

Yield potential exploratory analysis

This notebook presents and exploratory analysis of the yields maps. The direction of the analysis was guided by MLJE.

The analysis consists of 4 steps:

- 1. [Overview of fields and yield statistics](#)
- 2. [Visualizations of all yield maps for each field](#)
- 3. [Visualizations of yields divided into intervals](#)
- 4. [Visualizations of statistics across years](#)

The files outputted from this notebook has been stored at: T:\2018\160_PlanteInno\3979_PAF_Big Data i planteavlen\01_Arbejdsmappe\chso\181019_yield_potential

Overview of fields and yield statistics

Tables of field aggregated statistics

- Spatial average for each year (mean)
- Spatial standard deviation for each year (std)
- Spatial minimum for each year (min)
- Spatial maximum for each year (max)

Vinterhvede

In [7]:

```
gdf_vh_to_print = np.round(gdf_vh_overview / 1e3, 1)
gdf_vh_to_print.loc[:, idx['', 'Number of fields']] = gdf_vh_overview.loc[:, idx['', 'Number of fields']]
gdf_vh_to_print.style.applymap(color_non_nan_df)
```

Out[7]:

year		2011				2012				2013				2014			
statistic	Number of fields	mean	std	min	max	mean	std	min	max	mean	std	min	max	mean	std	min	max
ID_DDS_field																	
40.0	3	nan	nan	nan	nan	8.6	1.5	1.2	11.7	nan	nan	nan	nan	nan	nan	nan	nan
42.0	3	nan	nan	nan	nan	8.6	1.4	1.2	11.3	nan	nan	nan	nan	8.9	1.4	1.5	13
58.0	3	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	11.1	2.3	1.3	17.1
70.0	3	6.8	1.2	1.6	9.7	nan	nan	nan	nan	5.6	0.9	1.7	7.7	nan	nan	nan	nan
176.0	4	nan	nan	nan	nan	8.2	1.7	1.2	11.9	nan	nan	nan	nan	6.7	1.5	1.1	10.1
211.0	3	7.6	1.5	1.6	10.6	nan	nan	nan	nan	nan	nan	nan	nan	7.7	1.4	1.4	10.1
284.0	3	nan	nan	nan	nan	6.9	1.3	1.2	10.6	nan	nan	nan	nan	nan	nan	nan	nan

Majshelsaed

```
In [8]:

gdf_mh_to_print = np.round(gdf_mh_overview / 1e3, 1)
gdf_mh_to_print.loc[:, idx['', 'Number of fields']] = gdf_mh_overview.loc[:, idx['', 'Number of fields']]
gdf_mh_to_print.style.applymap(color_non_nan_df)
```

Out[8]:

year		2011				2012				2013				2014			
statistic	Number of fields	mean	std	min	max	mean	std	min	max	mean	std	min	max	mean	std	min	ma
ID_DDS_field																	
3.0	2	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan
18.0	2	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan
46.0	2	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan
49.0	2	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan
56.0	2	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan
64.0	2	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan
103.0	2	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan
124.0	2	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan
175.0	2	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan
192.0	2	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan
230.0	2	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan

Average field size

The field polygon changes between years. Thus, the field area also changes between years. The average area across all years for each field (in ha) is reported below.

Vinterhvede

```
In [9]:

vh_area_ser = gdf_fields.set_index('ID_DDS_field').loc[gdf_vh_overview.index, 'field_area'].groupby('ID_DDS_field').mean()
vh_area_ser
```

Out[9]:

```
ID_DDS_field
40.0      10.235077
42.0      10.873746
58.0       5.420770
70.0       6.003561
176.0      7.197205
211.0      9.536339
284.0     10.852312
Name: field_area, dtype: float64
```

Majhelsaed

In [10]:

```
mh_area_ser = gdf_fields.set_index('ID_DDS_field').loc[gdf_mh_overview.index, 'field_area'].groupby('ID_DDS_field').mean()
mh_area_ser
```

Out[10]:

```
ID_DDS_field
3.0      2.262486
18.0     7.282606
46.0     9.084553
49.0     4.294817
56.0     8.687333
64.0     5.905645
103.0    3.701472
124.0    10.581925
175.0    7.412463
192.0    13.501679
230.0    8.334780
Name: field_area, dtype: float64
```

Exploratory analysis

Show all yield maps for each field

All available yield maps for each field are displayed next to each other using a common colormap for the (continuous) yield measurements. The ID_DDS identifier specifies the field number.

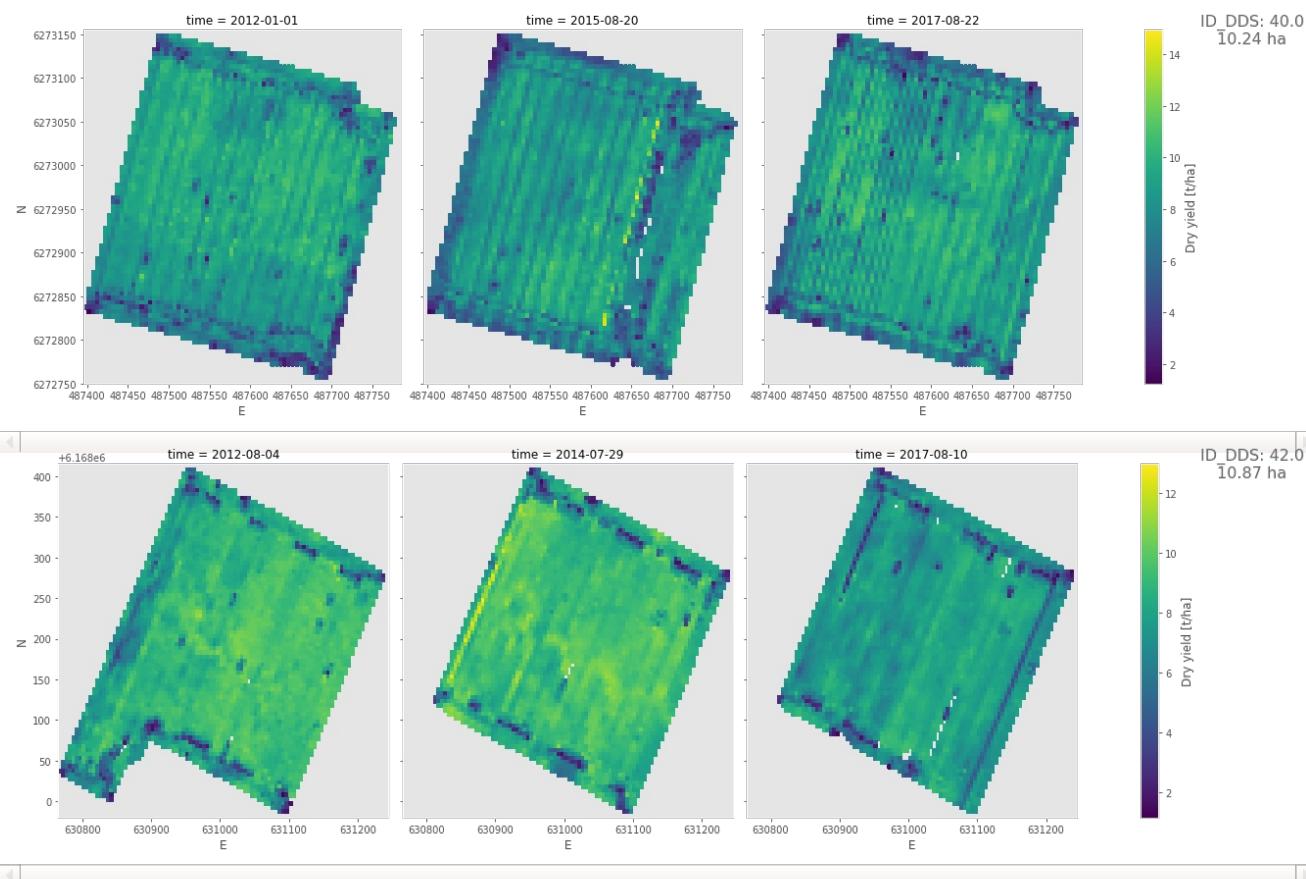
Some of the yield maps contain "holes" which are the result of insufficient data for the weighted distance interpolation method used to grid the yield maps.

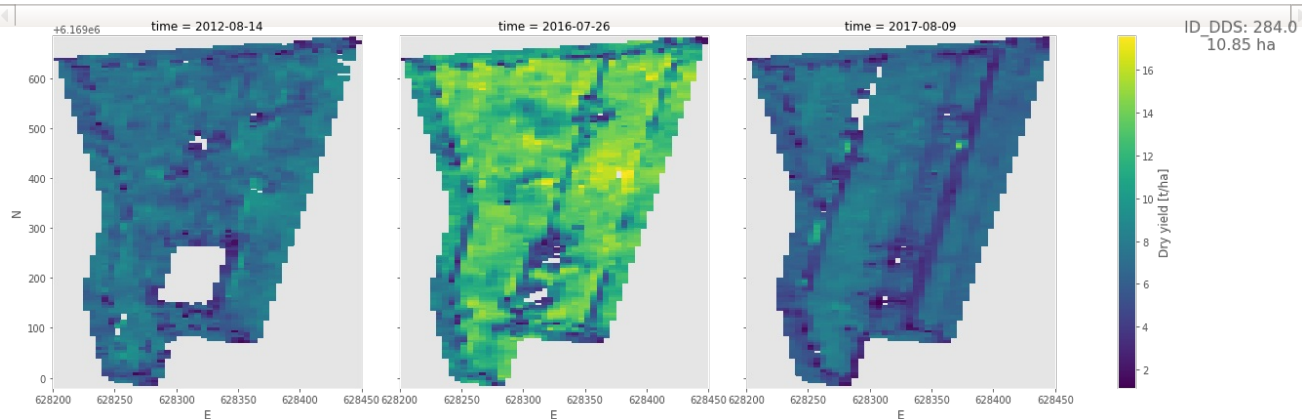
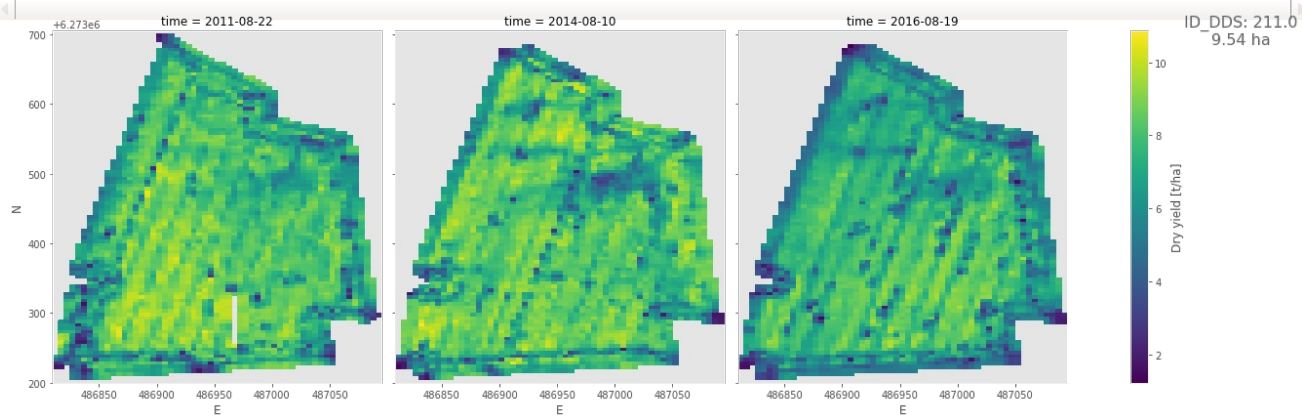
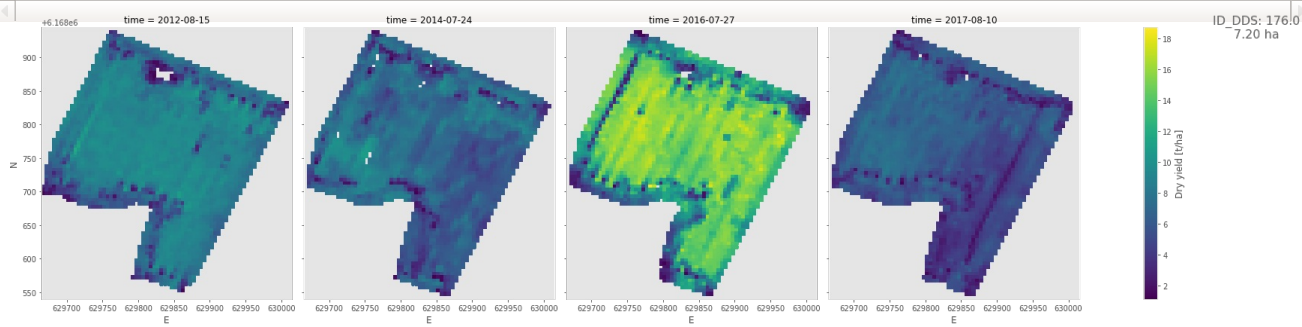
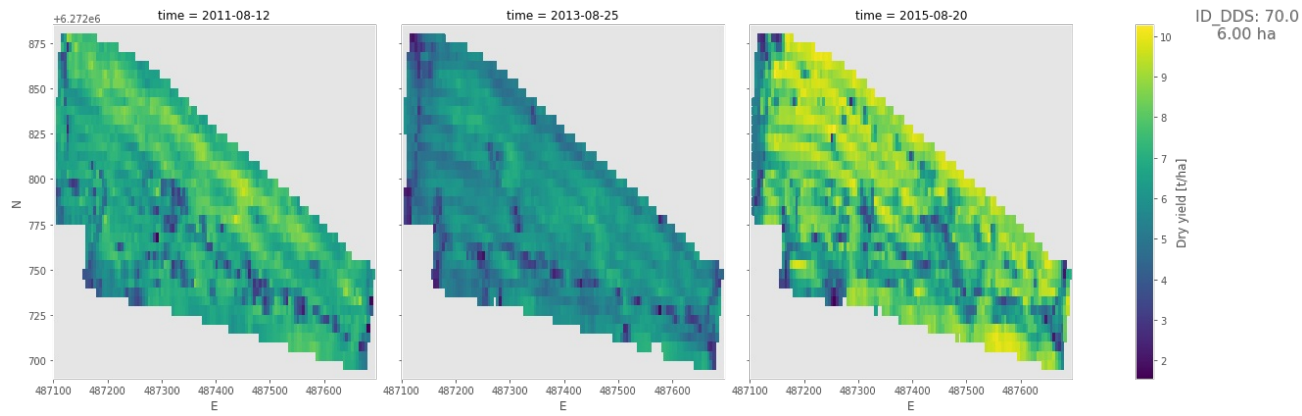
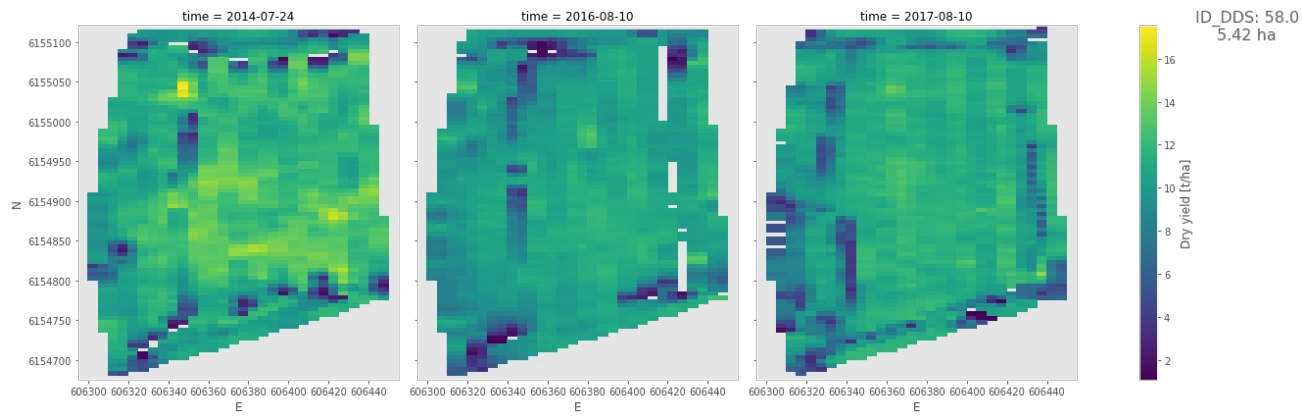
Vinterhvede

In [12]:

```
plot_yield_maps(gdf_vh_overview, vh_area_ser, output_path='./vinterhvede')

vh_yield_map_figures = sorted(glob.glob('./vinterhvede/*_yield_maps.pdf'), key=lambda name: int(name.split('_')[-3]))
subprocess.run(['pdfunite'] + vh_yield_map_figures + ['./vinterhvede/all_yield_maps_vinterhvede.pdf'])
```





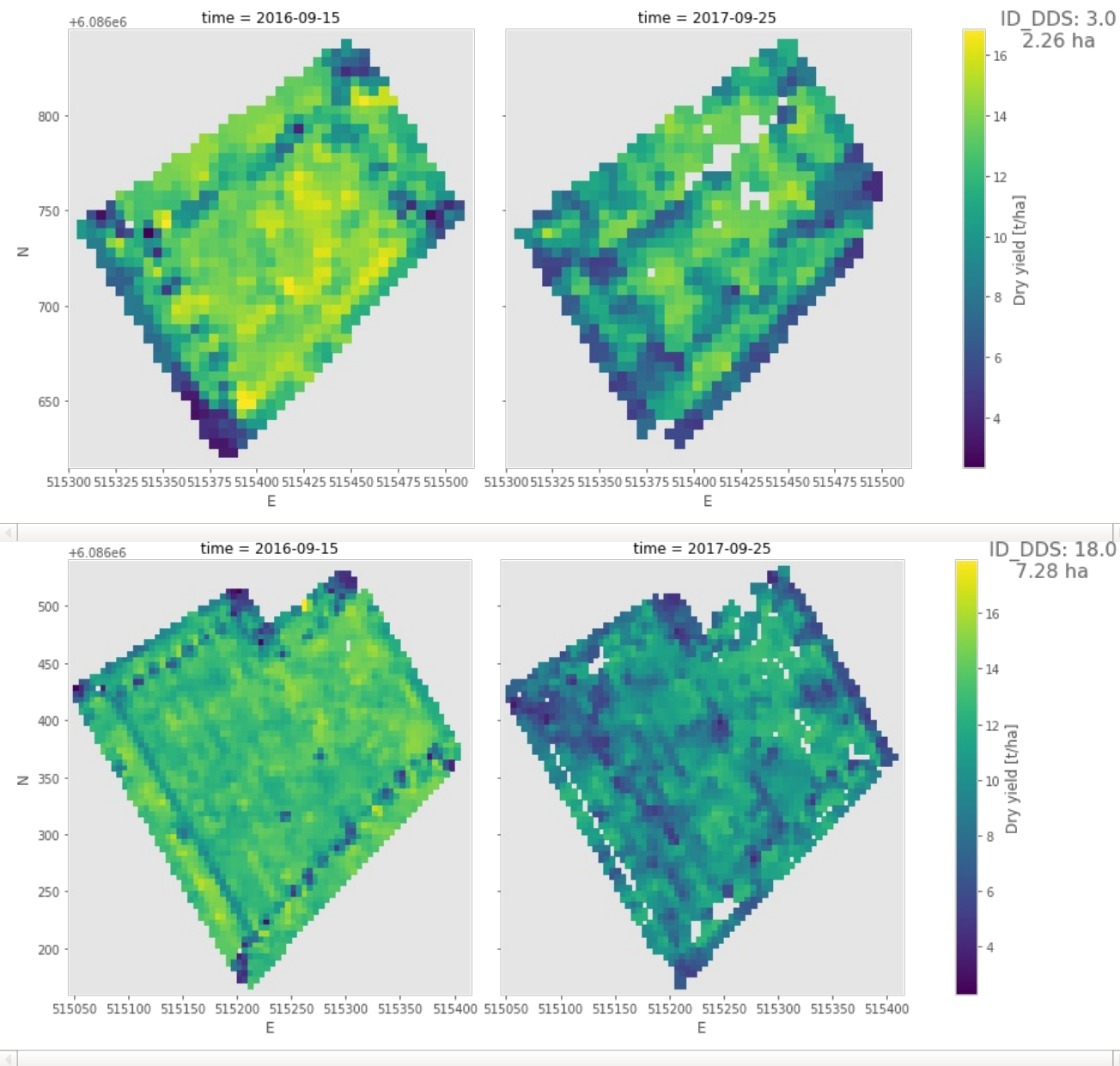
Out[12]:

