Observing Soil Moisture Anomalies for Wildfire Prediction

Soil moisture is one of the World Meteorological Organisation's *Essential Climate Variables* and is an important part of the water cycle through its role in regulating evapotranspiration. It has also been shown that there is a relationship between soil moisture deviations from a long-term mean value (or soil moisture anomalies) and the occurence and extend of wildfires. In particular, Bartsch et al. showed this relationship for forest fires in Siberia. They were able to demonstrate that positive soil moisture anomalies (unusually moist conditions) prevent the outbreak of fires and limit their extend. On the other hand, negative soil moisture anomalies (unusually dry conditions) were linked to a hightened likelihood of wildfire occurence. Large wildfires generally did not occur in areas where the soil moisture was at least 5% above the long-term mean for that area.

Using data from the ESA Climate Change Initiative (CCI), this notebook aims to show how soil moisture anomalies can be used in the context of wildfire prediction. This exercise focusses on the large-scale wildfires that broke out in [Western Russia in 2010 (https://en.wikipedia.org/wiki/2010_Russian_wildfires)](https://en.wikipedia.org/wiki/2010_Russian_wildfires). The fires started in late July 2010 and lasted until early September that same year. They are counted amongst the most devastating forest fires in Russian history.

In [1]:

```
# Importing all the necessary libraries

import numpy as np
from netCDF4 import Dataset
import matplotlib.pyplot as plt
import geopandas as gpd
import pandas as pd
from shapely.geometry import Point
import os
import matplotlib
import plotly
import plotly.express as px
import plotly.graph_objects as go
import math

from ipywidgets import interact, widgets
#%matplotlib notebook
%matplotlib inline
```

For this exercise, three datasets are used: a dataset containing soil moisture values for 2010, a dataset containing burned area values for August 2010 and a dataset containing the long-term (1991-2010) mean soil moisture. All three datasets can be obtained through the [CCI Open Data Portal (http://cci.esa.int/data)](http://cci.esa.int/data). While the burned area was obtained using MODIS data, the soil moisture data was obtained using a combination of different sensors (AMI-WS, ASCAT, SMMR, SSM/I, TMI, AMSR-E, WindSat, AMSR2 and SMOS satellite instruments). All three datasets are summarised into a grid of resolution 0.25x0.25 degrees, covering the entire Earth. The burned area dataset has a temporal resolution of one month while the soil moisture datasets have a daily resolution.

The data is stored in NetCDF format, a data format capable of handling multi-dimensional data and used mainly for scientific purposes. To read the files, we are using the netCDF4 Python library. The NetCDF files contain several variables which, for any dataset *dset* can be displayed using the function *dset.variables*. Variables contained in our datasets include the soil moisture (sm), latitude (lat) and longitude (lon) values of the measurement points.

In [2]:

```
# Reading Data

#Datapath
os.chdir('C:/Users/Nicholas Wagener/Desktop/nicolas/test_csv')  # CHANGE HERE

# Here we are reading the two datasets
mean1991_2010 = Dataset("STACKED_ESACCI-L3S-SSMV-COMBINED-CLIM-MEAN-1991-2010-DAILY-fv04.5.
daily2010 = Dataset("STACKED_ESACCI-SOILMOISTURE-L3S-SSMV-COMBINED-2010-DAILY-fv04.5.nc", '

# The quantities we're interested in are read out and stored in separate variables
with np.errstate(invalid='ignore'):
    sm_avg = mean1991_2010['sm'][:]
    sm_2010 = daily2010['sm'][:]

lat = mean1991_2010['lat'][:]
lon = mean1991_2010['lon'][:]
sm_avg = np.delete(sm_avg,59,axis=0)

# It is good practice to close the NetCDF files when they are no longer used
mean1991_2010.close()
daily2010.close()

print(sm_avg.shape)
print(sm_2010.shape)
```

```
(365, 720, 1440)
(365, 720, 1440)
```

We're mainly interested in the soil moisture anomalies before the 2010 fires. The idea is to compare the soil moisture values for the first 3 quarters of 2010 to the long term mean on a daily basis. I.e. for each day in our 2010 dataset we subtract the long term mean for that day.

