116e-323 0.6174749 0.7354532
## V3         3.452908e-02    3.639967e-06 8.399116e-323 0.6623448 0.6853175
## V4         6.891027e-02    2.947613e-06 7.905050e-323 0.7022797 0.7871031
## V5         6.340470e-03    9.349286e-06 7.905050e-323 0.6942960 0.7364682
## V6         1.878275e-03    3.794370e-03 7.905050e-323 0.6065230 0.6684830
##         ndvi  population        rh     slope       smi        sp      ssrd
## V1 0.4069018 0.007357039 0.2192440 0.9846432 0.2059364 0.5512044 0.7089050
```

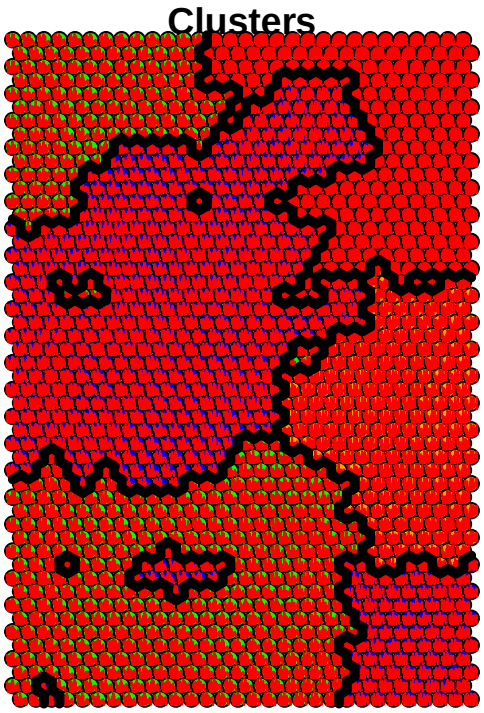
```
## V2 0.4608947 0.007186782 0.1962481 0.9819727 0.2068779 0.4797231 0.8099841
## V3 0.4254825 0.008013263 0.1582480 0.9577771 0.1839607 0.3142909 0.7659811
## V4 0.3811204 0.009200611 0.1481246 0.9828734 0.1795781 0.3806122 0.7557947
## V5 0.4193693 0.008981440 0.1520214 0.9722648 0.1933189 0.3352031 0.7402311
## V6 0.4572390 0.009086145 0.1781778 0.9811699 0.1881451 0.4461365 0.7519707
##          t2m          tp wind_direction wind_speed      year
## V1 0.6104830 0.007858834     0.16482651  0.2980448 0.5038169
## V2 0.6406428 0.007865410     0.17274127  0.2749129 0.3849385
## V3 0.7190333 0.007946676     0.11335035  0.2473229 0.2539000
## V4 0.7552871 0.008080301     0.08788862  0.2632724 0.1909433
## V5 0.7639188 0.007880663     0.07762189  0.2074077 0.1778203
## V6 0.6755849 0.008063792     0.18971336  0.2003617 0.3597994
##    cos_day_of_the_year sin_day_of_the_year  d2m_mean wind_speed_mean
## V1           0.2972482          0.04698629 0.3734759      0.09116571
## V2           0.1667768          0.13551074 0.3508664      0.04204168
## V3           0.2097651          0.09803202 0.2567166      0.25173054
## V4           0.2222619          0.08667098 0.3290575      0.26817441
## V5           0.2292261          0.08124536 0.2812473      0.28035929
## V6           0.2400778          0.08918078 0.3063428      0.10545124
##    wind_direction_mean   sp_mean ssrd_mean  t2m_mean   tp_mean   rh_mean
## V1           0.2670653 0.5349768 0.4087053 0.5169601 0.2338718 0.2385949
## V2           0.3302559 0.5032487 0.3187688 0.4524437 0.2358576 0.2959317
## V3           0.1956087 0.3255785 0.4330728 0.3586804 0.2972981 0.3200776
## V4           0.1746516 0.3980251 0.4553410 0.4128662 0.3183742 0.3209848
## V5           0.1764022 0.3541497 0.4769494 0.3923883 0.3216774 0.3064325
## V6           0.2586745 0.4522331 0.3623438 0.4346491 0.2407292 0.2779950
##    lst_day_mean lst_night_mean cos_month   sin_month repetitions roads_distance
## V1    0.5950088      0.6061614 0.4749258 0.006718606  0.03430172      0.2979056
## V2    0.4382927      0.5223117 0.2508369 0.066820030  0.04769539      0.1365446
## V3    0.4559422      0.3434094 0.2559202 0.065637705  0.07584301      0.1537541
## V4    0.5372818      0.5133403 0.2546877 0.065731235  0.08868530      0.1753072
## V5    0.5085023      0.3705383 0.2594165 0.064464163  0.09084651      0.1920793
## V6    0.4586689      0.4679265 0.3583393 0.038012322  0.02790466      0.1792600
```

```
clsEU00 <- cutree(hclust(dist(getCodes(SOM_00))), 4)
```

```
# Display the SOM-grid with different colors for each cluster:
# Colour palette definition
pretty_palette <- c("green", 'blue', 'orange', 'red',

'yellow', 'grey', 'brown')

plot(SOM_00, bgcol = pretty_palette[clsEU00], main = "Clusters")
add.cluster.boundaries(SOM_00, clsEU00)
```


Clusters