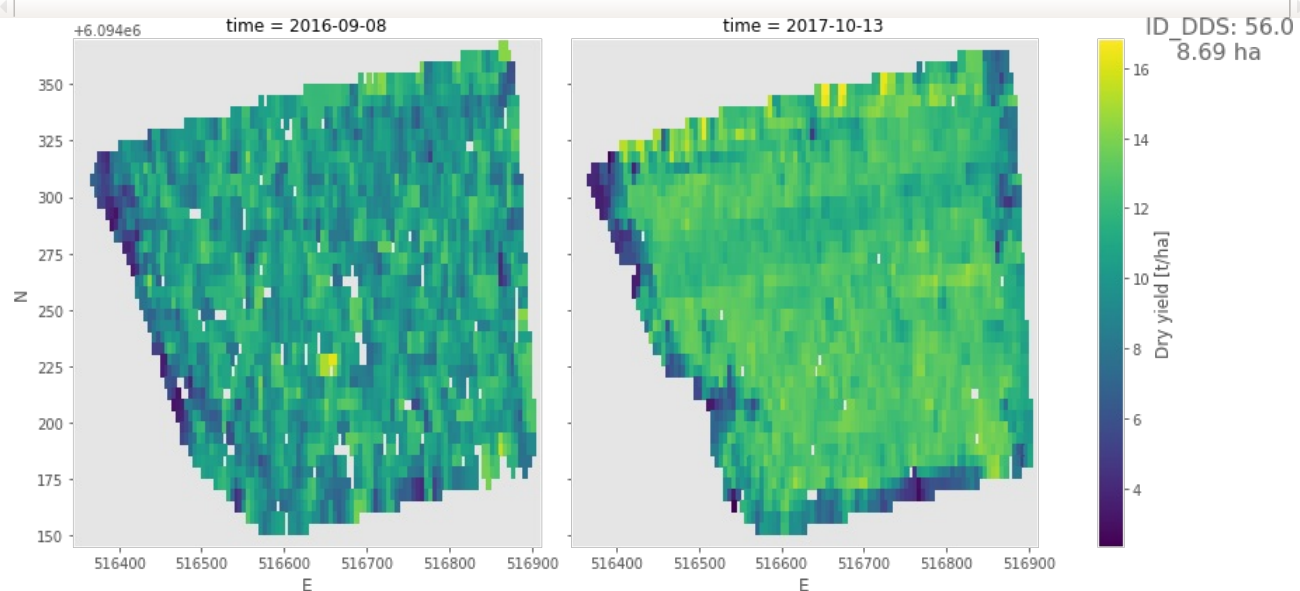
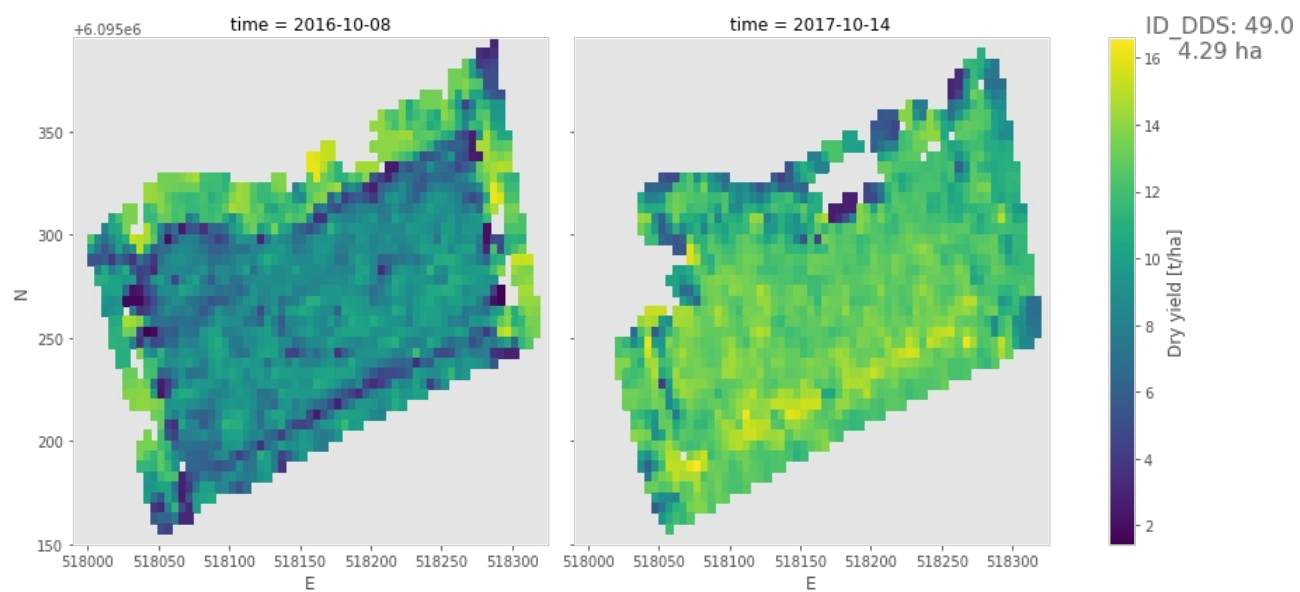
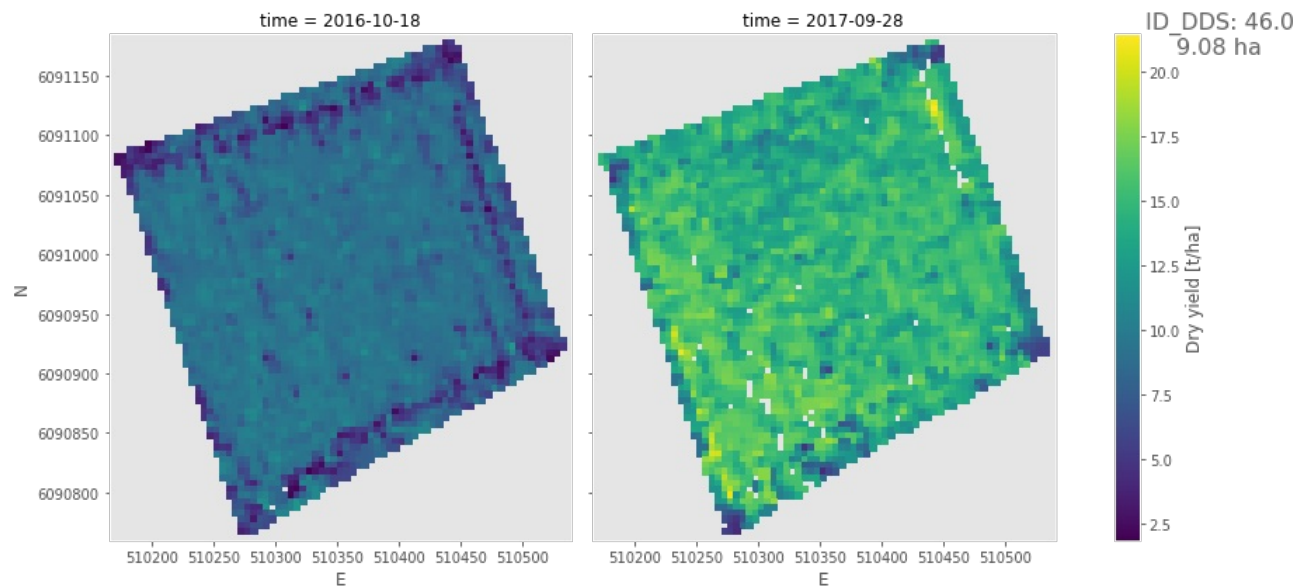
```
CompletedProcess(args=['pdfunite', './vinterhvede/DDS_field_40_yield_maps.pdf', './vinterhvede/DDS_f  
ield_42_yield_maps.pdf', './vinterhvede/DDS_field_58_yield_maps.pdf', './vinterhvede/DDS_field_70_yi  
eld_maps.pdf', './vinterhvede/DDS_field_176_yield_maps.pdf', './vinterhvede/DDS_field_211_yield_maps  
.pdf', './vinterhvede/DDS_field_284_yield_maps.pdf', './vinterhvede/all_yield_maps_vinterhvede.pdf']  
, returncode=0)
```

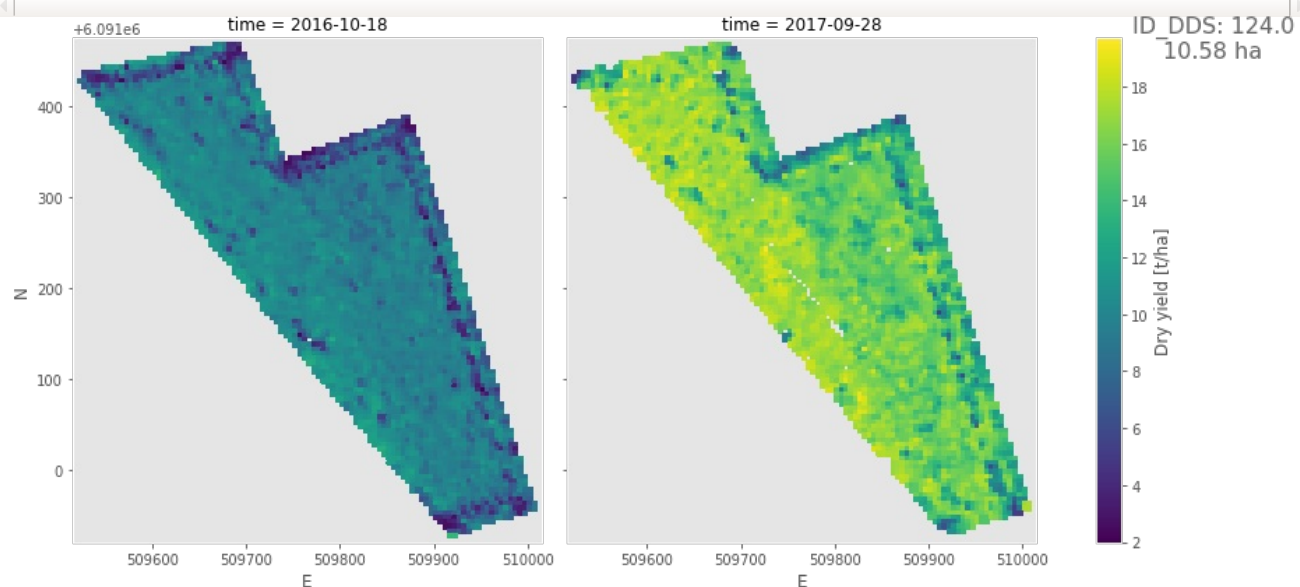
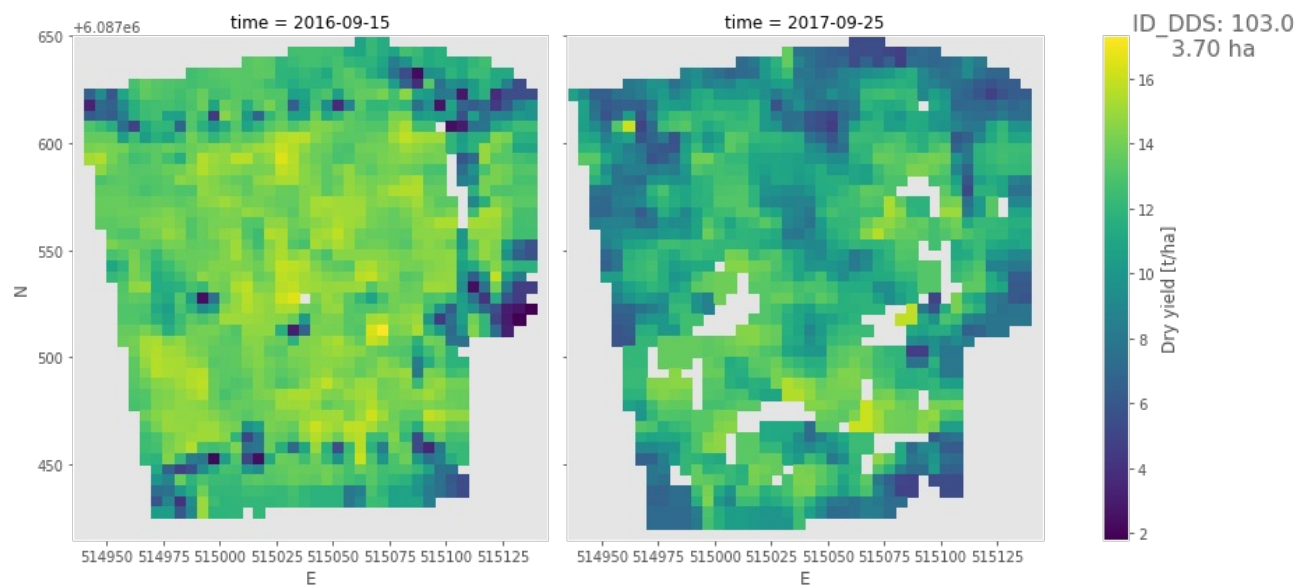
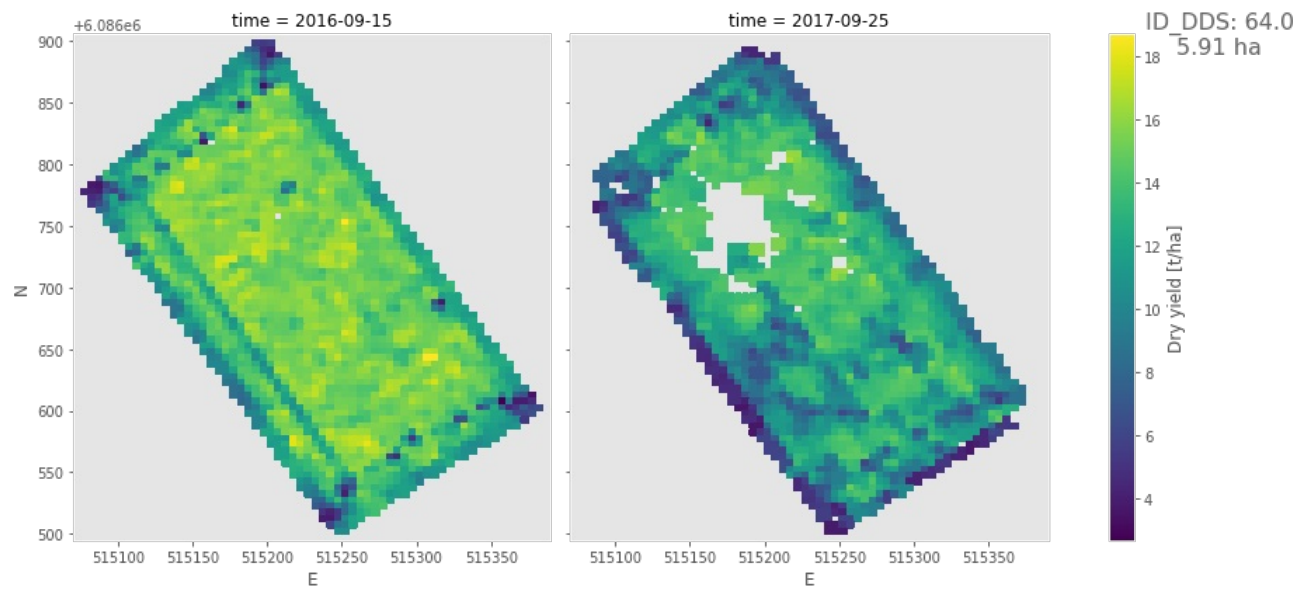
Majshelsaed