The fires broke out in late July 2010 and burned until the beginning of September that same year. After the difference of the 2010 values from the historical mean is calculated, we therefore reduce the resulting anomaly dataset to the months June to September.

In [3]:

```
# Calculating soil moisture anomalies for summer 2010

sm_avg = np.moveaxis(sm_avg, 0, -1)
sm_2010 = np.moveaxis(sm_2010, 0, -1)
sm_anomalies = sm_2010 - sm_avg

sm_anomalies = sm_anomalies[:,:,151:271:10] # Reducing number of datapoints to 3 per month

del(sm_avg)
del(sm_2010)
```

We now have the soil moisture anomalies, as calculated in the previous step, and the geographic coordinates each as a separate variable. In order to get a first impression of the soil moisture anomalies leading up to the 2010 wildfires, we will plot these data points in a next step. To do that, we will first store all data in a geopandas dataframe. geopandas is a Python library based on the popular pandas library but with added support for geographic data.

The first step will be to convert the latitude and longitude coordinates, which are currently stored as float numbers, into a geometry object which can be interpreted by geopandas.

In [4]:

```
# Extracting coordinates of observations

shape = sm_anomalies.shape[:2]
coordinates = np.zeros((shape+(2,)))

# Creating the list of acquisition dates
datelist = [d.strftime('%Y-%m-%d')
            for d in pd.date_range('2010-06-01', periods=sm_anomalies.shape[2], freq='10D')

for i in range(lat.shape[0]):
    for j in range(lon.shape[0]):
        coordinates[i,j]=[lat[i],lon[j]]

sm_anomalies = sm_anomalies.reshape(-1, sm_anomalies.shape[-1])
coordinates = coordinates.reshape(-1,coordinates.shape[-1])

# Further reducing data volume using a bounding box
bounding = (coordinates[:,0]>=40) & (coordinates[:,0]<=82)
sm_anomalies = sm_anomalies[bounding]
coordinates = coordinates[bounding]

# Converting float number coordinates to geometry
geometry = [Point(xy) for xy in zip(coordinates[:,1], coordinates[:,0])]
```

Now the gopandas dataframe is created using the geometry and the soil moisture anomalies calculated previously. We're using the WGS84 geographic coordinate system.

In [5]:

```
# Creating Geopandas Dataframes

# Coordinate reference system : WGS84
crs = {'init': 'epsg:4326'}

sm_anomalies_gdf = gpd.GeoDataFrame(data=sm_anomalies, columns=datelist, crs=crs, geometry=
sm_anomalies_gdf['lat'] = coordinates[:,0]
sm_anomalies_gdf['lon'] = coordinates[:,1]

del(sm_anomalies)
del(coordinates)
```

Our dataset contains soil moisture measurements covering the whole earth. Since we're interested in Russia only, we're going to extract the measurement points lying within Russia. naturalearth_lowres is a geopandas dataset containing the contours of countries which we will use to obtain the approximate country boundaries. In the next step, a mask is created masking out all values outside Russia. **This may take a while to run**.

In [6]:

```
# Keep only datapoints in Russia

world = gpd.read_file(gpd.datasets.get_path('naturalearth_lowres'))
russia =  world[world.name == "Russia"]
del(world)

russia_mask = sm_anomalies_gdf['geometry'].within(russia['geometry'].iloc[0])
sm_russia_anom = sm_anomalies_gdf.loc[russia_mask]
```

Now we wil transform the geodataframe so that all soil moisture anomalies are in one column:

In [7]:

```
sm_russia_anom = pd.melt(sm_russia_anom,
                         id_vars=['lat','lon','geometry'],
                         var_name='date',
                         value_name='sm')

sm_russia_anom = sm_russia_anom.dropna(subset=['sm'])
```

In the next step, we will plot the soil moisture anomalies for Russia in 2010 before and during the wildfires. Run the cell below and use the slider to change the date. The anomalies are displayed in %/100 (e.g. 0.2 means 20% above long-term average).

In [14]:

```
colorscale=[[0.0, "rgb(165,0,38)"],
                [0.0666666666666666, "rgb(215,48,39)"],
                [0.1111111111111111, "rgb(244,109,67)"],
                [0.1777777777777777, "rgb(253,174,97)"],
                [0.2222222222222222, "rgb(254,224,144)"],
                [0.3333333333333333, "rgb(224,243,248)"],
                [0.5555555555555555, "rgb(171,217,233)"],
                [0.7777777777777778, "rgb(116,173,209)"],
                [0.8888888888888888, "rgb(69,117,180)"],
                [1.0, "rgb(49,54,149)"]]

fig = px.scatter_mapbox(sm_russia_anom,
                        lat="lat",
                        lon="lon",
                        color='sm',
                        color_continuous_scale=colorscale,
                        zoom=1.5,
                        animation_frame='date',
                        title='Soil moisture anomalies 2010')
fig.update_layout(mapbox_style="open-street-map")
fig.show()
```

Soil moisture anomalies 2010

Keep in mind that the first fires in Western Russia broke out on July 29th 2010. As we can see, negative soil moisture anomalies started appearing before the first fires broke out.

Now let's load the burned area dataset. This dataset contains information on burned areas in August 2010. Again, we extract the burned area, latitude and longitude each into separate variables. Then we largely repeat the steps we applied to the soil moisture datasets.

In [9]:

```
fire = Dataset("burned_area/20100801-ESACCI-L4_FIRE-BA-MODIS-fv5.1.nc", "r", format="NETCDF

with np.errstate(invalid='ignore'):
    burned_area = fire['burned_area'][:] #total burned area in m2

lat2 = fire['lat'][:]
lon2 = fire['lon'][:]

burned_area = np.moveaxis(burned_area, 0, -1)
burned_area = burned_area.flatten()

fire.close()
```

In [10]:

```
coordinates2 = np.zeros((shape+(2,)))

for i in range(lat.shape[0]):
    for j in range(lon.shape[0]):
        coordinates2[i,j]=[lat2[i],lon2[j]]

coordinates2 = coordinates2.reshape(-1,coordinates2.shape[-1])

bounding2 = (coordinates2[:,0]>=40) & (coordinates2[:,0]<=82)
burned_area = burned_area[bounding2]
coordinates2 = coordinates2[bounding2]

# Converting float number coordinates to geometry
geometry2 = [Point(xy) for xy in zip(coordinates2[:,1], coordinates2[:,0])]
```

In [11]:

```
fire_gdf = gpd.GeoDataFrame({'total': burned_area,
                             'lat':coordinates2[:,0],
                             'lon':coordinates2[:,1]},
                            columns=['total','lat','lon'],
                            crs=crs,
                            geometry=geometry2)

# masking out data outside of Russia
russia_mask2 = fire_gdf.within(russia['geometry'].iloc[0])
russia_fire = fire_gdf.loc[russia_mask2]

del(burned_area,fire_gdf,geometry2,coordinates2)
```

Now we will plot the areas that were burned in August 2010. This is to get a first visual impression of where the 2010 fires were concentrated. Run the cell below to see the map.