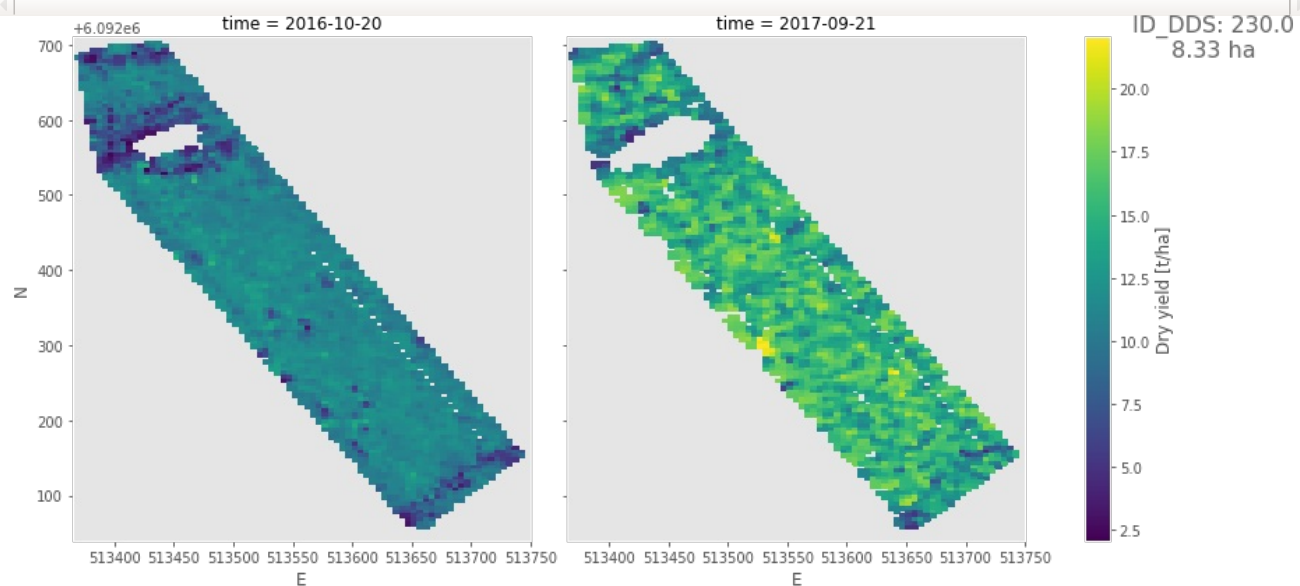
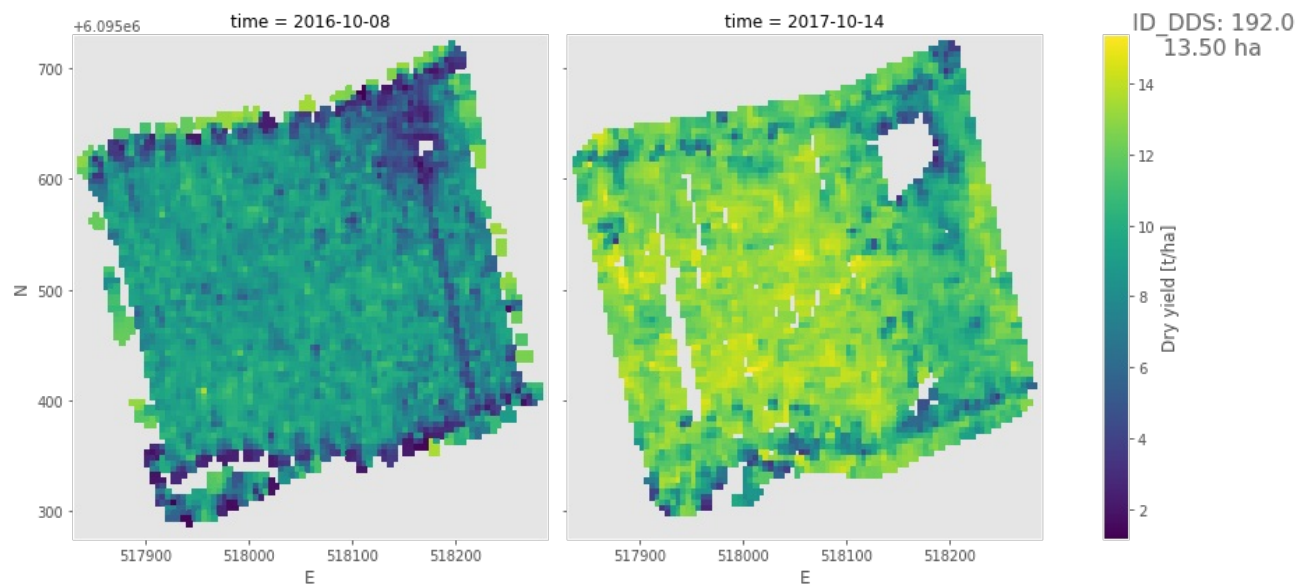
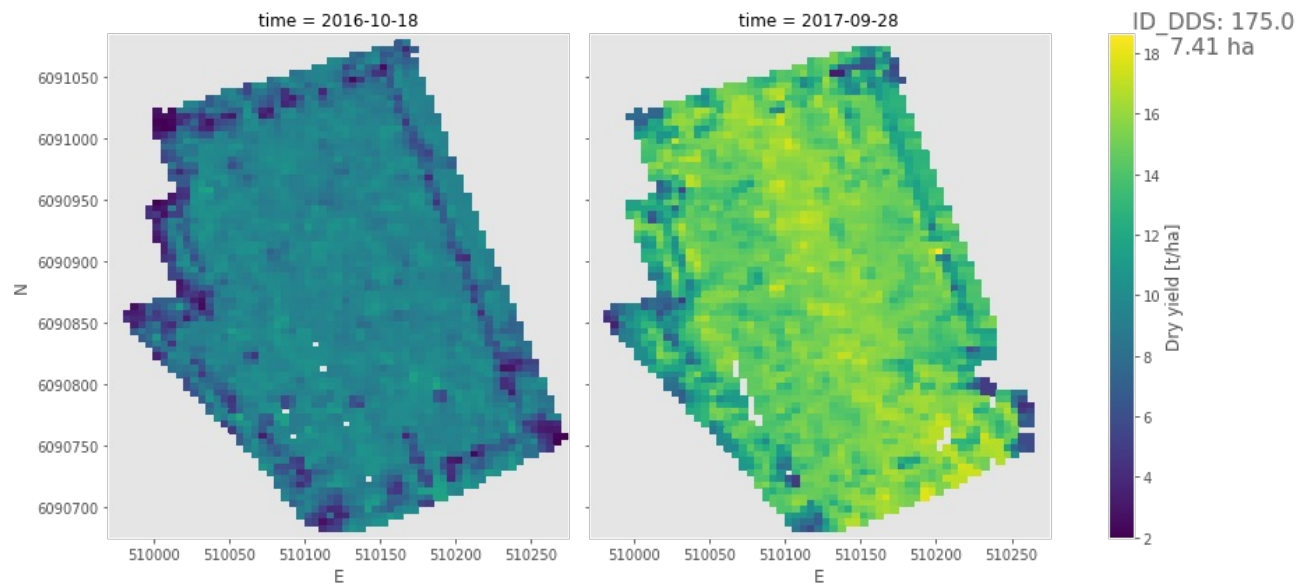
In [13]:

```
plot_yield_maps(gdf_mh_overview, mh_area_ser, output_path='./majshelsaed')  
  
mh_yield_map_figures = sorted(glob.glob('./majshelsaed/*_yield_maps.pdf'), key=lambda name: int(name.split('_')[-  
3]))  
subprocess.run(['pdfunite'] + mh_yield_map_figures + ['./majshelsaed/all_yield_maps_majshelsaed.pdf'])
```









Out[13]:

```
CompletedProcess(args=['pdfunite', './majshelsaed/DDS_field_3_yield_maps.pdf', './majshelsaed/DDS_fi
eld_18_yield_maps.pdf', './majshelsaed/DDS_field_46_yield_maps.pdf', './majshelsaed/DDS_field_49_yie
ld_maps.pdf', './majshelsaed/DDS_field_56_yield_maps.pdf', './majshelsaed/DDS_field_64_yield_maps.pd
f', './majshelsaed/DDS_field_103_yield_maps.pdf', './majshelsaed/DDS_field_124_yield_maps.pdf', './m
ajshelsaed/DDS_field_175_yield_maps.pdf', './majshelsaed/DDS_field_192_yield_maps.pdf', './majshelsa
ed/DDS_field_230_yield_maps.pdf', './majshelsaed/all_yield_maps_majshelsaed.pdf'], returncode=0)
```

Divide yield into intervals

All available yield maps are binned into 4-5 yield intervals. The number of intervals for the vinterhvede fields has been settled by MLJE:

- 40: 4 bins
- 42: 4 bins
- 58: 5 bins
- 70: 4 bins
- 176: 5 bins
- 211: 4 bins
- 284: 5 bins

All majshelsæd fields are binned using 4 bins.

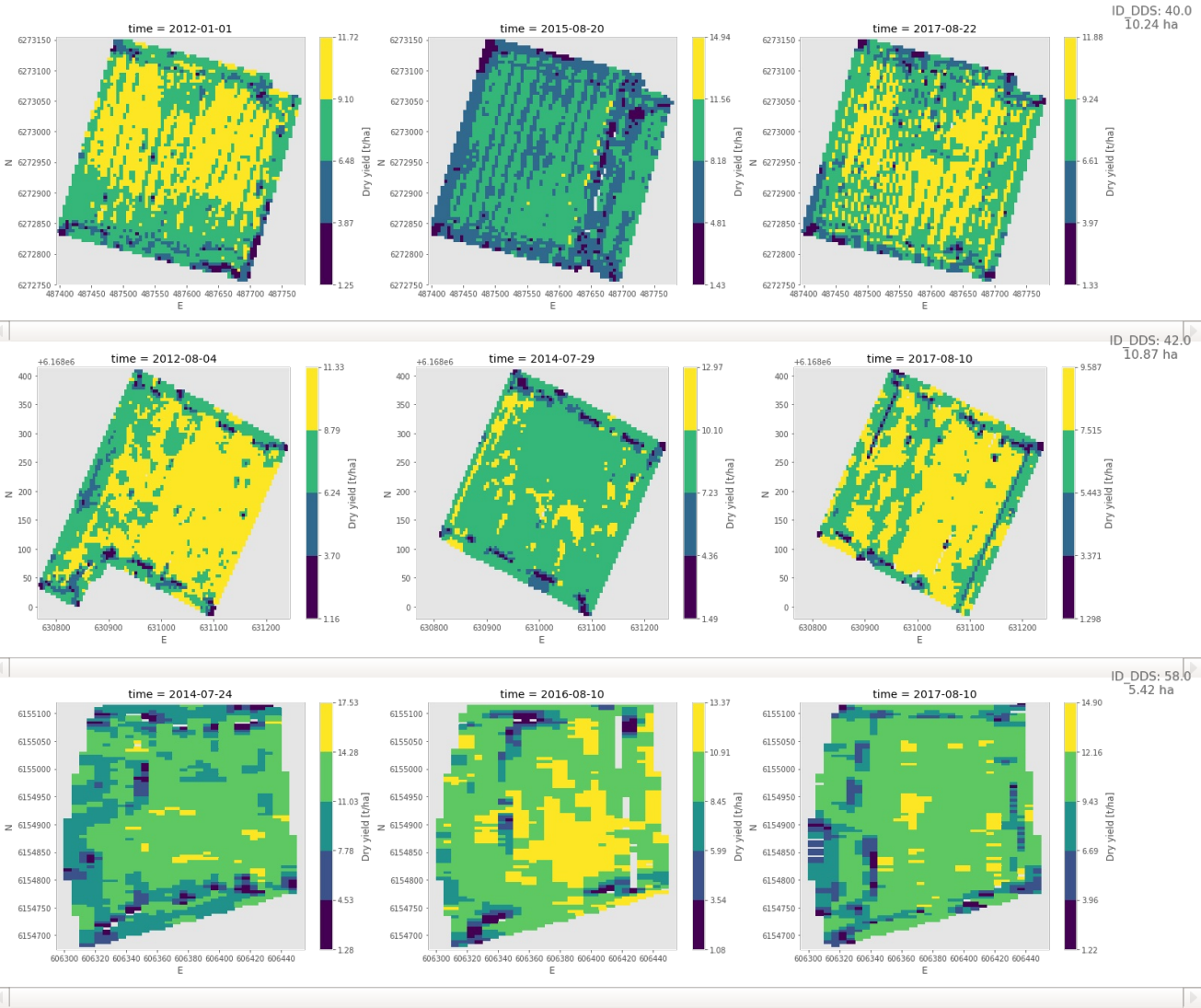
All the binned yield maps for each field are shown next to each other. Note that each yield maps has its own unique interval borders.

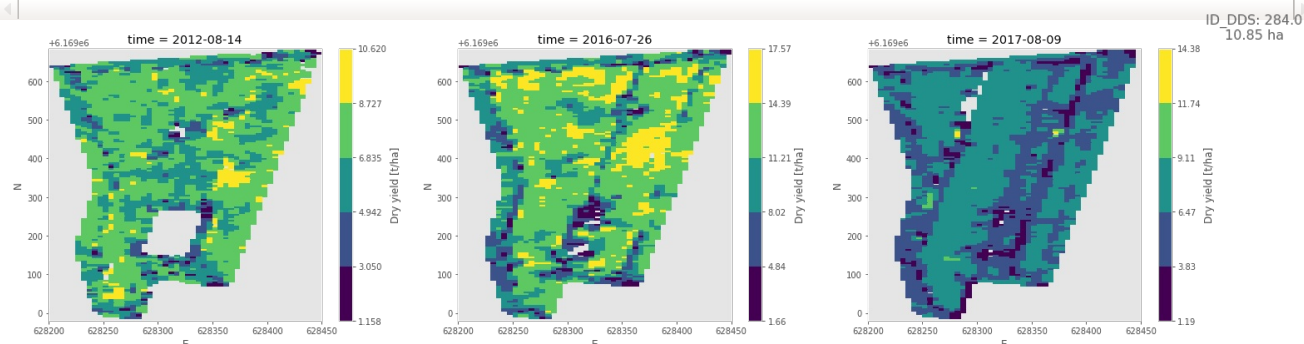
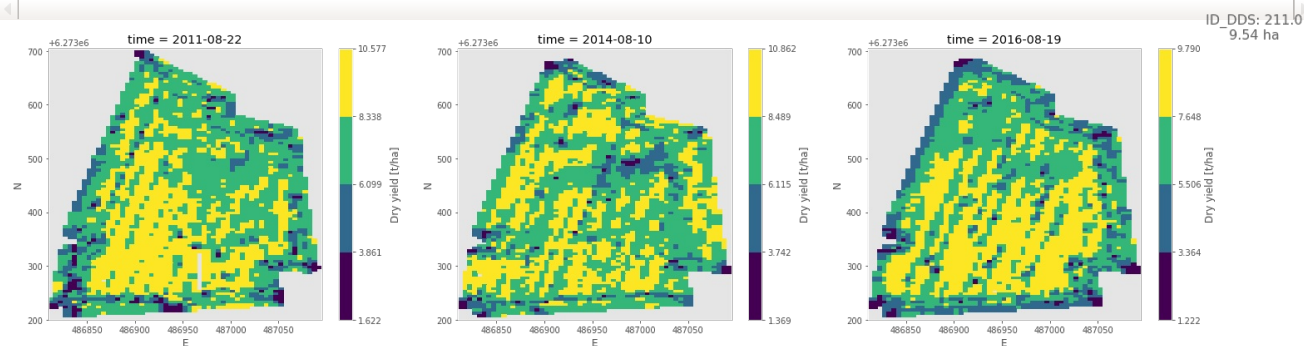
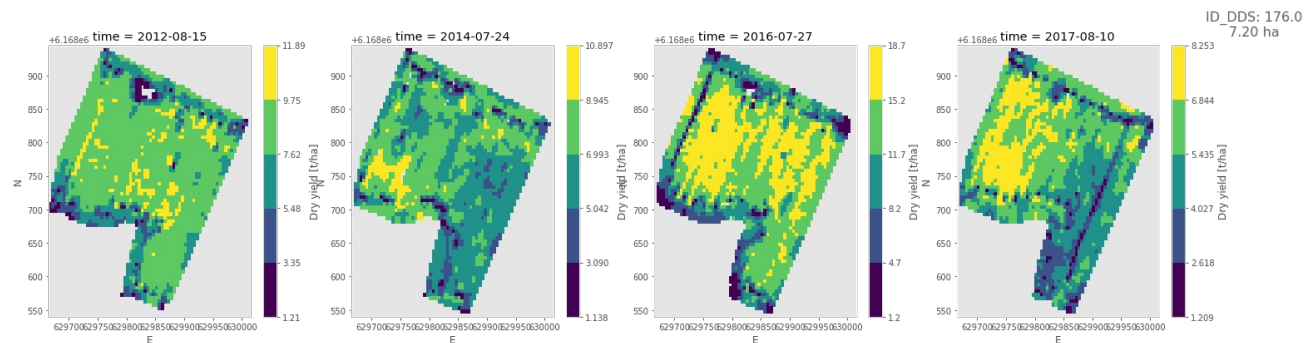
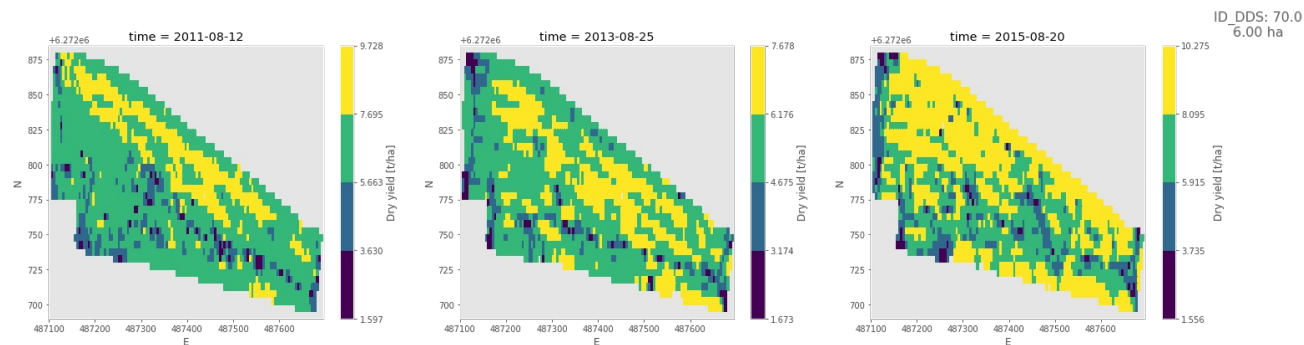
Vinterhvede

In [14]:

```
vh_harvest_dry_yield_intervals = {field: intervals for field, intervals in zip(gdf_vh_overview.index, [4, 4, 5, 4, 5, 4, 5])} # As requested by MLJE
plot_yield_intervals(gdf_vh_overview, vh_area_ser, vh_harvest_dry_yield_intervals, output_path='./vinterhvede')

vh_yield_intervals_figures = sorted(glob.glob('./vinterhvede/*_yield_intervals.pdf'), key=lambda name: int(name.split('_')[-3]))
subprocess.run(['pdfunite'] + vh_yield_intervals_figures + ['./vinterhvede/all_yield_intervals_vinterhvede.pdf'])
```





Out[14]:

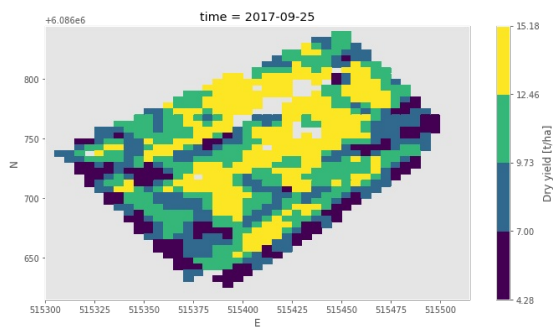
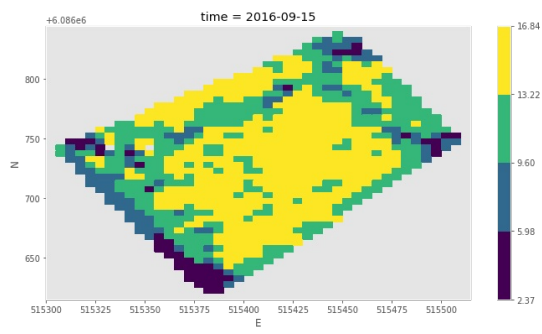
```
CompletedProcess(args=['pdfunite', './vinterhvede/DDS_field_40_yield_intervals.pdf', './vinterhvede/DDS_field_42_yield_intervals.pdf', './vinterhvede/DDS_field_58_yield_intervals.pdf', './vinterhvede/DDS_field_70_yield_intervals.pdf', './vinterhvede/DDS_field_176_yield_intervals.pdf', './vinterhvede/DDS_field_211_yield_intervals.pdf', './vinterhvede/DDS_field_284_yield_intervals.pdf', './vinterhvede/all_yield_intervals_vinterhvede.pdf'], returncode=0)
```

Majshelsaед

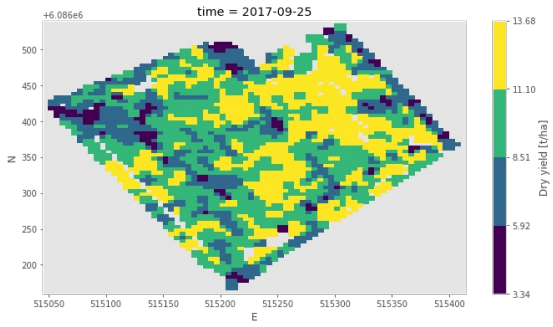
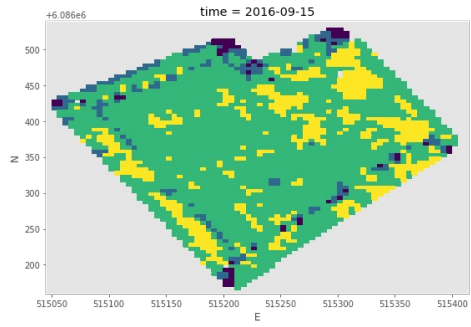
In [15]:

```
mh_harvest_dry_yield_intervals = {field: intervals for field, intervals in zip(gdf_mh_overview.index, [4] * len(gdf_mh_overview.index))} # As requested by MLJE
plot_yield_intervals(gdf_mh_overview, mh_area_ser, mh_harvest_dry_yield_intervals, output_path='./majshelsaед')

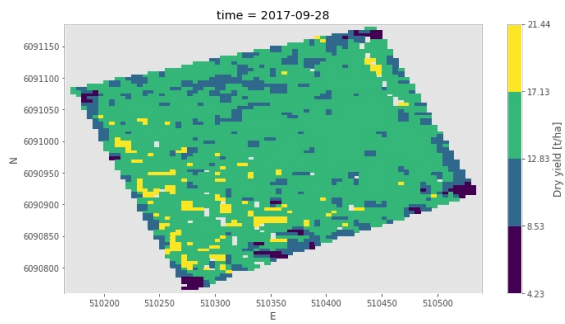
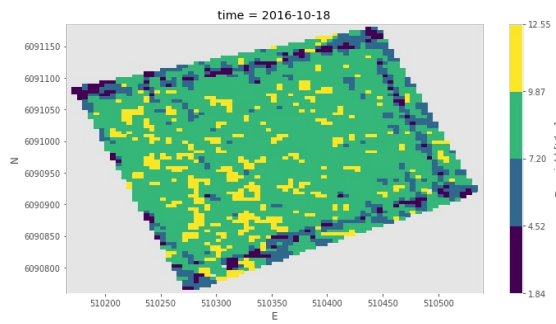
mh_yield_intervals_figures = sorted(glob.glob('./majshelsaед/*_yield_intervals.pdf'), key=lambda name: int(name.split('_')[-3]))
subprocess.run(['pdfunite'] + mh_yield_intervals_figures + ['./majshelsaед/all_yield_intervals_majshelsaед.pdf'])
```



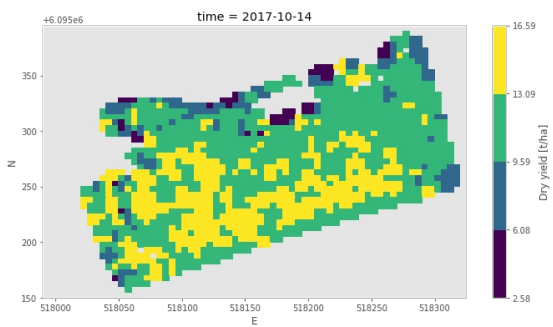
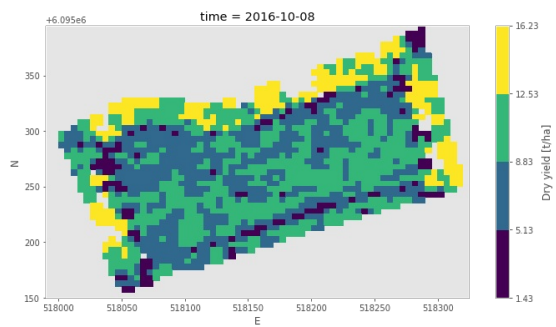
ID DDS: 3.0
2.26 ha



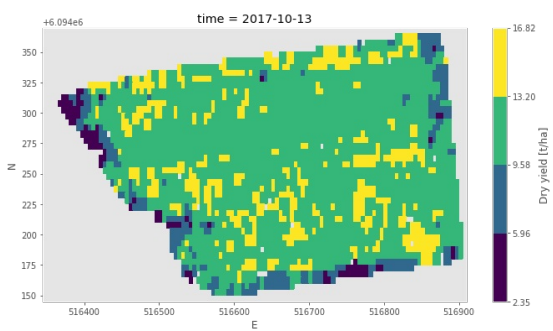
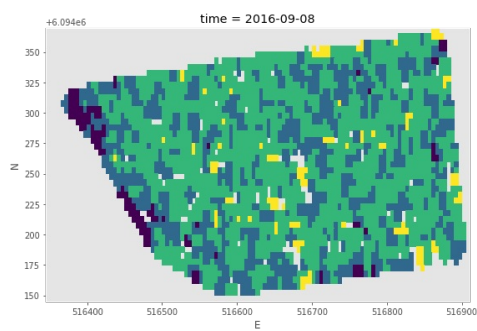
ID DDS: 18.0
7.28 ha



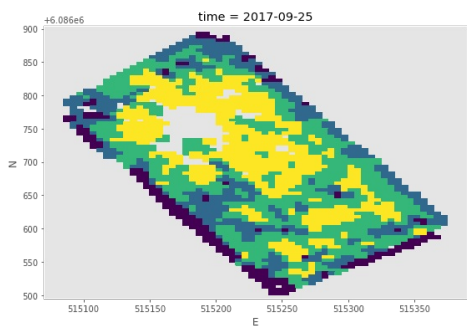
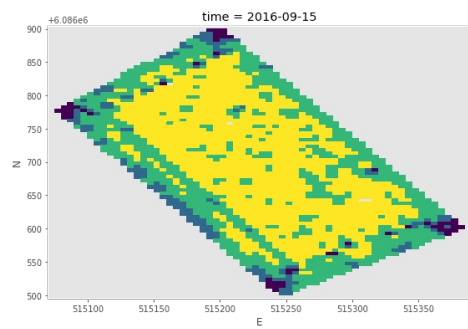
ID DDS: 46.0
9.08 ha



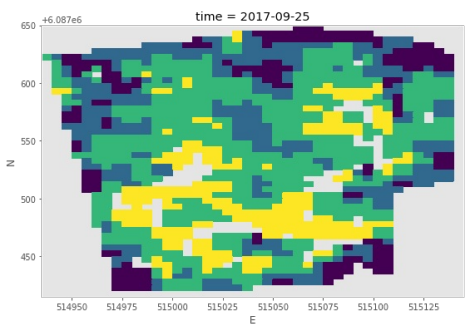
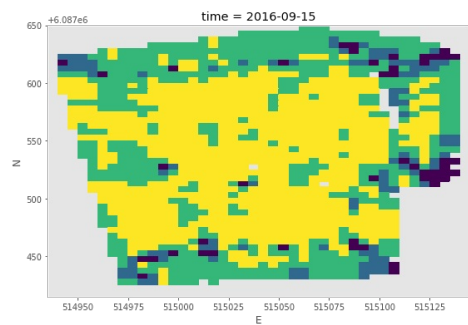
ID DDS: 49.0
4.29 ha



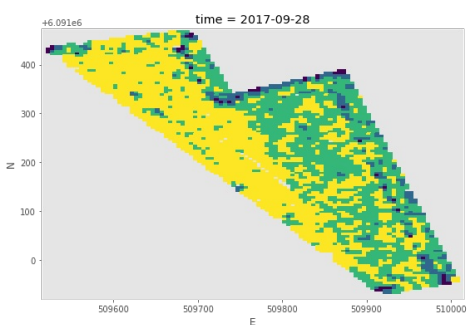
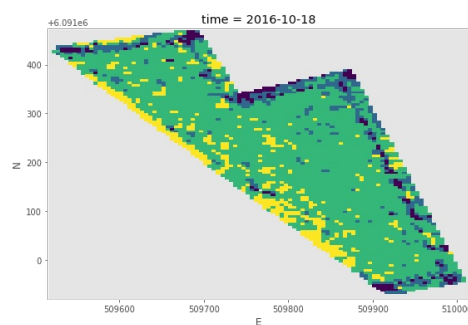
ID DDS: 56.0
8.69 ha



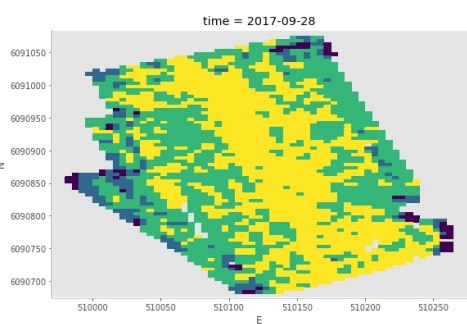
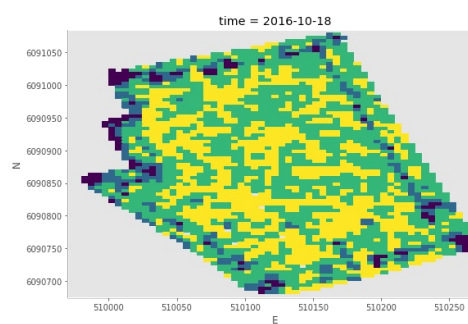
ID_DDS: 64.0
5.91 ha



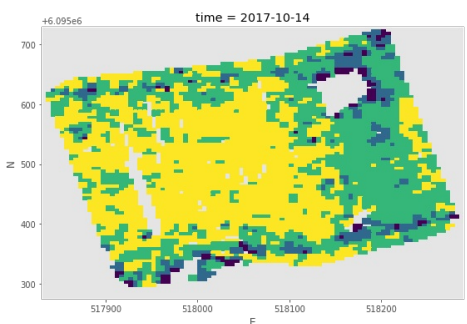
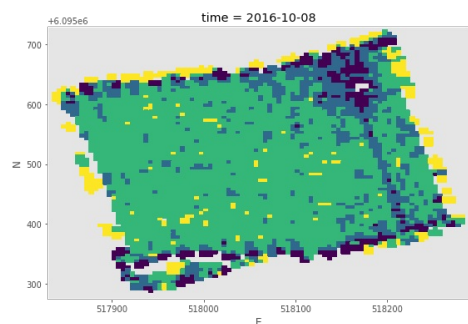
ID_DDS: 103.0
3.70 ha



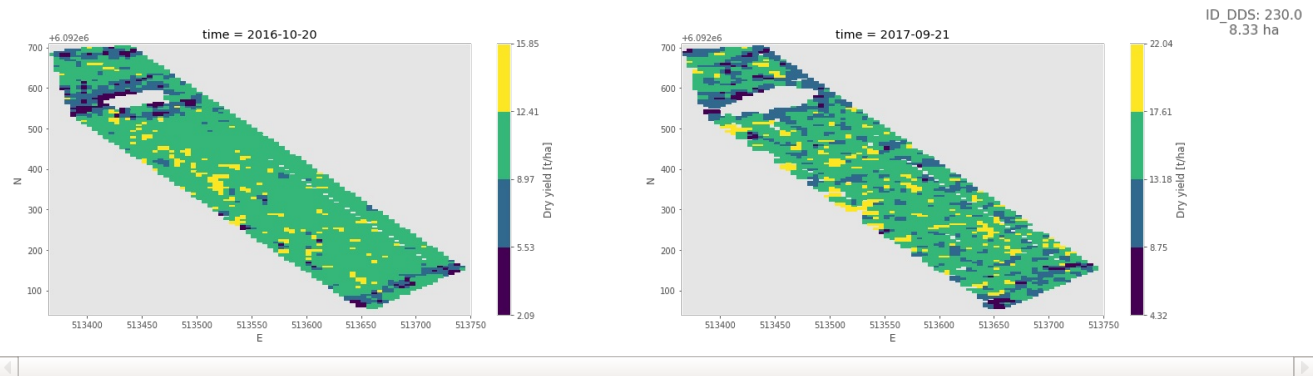
ID_DDS: 124.0
10.58 ha



ID_DDS: 175.0
7.41 ha



ID_DDS: 192.0
13.50 ha



Out[15]:

```
CompletedProcess(args=['pdfunite', './majshelsaed/DDS_field_3_yield_intervals.pdf', './majshelsaed/DDS_field_18_yield_intervals.pdf', './majshelsaed/DDS_field_46_yield_intervals.pdf', './majshelsaed/DDS_field_49_yield_intervals.pdf', './majshelsaed/DDS_field_56_yield_intervals.pdf', './majshelsaed/DDS_field_64_yield_intervals.pdf', './majshelsaed/DDS_field_103_yield_intervals.pdf', './majshelsaed/DDS_field_124_yield_intervals.pdf', './majshelsaed/DDS_field_175_yield_intervals.pdf', './majshelsaed/DDS_field_192_yield_intervals.pdf', './majshelsaed/DDS_field_230_yield_intervals.pdf', './majshelsaed/all_yield_intervals_majshelsaed.pdf'], returncode=0)
```

Statistics across years

Temporal statistics (aggregations over time) are computed and displayed for each field. The computed statistics are:

- Temporal mean (mean)
- Temporal standard deviation (std)
- Temporal minimum (amin)
- Temporal maximum (amax)

Two sets of figures are shown:

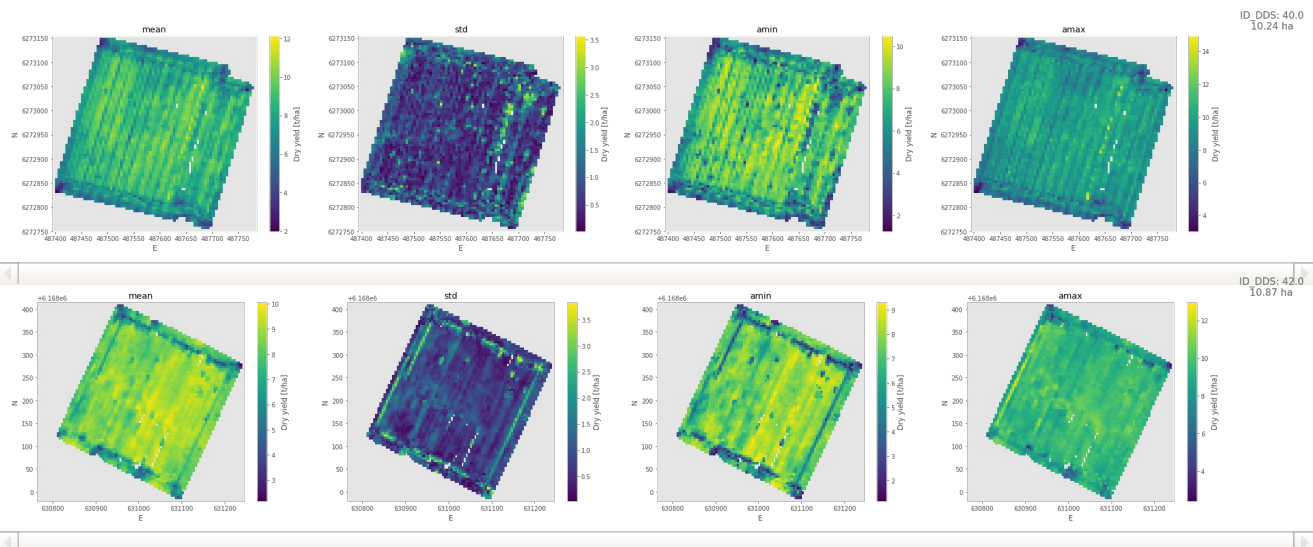
1. For each field the statistics are shown as continuous variables.
2. For each field the statistics are binned to 4 bins and shown as such. The percentage of the area falling in each bin is shown in a grey colored text next to the interval in the colorbar showing the interval colors and their borders.