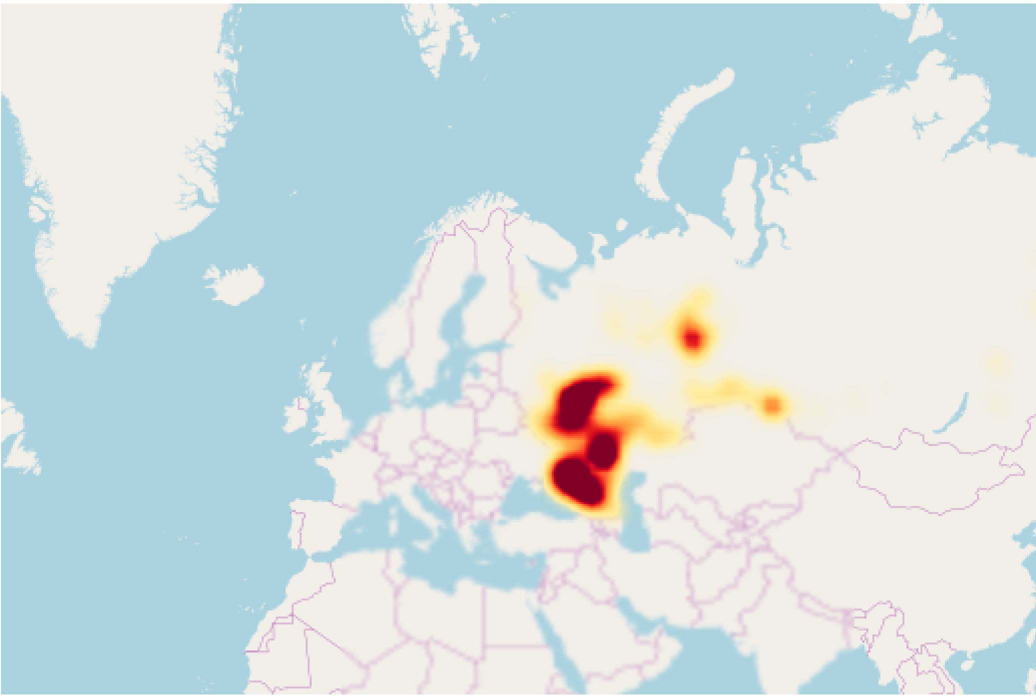
In [12]:

```
burned = russia_fire[russia_fire['total']!=0]

fig = go.Figure(go.Densitymapbox(lat=burned['lat'],
                                 lon=burned['lon'],
                                 z=burned['total'],
                                 radius=12,
                                 colorbar=dict(title='Burned area [m²]'),
                                 colorscale='YlOrRd'))

fig.update_layout(mapbox_style="open-street-map",
                  mapbox_center_lon=90,
                  mapbox_center_lat=60,
                  title='Total burned area [m²] August 2010')
fig.show()
```

Total burned area [m^2] August 2010



From a first visual inspection it seems, that the occurence and intensity of the 2010 wildfires correlates with the areas that demonstrated high soil moisture anomalies before the event. In order to probe this theory, let's look at the soil moisture anomalies on July 21st 2010, just before the first fires broke out. We will categorize our data into bins of different soil moisture anomaly thresholds and check for each bin, in how many cases (%) fires broke out and also, how much area was burned on average. To do this, we will first join the soil moisture dataset for the 21st of July 2010 with the burned area dataset for August 2010.

In [13]:

```
# Soil moisture anomaly, % of fire, total area mean

# Soil moisture and burned area dataframes are merged.
join = pd.merge(sm_russia_anom.loc[sm_russia_anom['date'] == '2010-07-21'],russia_fire,on=[
# List of threshold values to use: -10%, -8%, -5%, 0%, +5%
thresholds = [-0.1,-0.08,-0.05,0,0.05]

output = np.zeros((len(thresholds)+1,2))
iteration = 0

for i in range(len(thresholds)+1):
    if iteration == 0:
        dat = join.loc[(join['sm'] < thresholds[i])]
    elif iteration == len(thresholds):
        dat = join.loc[(join['sm'] >= thresholds[i-1])]
    else:
        dat = join.loc[(join['sm'] >= thresholds[i-1]) & (join['sm'] < thresholds[i])]
    if dat.shape[0] != 0:
        output[i,0] = dat[dat['total'] != 0].shape[0]/dat.shape[0]
    else:
        output[i,0] = dat.shape[0]
    output[i,1] = dat.loc[dat['total'] != 0]['total'].mean()
    iteration +=1

print(output)
```

```
[[4.26229508e-01 7.03626350e+06]
 [4.49645390e-01 9.41366300e+06]
 [2.09174312e-01 8.09124250e+06]
 [6.16410465e-02 1.54960610e+07]
 [3.71253914e-02 1.92169060e+07]
 [1.92644483e-02 2.33221780e+07]]
```

The table below shows the results of merging the two datasets. For each category of soil moisture anomaly in July 2010, the likeliness of a fire breaking out the following month and the average size of the area burned was calculated.