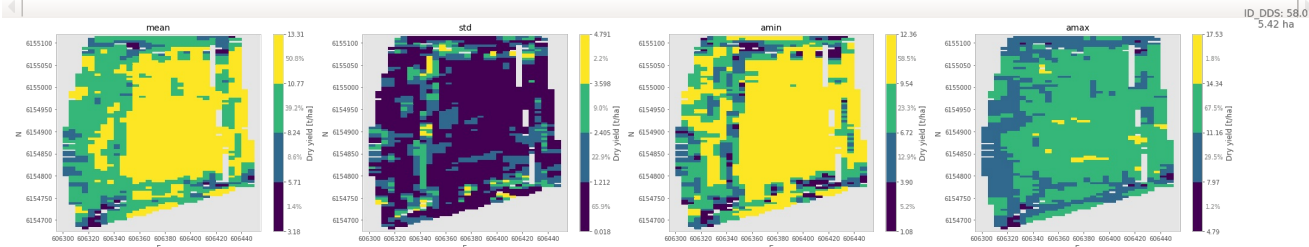
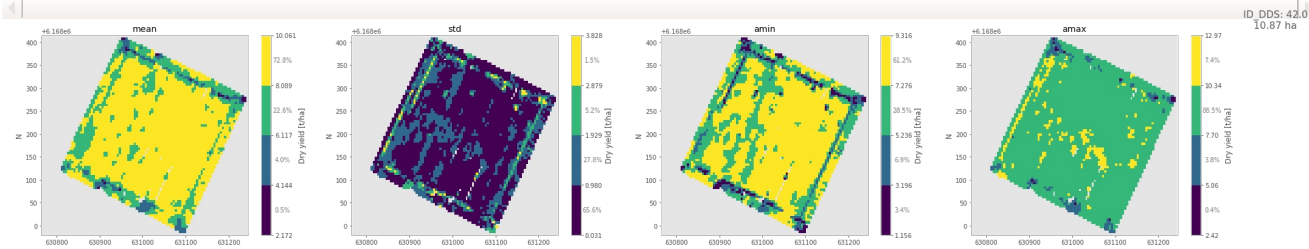
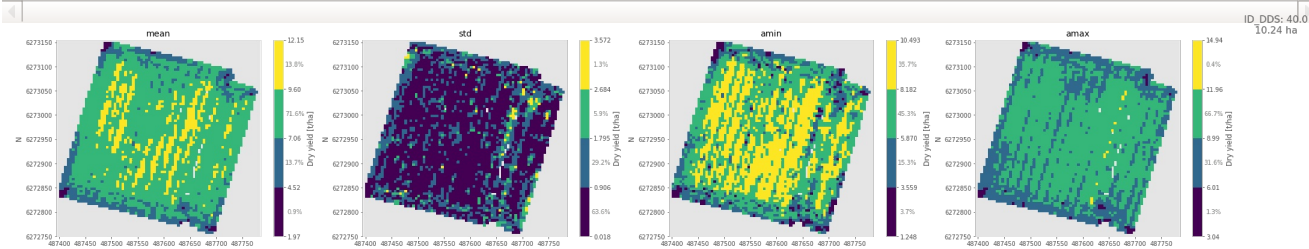
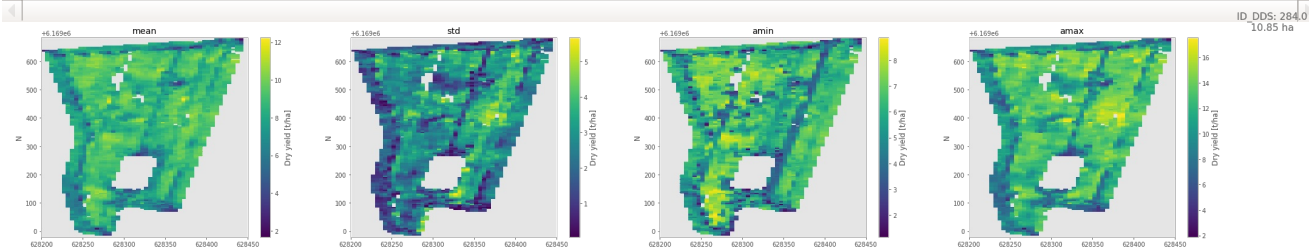
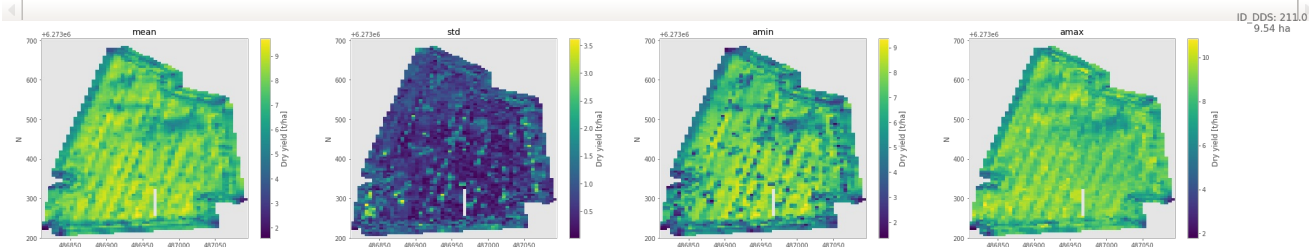
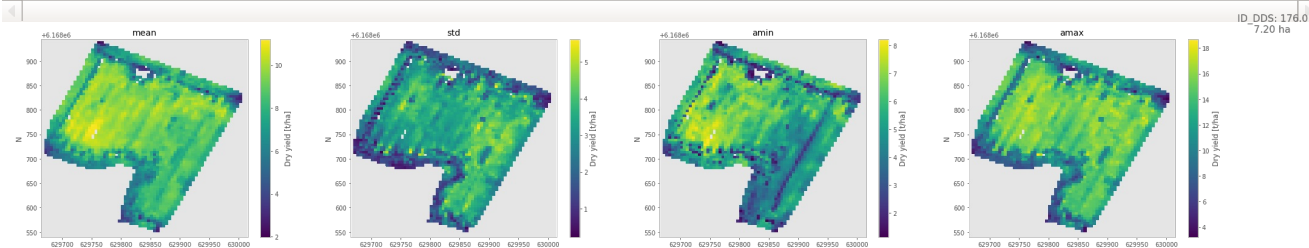
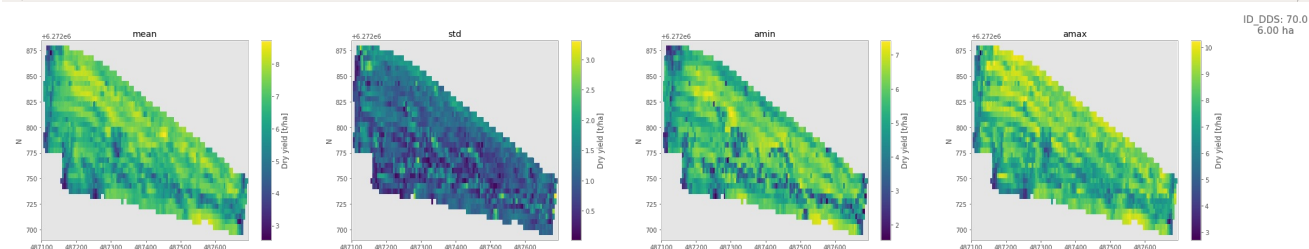
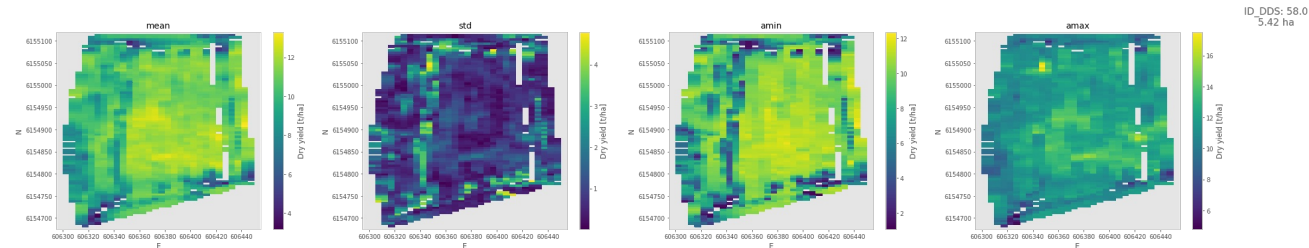
Vinterhvede

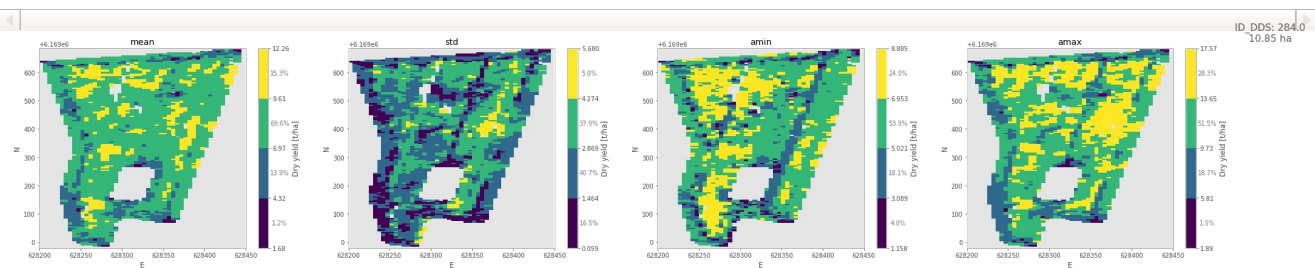
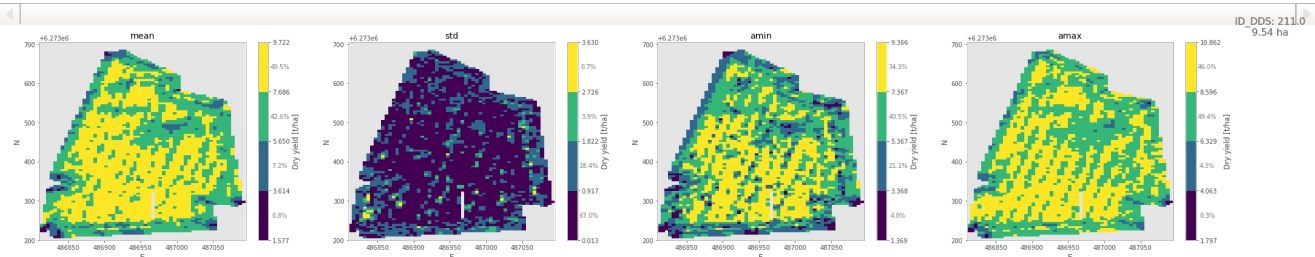
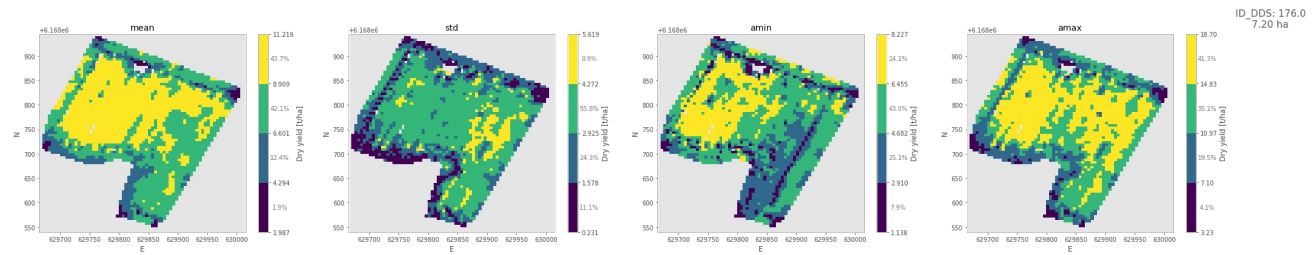
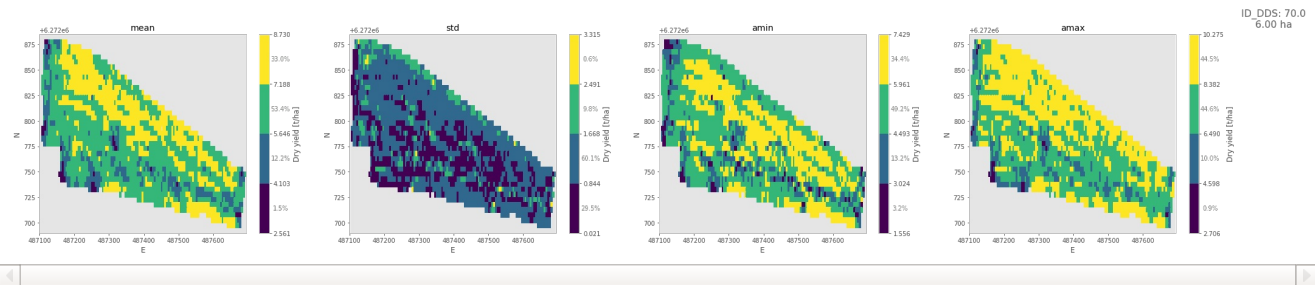
In [16]:

```
with warnings.catch_warnings():
    warnings.simplefilter("ignore")
    plot_yield_time_statistics(gdf_vh_overview, vh_area_ser, output_path='./vinterhvede')
    plot_yield_time_statistics(gdf_vh_overview, vh_area_ser, intervals=4, output_path='./vinterhvede')
```

```
vh_yield_time_statistics_figures = sorted(glob.glob('./vinterhvede/*_yield_time_statistics.pdf'), key=lambda name: int(name.split('_')[5]))
subprocess.run(['pdfunite'] + vh_yield_time_statistics_figures + ['./vinterhvede/all_yield_time_statistics_vinterhvede.pdf'])
```







Out[16]:

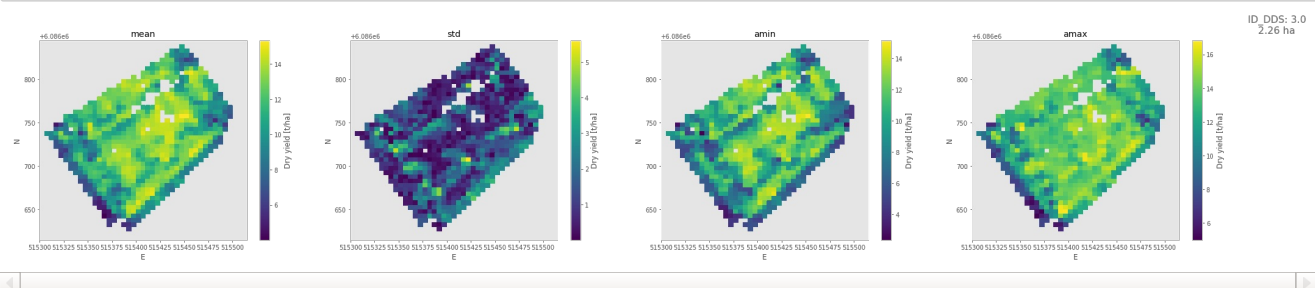
```
CompletedProcess(args=['pdfunite', './vinterhvede/DDS_field_40_intervals_yield_time_statistics.pdf',
'./vinterhvede/DDS_field_40_full_yield_time_statistics.pdf', './vinterhvede/DDS_field_42_full_yield
_time_statistics.pdf', './vinterhvede/DDS_field_42_intervals_yield_time_statistics.pdf', './vinterhv
ede/DDS_field_58_full_yield_time_statistics.pdf', './vinterhvede/DDS_field_58_intervals_yield_time_s
tatistics.pdf', './vinterhvede/DDS_field_70_intervals_yield_time_statistics.pdf', './vinterhvede/DDS
_field_70_full_yield_time_statistics.pdf', './vinterhvede/DDS_field_176_full_yield_time_statistics.p
df', './vinterhvede/DDS_field_176_intervals_yield_time_statistics.pdf', './vinterhvede/DDS_field_211
_full_yield_time_statistics.pdf', './vinterhvede/DDS_field_211_intervals_yield_time_statistics.pdf',
'./vinterhvede/DDS_field_284_full_yield_time_statistics.pdf', './vinterhvede/DDS_field_284_interval
s_yield_time_statistics.pdf', './vinterhvede/all_yield_time_statistics_vinterhvede.pdf'], returncode
=0)
```

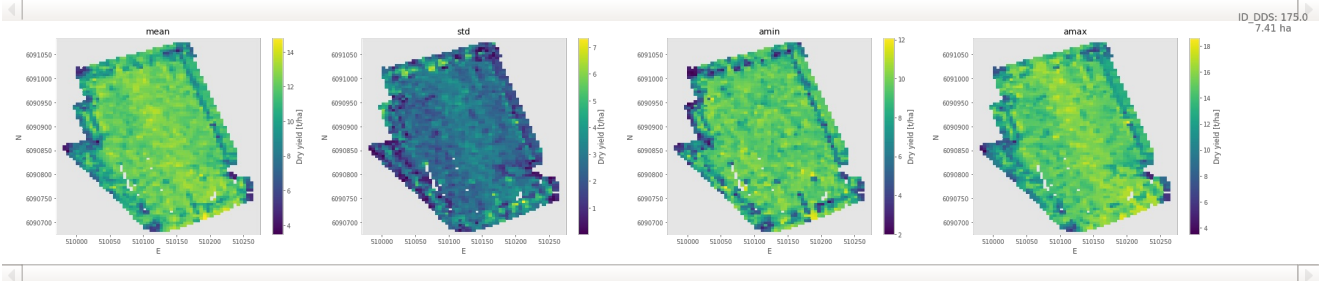
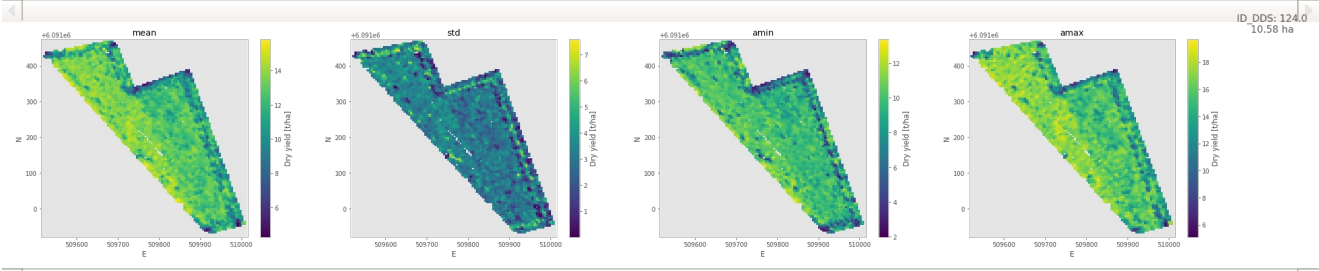
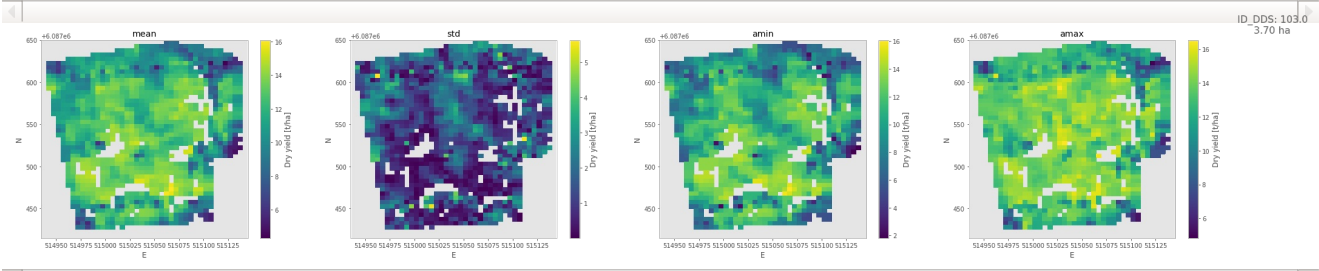
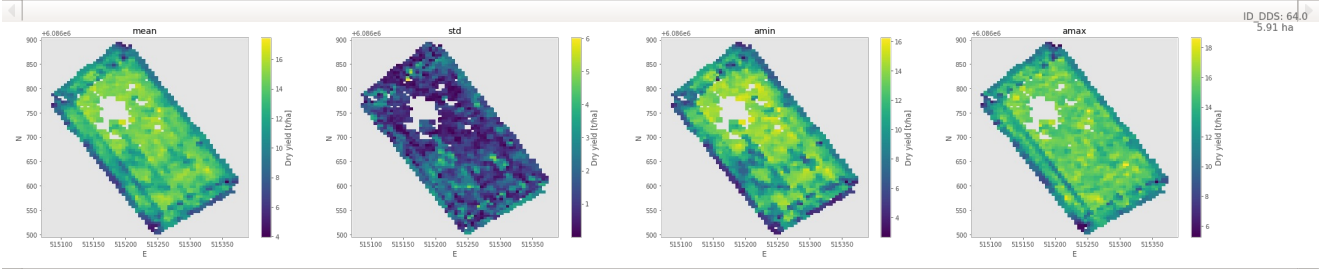
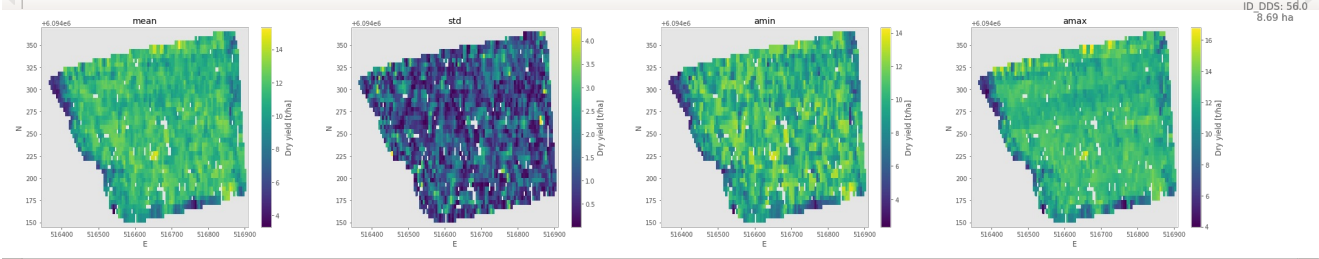
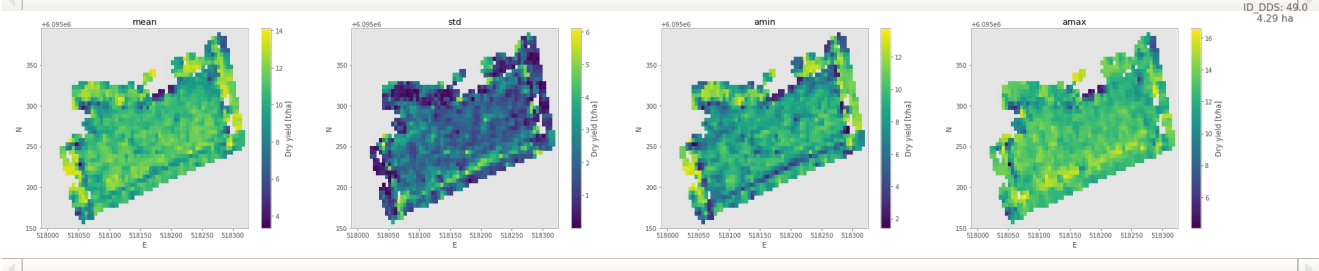
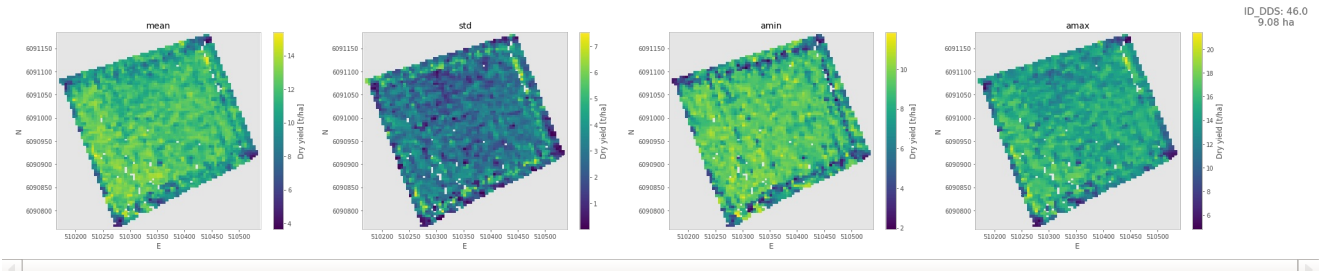
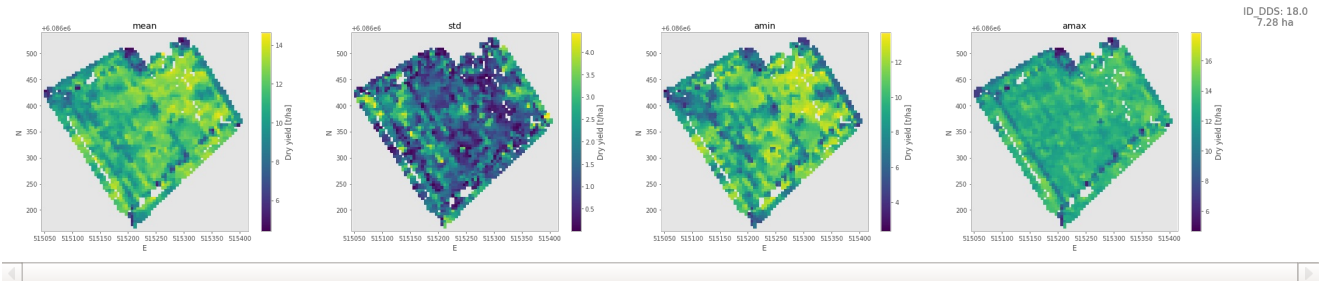
Majshelsaед

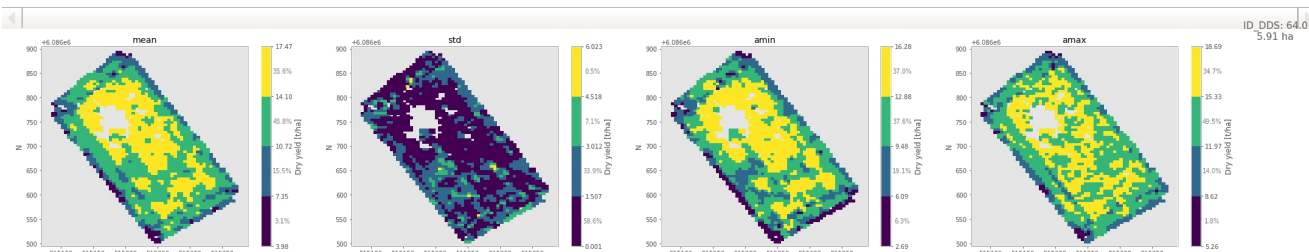
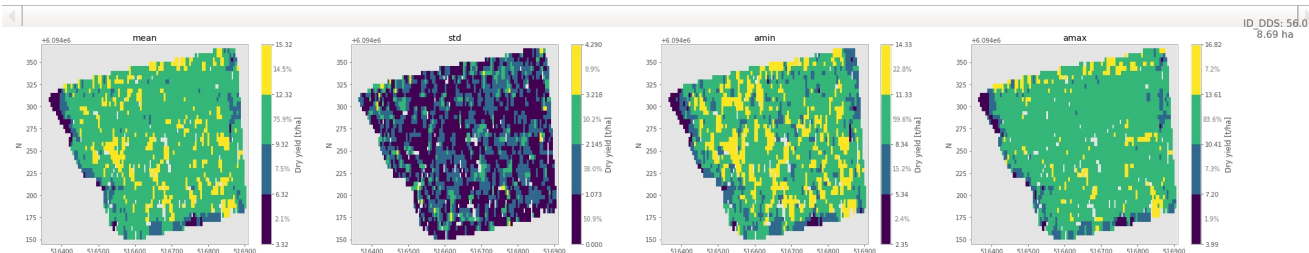
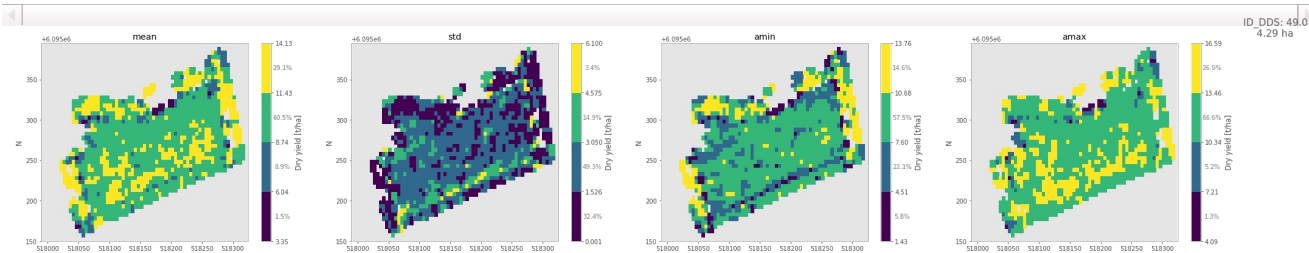
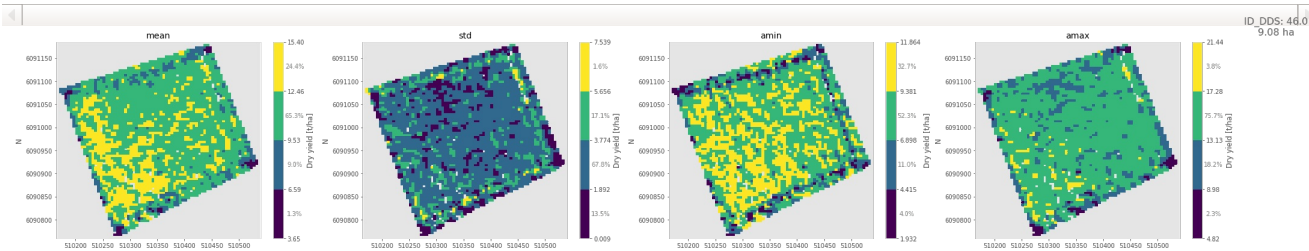
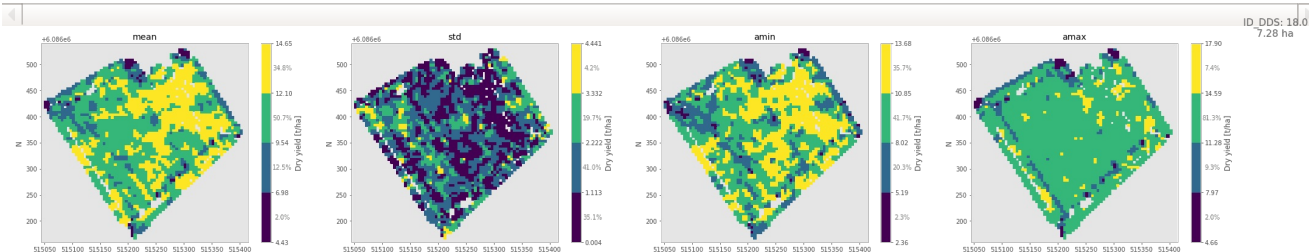
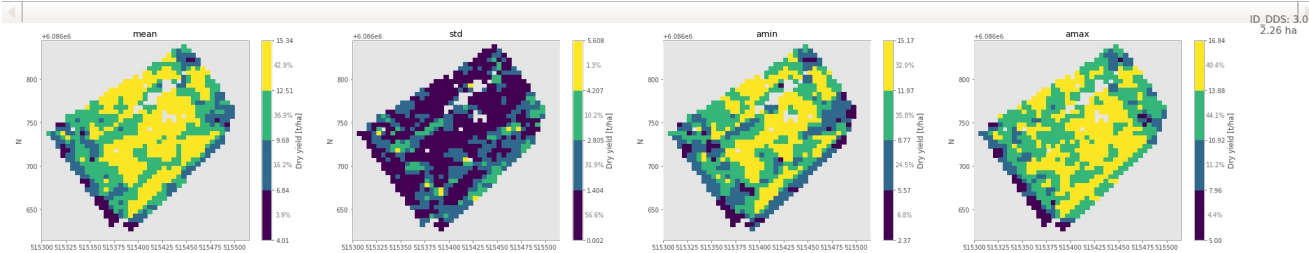
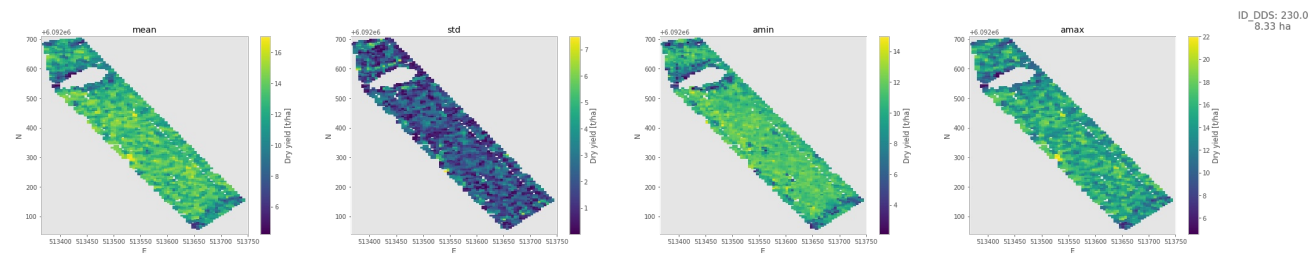
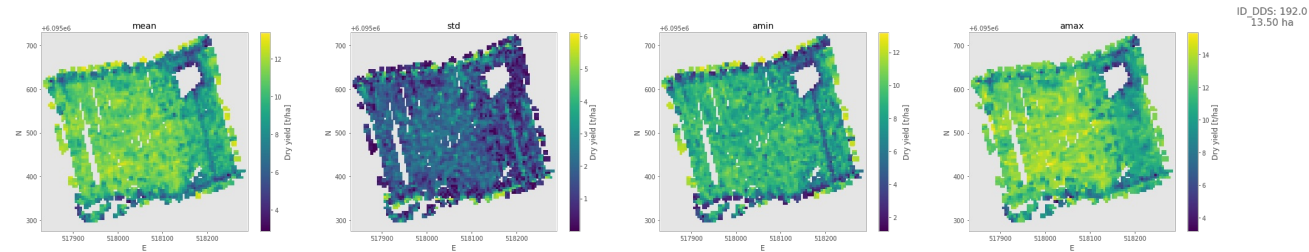
In [17]:

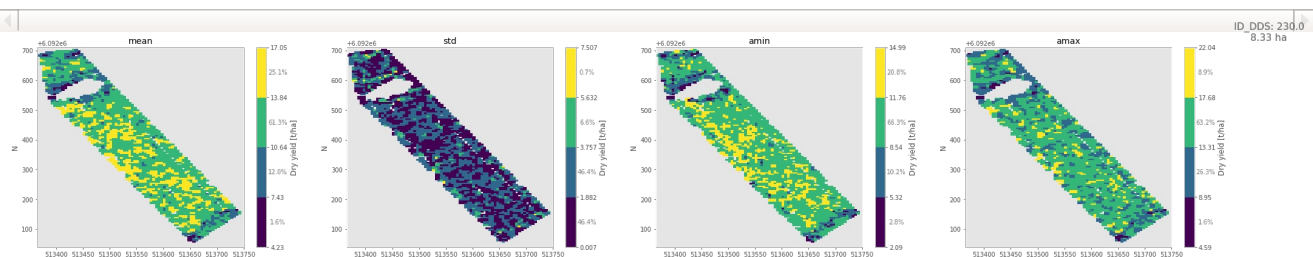
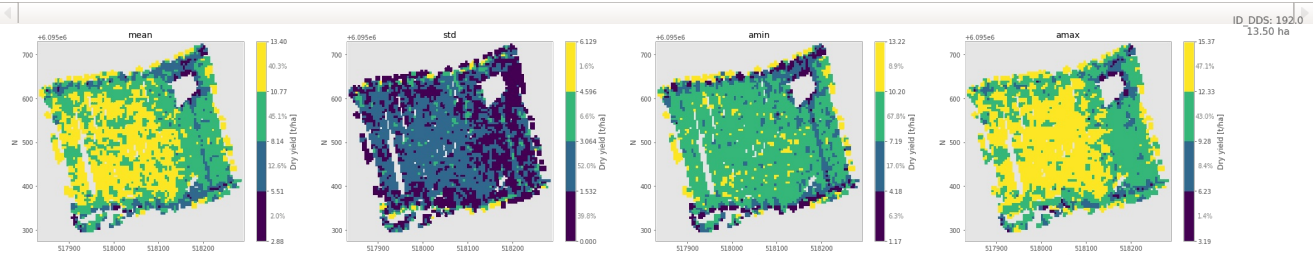
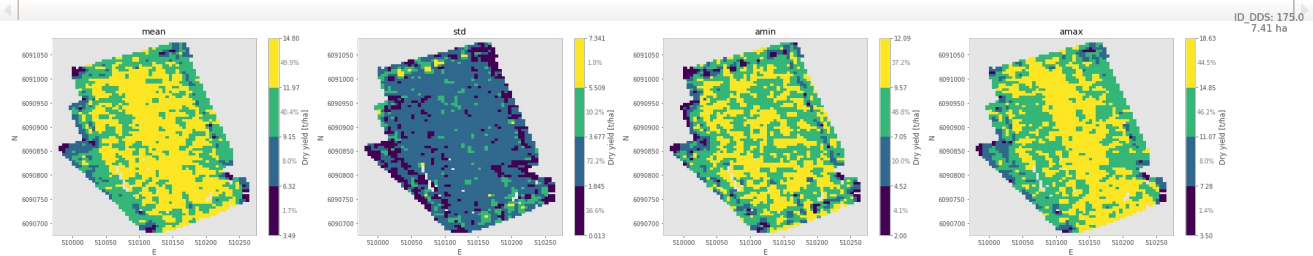
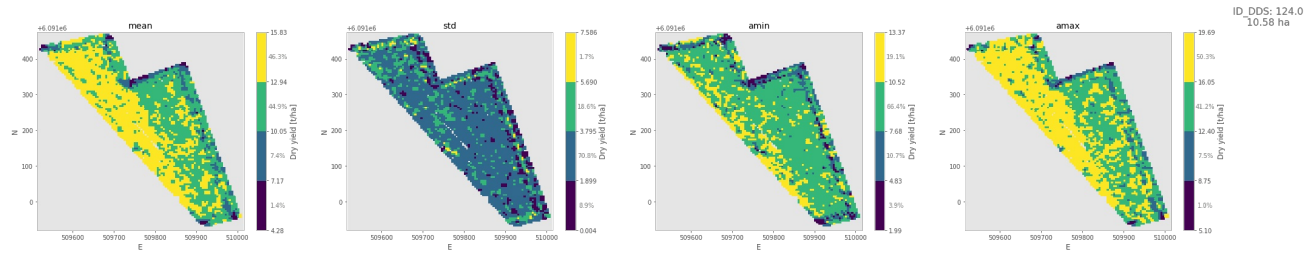
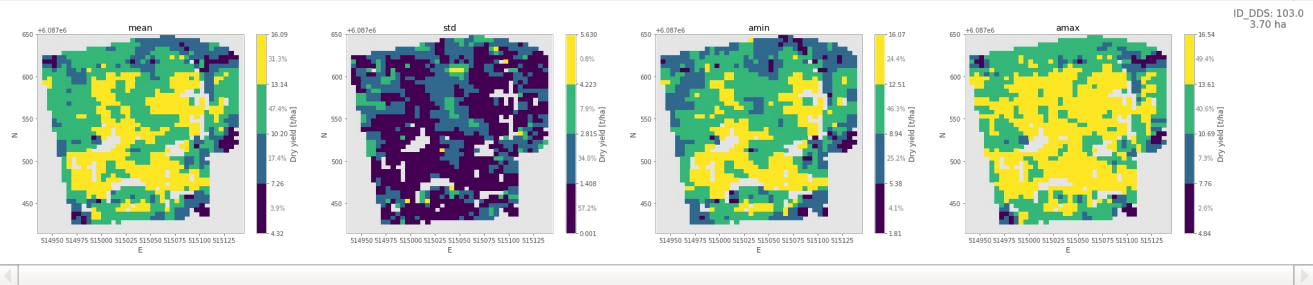
```
with warnings.catch_warnings():
    warnings.simplefilter("ignore")
    plot_yield_time_statistics(gdf_mh_overview, mh_area_ser, output_path='./majshelsaед')
    plot_yield_time_statistics(gdf_mh_overview, mh_area_ser, intervals=4, output_path='./majshelsaед')

mh_yield_time_statistics_figures = sorted(glob.glob('./majshelsaед/*_yield_time_statistics.pdf'), key=lambda name
: int(name.split('_')[-5]))
subprocess.run(['pdfunite'] + mh_yield_time_statistics_figures + ['./majshelsaед/all_yield_time_statistics_majshe
lsaед.pdf'])
```









Out[17]:

```
CompletedProcess(args=['pdfunite', './majshelsaed/DDS_field_3_intervals_yield_time_statistics.pdf',
'./majshelsaed/DDS_field_3_full_yield_time_statistics.pdf', './majshelsaed/DDS_field_18_intervals_yi
eld_time_statistics.pdf', './majshelsaed/DDS_field_18_full_yield_time_statistics.pdf', './majshelsae
d/DDS_field_46_full_yield_time_statistics.pdf', './majshelsaed/DDS_field_46_intervals_yield_time_sta
tistics.pdf', './majshelsaed/DDS_field_49_full_yield_time_statistics.pdf', './majshelsaed/DDS_field_
49_intervals_yield_time_statistics.pdf', './majshelsaed/DDS_field_56_full_yield_time_statistics.pdf'
, './majshelsaed/DDS_field_56_intervals_yield_time_statistics.pdf', './majshelsaed/DDS_field_64_full
_yield_time_statistics.pdf', './majshelsaed/DDS_field_64_intervals_yield_time_statistics.pdf', './ma
jshelsaed/DDS_field_103_intervals_yield_time_statistics.pdf', './majshelsaed/DDS_field_103_full_yiel
d_time_statistics.pdf', './majshelsaed/DDS_field_124_full_yield_time_statistics.pdf', './majshelsaed
/DDS_field_124_intervals_yield_time_statistics.pdf', './majshelsaed/DDS_field_175_full_yield_time_st
atistics.pdf', './majshelsaed/DDS_field_175_intervals_yield_time_statistics.pdf', './majshelsaed/DDS
_field_192_full_yield_time_statistics.pdf', './majshelsaed/DDS_field_192_intervals_yield_time statis
tics.pdf', './majshelsaed/DDS_field_230_intervals_yield_time_statistics.pdf', './majshelsaed/DDS_fie
ld_230_full_yield_time_statistics.pdf', './majshelsaed/all_yield_time_statistics_majshelsaed.pdf'],
returncode=0)
```

Yield variance classes across years

For each field, we compute the temporal (across year) mean and standard deviation yield. We then group each cell into one of three groups:

1. **Low mean, low variance** (red): The temporal mean in the cell is smaller than the field (spatial and temporal) mean AND the temporal standard deviation in the cell is smaller than the spatial field average of the temporal standard deviation.
2. **High variance** (yellow): The temporal standard deviation in the cell is larger than (or equal to) spatial field average of the the temporal standard deviation.
3. **High mean, low variance** (green): The temporal mean in the cell is larger than (or equal to) the field (spatial and temporal) mean AND the temporal standard deviation in the cell is smaller than the spatial field average of the temporal standard deviation.