Soil moisture anomaly	Likeliness of fire [%]	Mean burned area [m²]
<-10%	42.62	7,036,263
>=-10%<-8%	44.96	9,413,663
>=-8%<-5%	20.92	8,091,242
>=-5%<0%	6.16	15,496,061
>=0%<+5%	3.71	19,216,906
>=+5%	1.93	23,322,178

Conclusions

It could be shown that the large-scale wildfires in Western Russia in 2010 developed under unusually dry soil conditions. These conditions could be observed already days before the first fires broke out. We also saw that the chances of a fire breaking out are highest under unusually dry conditions. In over 40% of all areas with soil moisture anomalies < -8%, fires broke out. But only in 1.9% of areas with soil moisture anomalies of > +5% (i.e.

unusually wet conditions) fires broke out. However, fires that broke out under such more 'moist' conditions, are likely to be larger in size. The average burned area in areas with soil moisture anomalies of >+5% is more than 23 million m^2 (23 Km^2) while in areas with soil moisture anomalies below -10% (i.e. unusually dry conditions) the average burned area is *only* around 7 million m^2. This indicates, that the size of fires is influenced by other factors (e.g. availability of burnable material, anthropogenic factors, etc.). The fact that many fires also broke out in areas with low negative or even positive soil moisture anomalies shows that soil moisture cannot be the only variable in a fire prediction system but only one of many.

References

1. Bartsch, Annett & Balzter, Heiko & George, Charles. (2009). The influence of regional surface soil moisture anomalies on forest fires in Siberia observed from satellites. Environmental Research Letters. 4. 045021. 10.1088/1748-9326/4/4/045021.
2. Chuvieco, E.; Pettinari, M.L.; Lizundia Loiola, J.; Storm, T.; Padilla Parellada, M. (2019): ESA Fire Climate Change Initiative (Fire_cci): MODIS Fire_cci Burned Area Grid product, version 5.1. Centre for Environmental Data Analysis, 08 February 2019. doi:10.5285/3628cb2fdba443588155e15dee8e5352. [http://dx.doi.org/10.5285/3628cb2fdba443588155e15dee8e5352 (http://dx.doi.org/10.5285/3628cb2fdba443588155e15dee8e5352)](http://dx.doi.org/10.5285/3628cb2fdba443588155e15dee8e5352)
3. Dorigo, W.A., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., Chung, D., Ertl, M., Forkel, M., Gruber, A., Haas, E., Hamer, D. P. Hirschi, M., Ikonen, J., De Jeu, R. Kidd, R. Lahoz, W., Liu, Y.Y., Miralles, D., Lecomte, P. (2017). ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions. In Remote Sensing of Environment, 2017, ISSN 0034-4257, [https://doi.org/10.1016/j.rse.2017.07.001 (https://doi.org/10.1016/j.rse.2017.07.001)](https://doi.org/10.1016/j.rse.2017.07.001)
4. Gruber, A., Dorigo, W. A., Crow, W., Wagner W. (2017). Triple Collocation-Based Merging of Satellite Soil Moisture Retrievals. IEEE Transactions on Geoscience and Remote Sensing. PP. 1-13. 10.1109/TGRS.2017.2734070
5. Liu, Y.Y., Dorigo, W.A., Parinussa, R.M., de Jeu, R.A.M. , Wagner, W., McCabe, M.F., Evans, J.P., van Dijk, A.I.J.M. (2012). Trend-preserving blending of passive and active microwave soil moisture retrievals, Remote Sensing of Environment, 123, 280-297, doi: 10.1016/j.rse.2012.03.014