Only cells for which yield data is available for all years are shown.

For each class, the following is shown as part of the class label:

- The percentage of the area covered by the cells in the class.
- The average (temporal and spatial mean) yield of the cells in the class.
- The standard deviation (temporal and spatial SD) of the yield of the cells in the class.

Finally, the field average (temporal and spatial mean) yield and spatial average of the temporal standard deviation (SD) of the yield is shown in the upper right corner of the figure.

Note

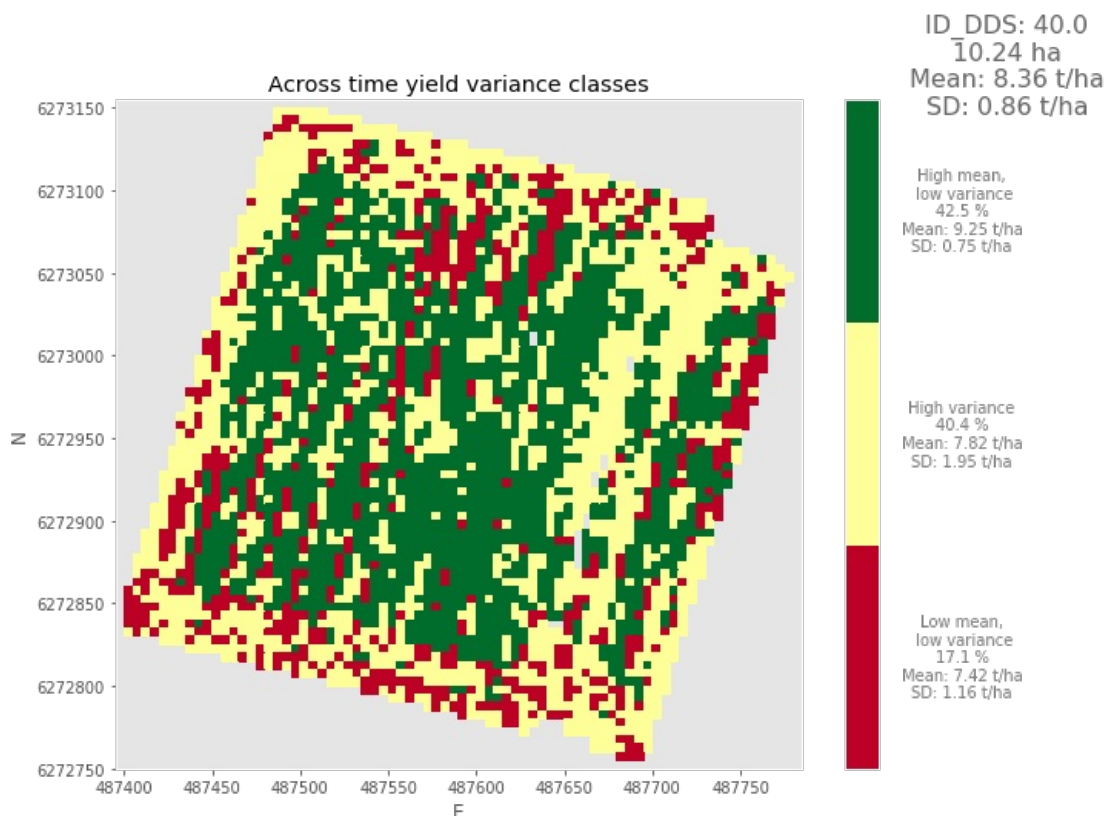
- Computing the average is a linear operation. Thus, one is always guaranteed that the mean yield of each class is lower, respectively higher, than the field mean yield.
- Computing standard deviation is, on the other hand, not a linear operation. Thus, the class standard deviation may be lower, respectively higher, than the field spatial average of the temporal standard deviation, e.g. as is seen the first figure where the "low mean, low variance" class has a higher standard deviation than the field spatial average of the temporal standard deviation.

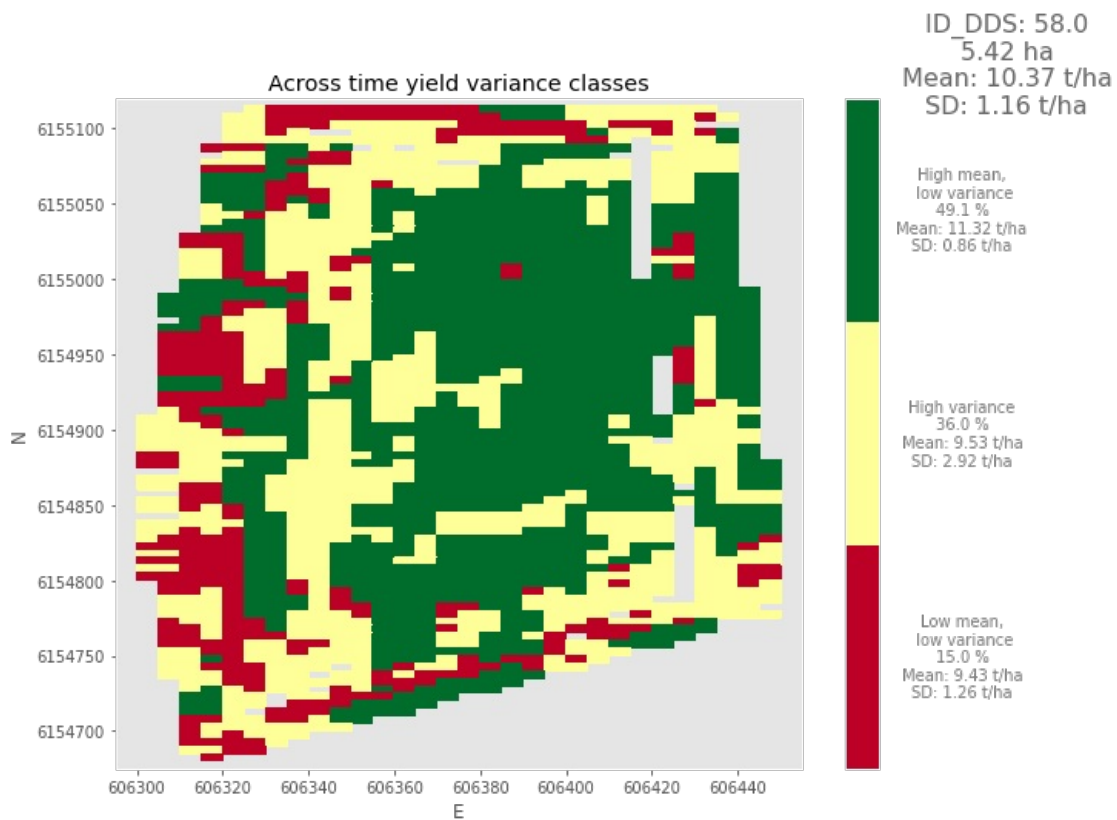
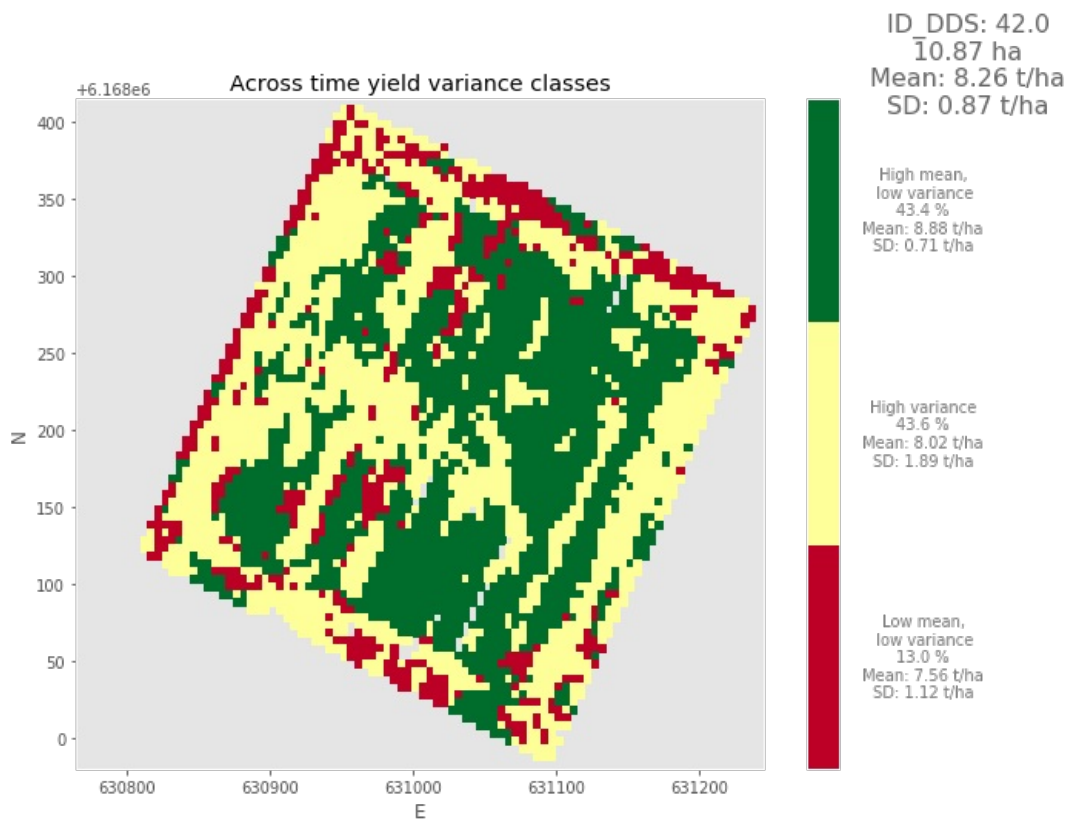
Vinterhvede

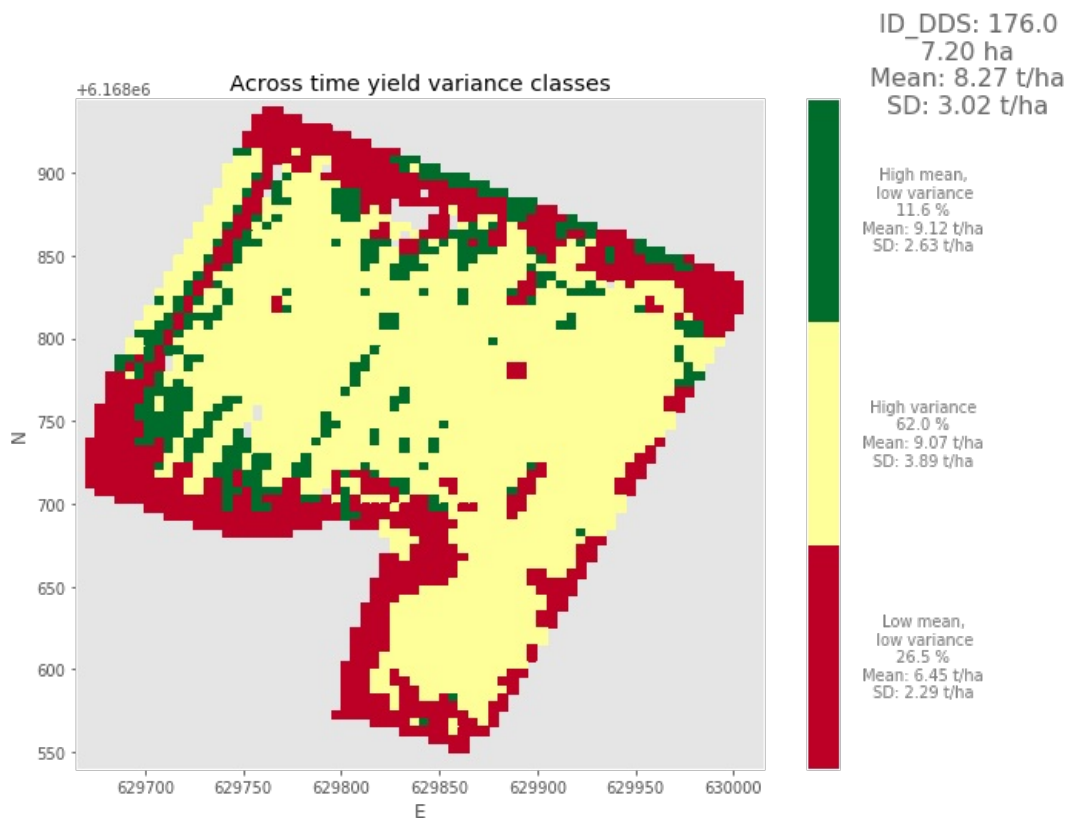
In [18]:

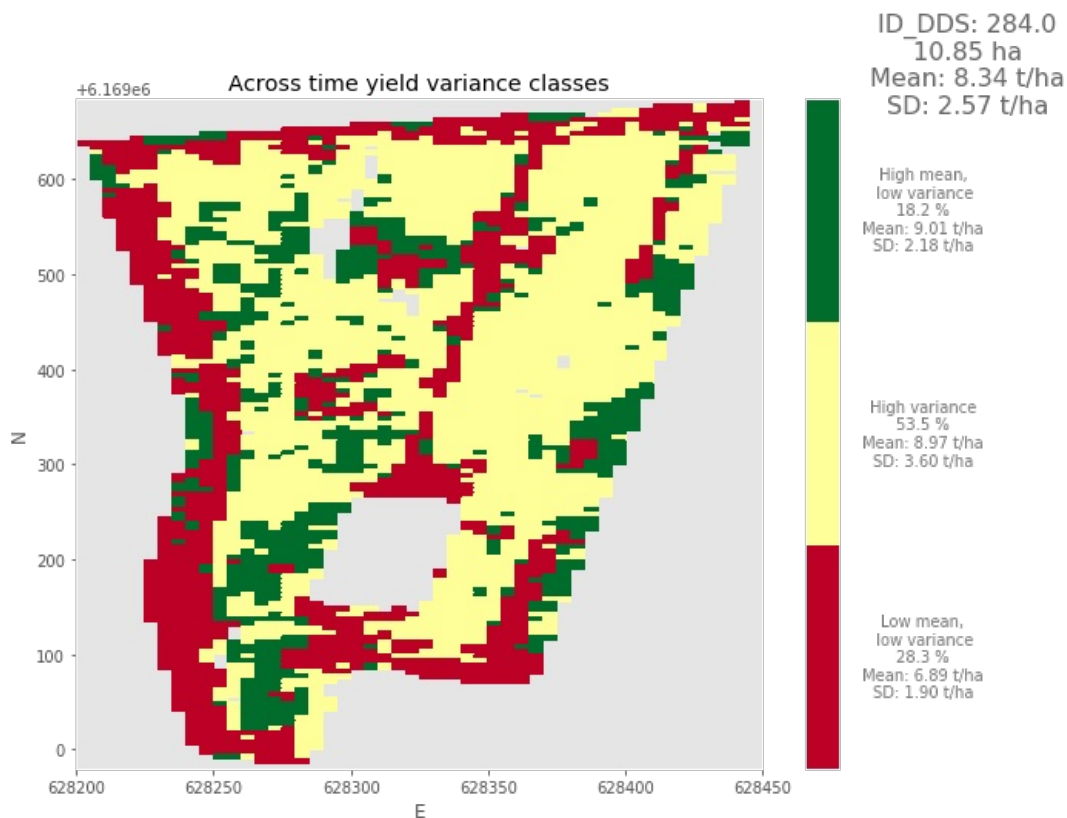
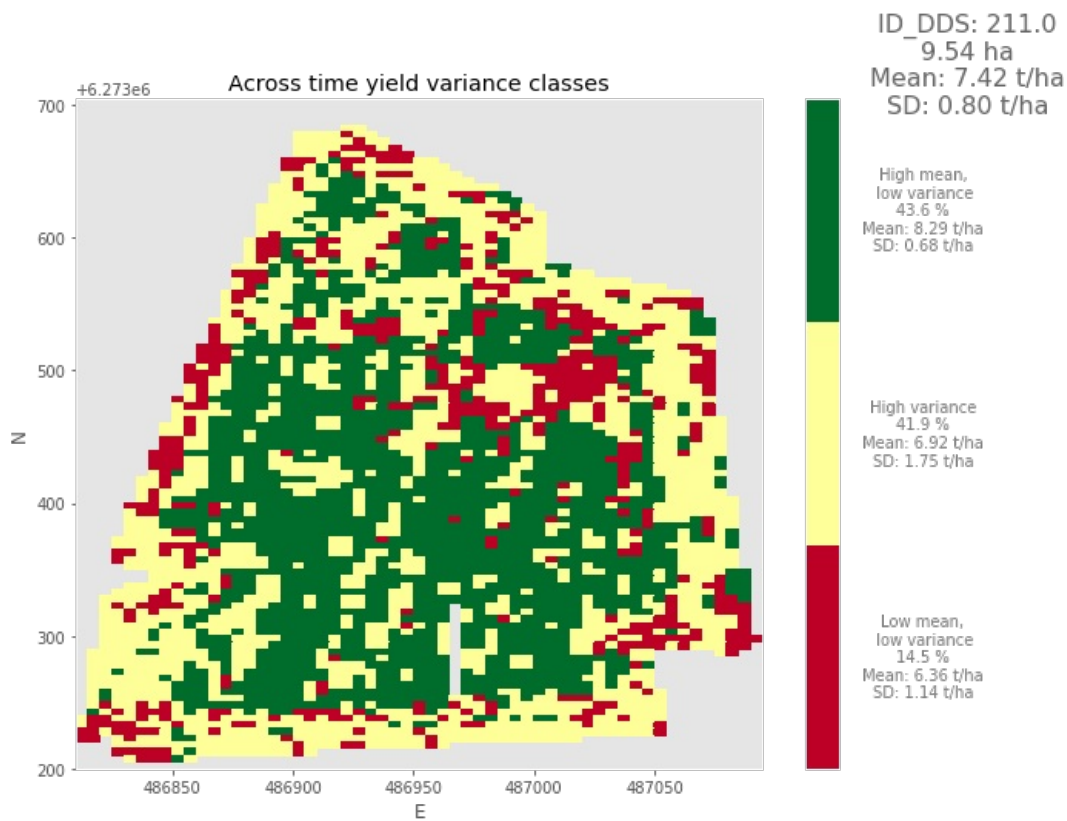
```
plot_yield_variance_classes(gdf_vh_overview, vh_area_ser, output_path='./vinterhvede')

vh_yield_time_statistics_figures = sorted(glob.glob('./vinterhvede/*_yield_variance_classes.pdf'), key=lambda name: int(name.split('_')[-4]))
subprocess.run(['pdfunite'] + vh_yield_time_statistics_figures + ['./vinterhvede/all_yield_variance_classes_vinterhvede.pdf'])
```









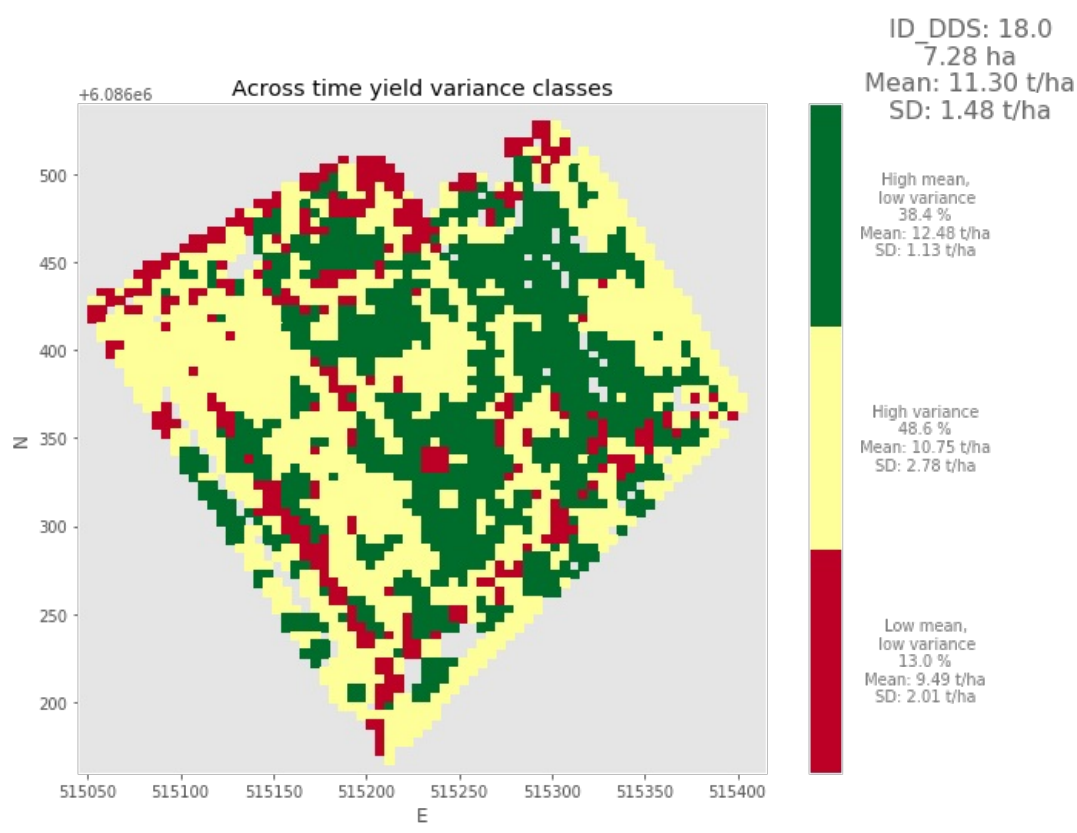
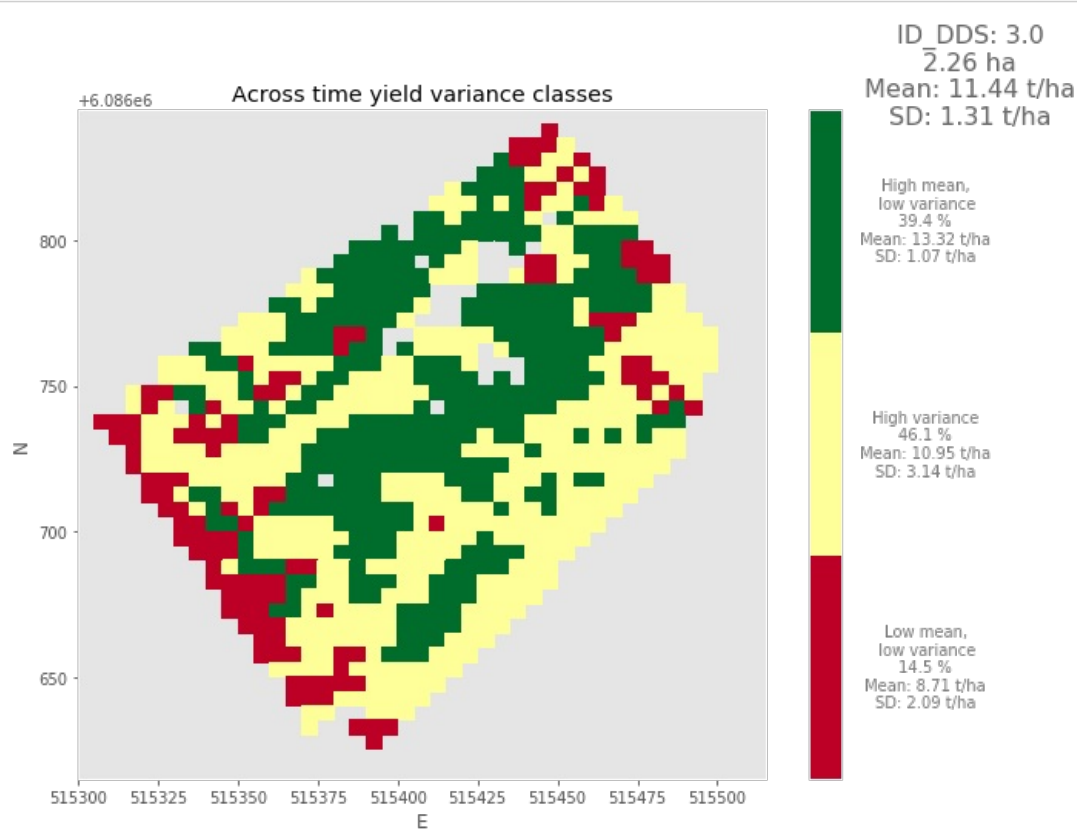
Out[18]:

```
CompletedProcess(args=['pdfunite', './vinterhvede/DDS_field_40_yield_variance_classes.pdf', './vinterhvede/DDS_field_42_yield_variance_classes.pdf', './vinterhvede/DDS_field_58_yield_variance_classes.pdf', './vinterhvede/DDS_field_70_yield_variance_classes.pdf', './vinterhvede/DDS_field_176_yield_variance_classes.pdf', './vinterhvede/DDS_field_211_yield_variance_classes.pdf', './vinterhvede/DDS_field_284_yield_variance_classes.pdf', './vinterhvede/all_yield_variance_classes_vinterhvede.pdf'], returncode=0)
```

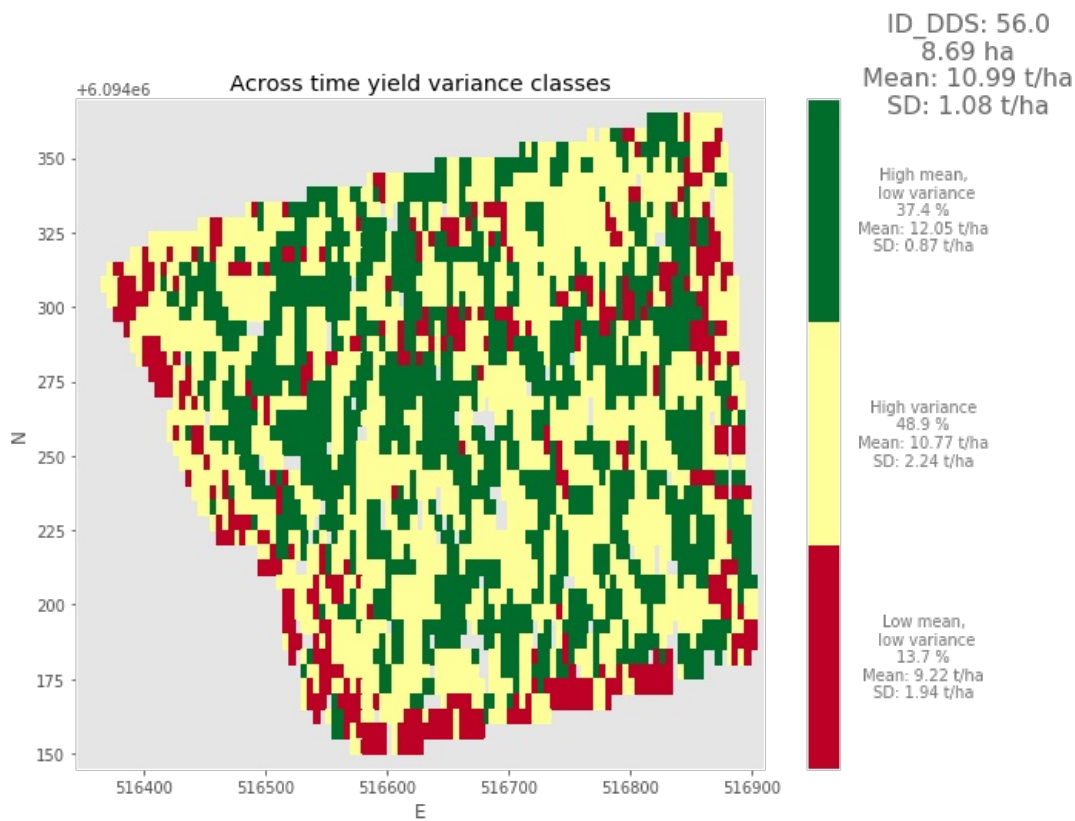
Majshelsaed

```
In [19]:
plot_yield_variance_classes(gdf_mh_overview, mh_area_ser, output_path='./majshelsaed')

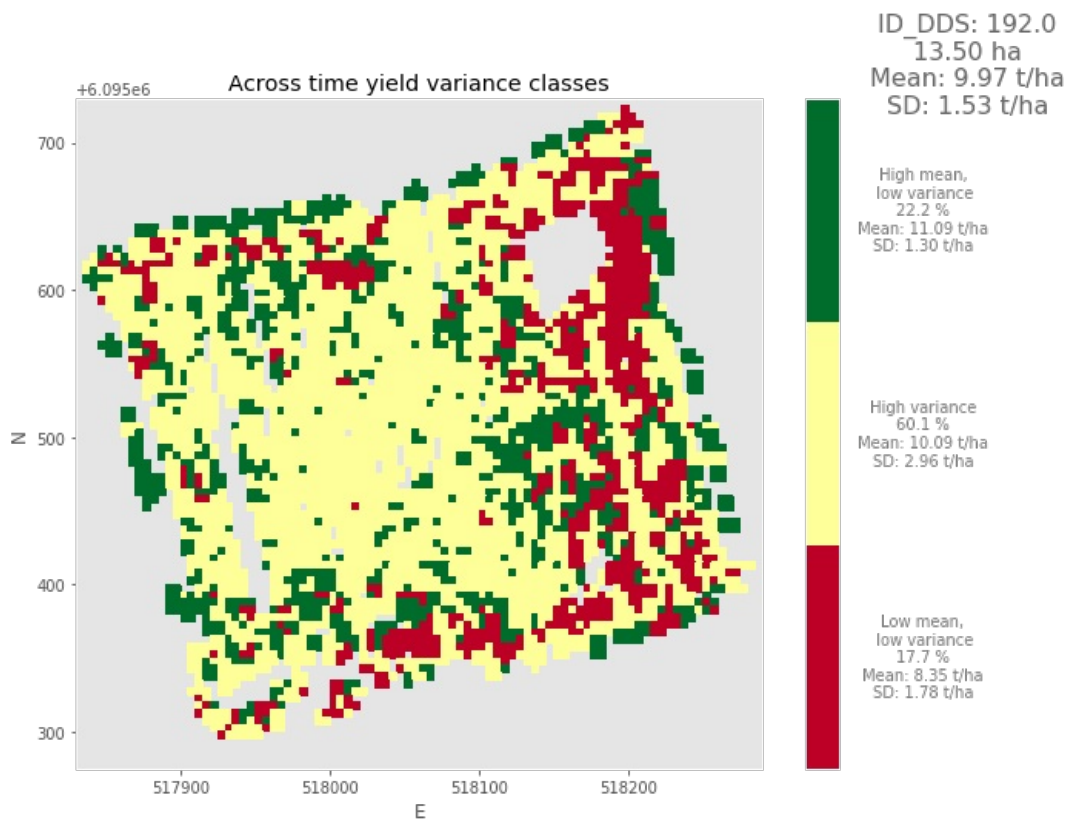
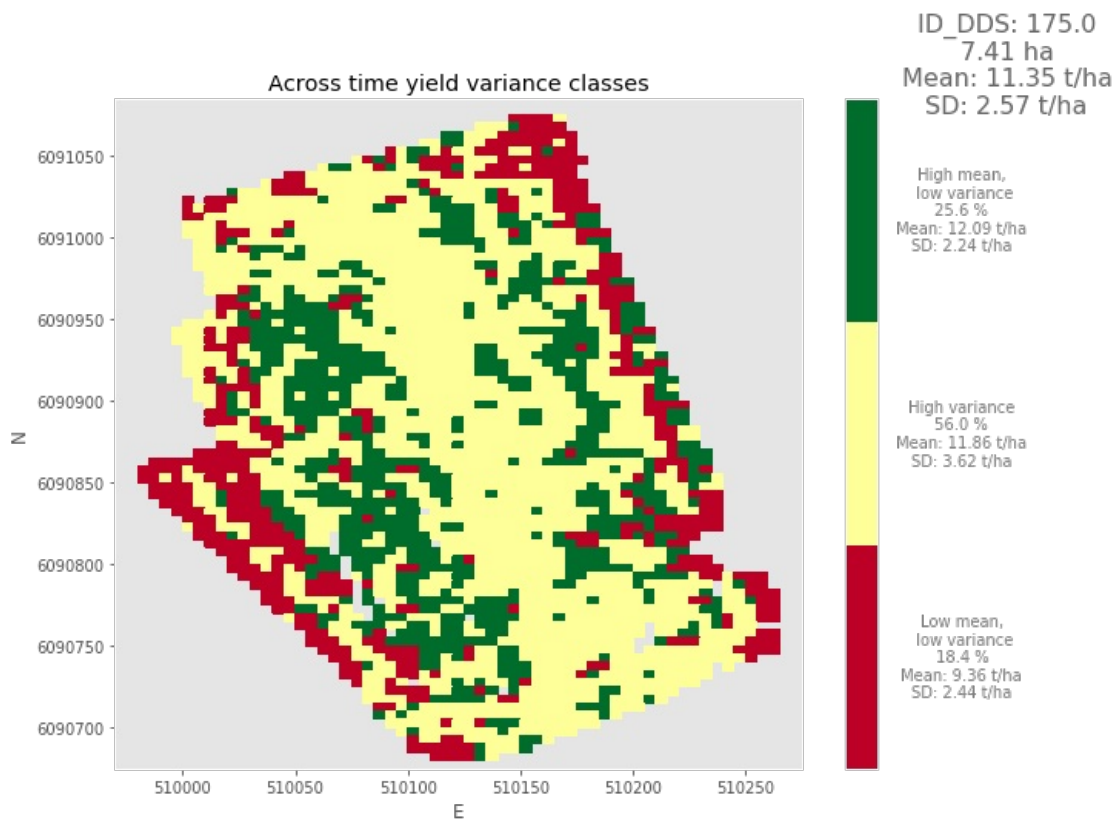
mh_yield_time_statistics_figures = sorted(glob.glob('./majshelsaed/*_yield_variance_classes.pdf'), key=lambda name: int(name.split('_')[-4]))
subprocess.run(['pdfunite'] + mh_yield_time_statistics_figures + ['./majshelsaed/all_yield_variance_classes_majshelsaed.pdf'])
```

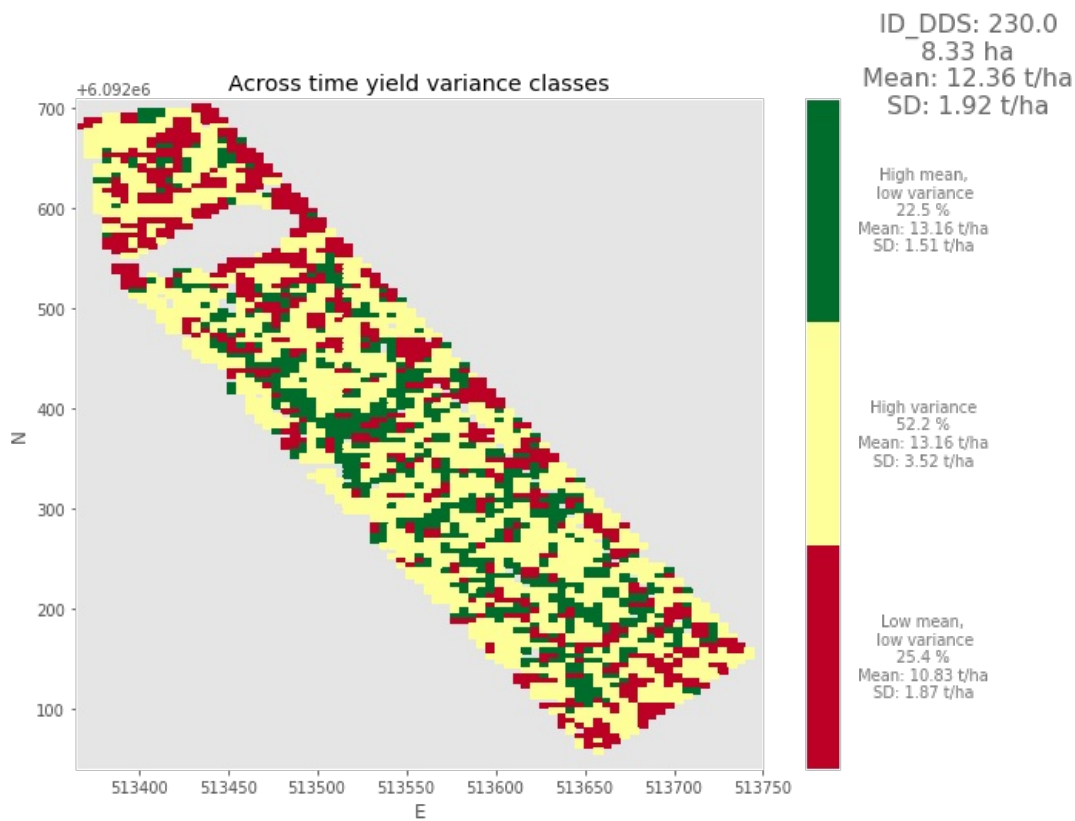












Out[19]:

```
CompletedProcess(args=['pdfunite', './majshelsaed/DDS_field_3_yield_variance_classes.pdf', './majshelsaed/DDS_field_18_yield_variance_classes.pdf', './majshelsaed/DDS_field_46_yield_variance_classes.pdf', './majshelsaed/DDS_field_49_yield_variance_classes.pdf', './majshelsaed/DDS_field_56_yield_variance_classes.pdf', './majshelsaed/DDS_field_64_yield_variance_classes.pdf', './majshelsaed/DDS_field_103_yield_variance_classes.pdf', './majshelsaed/DDS_field_124_yield_variance_classes.pdf', './majshelsaed/DDS_field_175_yield_variance_classes.pdf', './majshelsaed/DDS_field_192_yield_variance_classes.pdf', './majshelsaed/DDS_field_230_yield_variance_classes.pdf', './majshelsaed/all_yield_variance_classes_majshelsaed.pdf'], returncode=0)
```