In [1]:
```
import os
print(os.environ['SSS_TAG_DIR'])
import iris
print(iris.__version__)
import numpy as np

import warnings
warnings.filterwarnings('ignore')
```

```
/net/project/ukmo/scitools/opt_scitools/conda/deployments/default-2022_11
_22
3.3.1
```

In [2]:
```
# Fix original climatology
def reformat(orig, field, filename):

    orig.rename("Original climatology")

    if orig.attributes['invalid_units'] == 'm^3m^-3':
        orig.units = 'm3 m-3'
        del orig.attributes['invalid_units']

    # Roll 180 degrees and flip N-S
    nlon = len(orig.coord('longitude').points)
    nlat = len(orig.coord('latitude').points)
    orig.data = np.roll(orig.data, nlon // 2)
    orig.coord('longitude').points = orig.coord('longitude').points + 180

    orig.data = np.flip(orig.data, 0)
    orig.coord('latitude').points = np.flip(orig.coord('latitude').points
```

In [3]:
```
# New climatology pre-processed by ESMValTool
dname = '/scratch/hadtq/esmvaltool_output/recipe_autoassess_landsurface_s
fname = 'OBS_ESACCI-SOILMOISTURE_sat_L3S-SSMV-COMBINED-v4.2_Lmon_sm_1999-
import os.path
new_all_seasons = iris.load_cube(os.path.join(dname,fname))
new_all_seasons.rename("New climatology")
```

In [4]:
```
import matplotlib.pyplot as plt
%matplotlib inline
import iris.quickplot as qplt
import matplotlib.ticker as mticker
import matplotlib.cm as mpl_cm

vmin = 0
vmax = 0.6
levels = mticker.MaxNLocator(nbins=21).tick_values(vmin, vmax)

def main_plots(new, orig):
    # orig
    qplt.contourf(orig, levels=levels)
    plt.gca().coastlines()
    plt.show()

    # new
    qplt.contourf(new, levels=levels)
    plt.gca().coastlines()
    plt.show()
```

In [5]:
```
# Absolute difference
def absolute_diff(new, orig):
    diff = new.copy()
    diff.rename('Absolute difference: new vs orig')
    diff.data = new.data - orig.data

    import iris.analysis.cartography
    grid_areas = iris.analysis.cartography.area_weights(diff)
    print('Mean diff:', diff.collapsed(['longitude', 'latitude'], iris.an
    print('Median absolute diff:', np.ma.median(np.ma.absolute(diff.data)
    print('Standard deviation of diff:', diff.collapsed(['longitude', 'la
    print('Min diff:', diff.data.min())
    print('Max diff:', diff.data.max())

    print('Standard deviation of orig:', orig.collapsed(['longitude', 'la
    print('Standard deviation of new:', new.collapsed(['longitude', 'lati

    cmap = mpl_cm.get_cmap("brewer_RdBu_11")

    vmin = -0.06
    vmax = 0.06
    levels = mticker.MaxNLocator(nbins=14, symmetric=True).tick_values(vm
    qplt.contourf(diff, levels=levels, cmap=cmap)
    plt.gca().coastlines()
    plt.show()
```

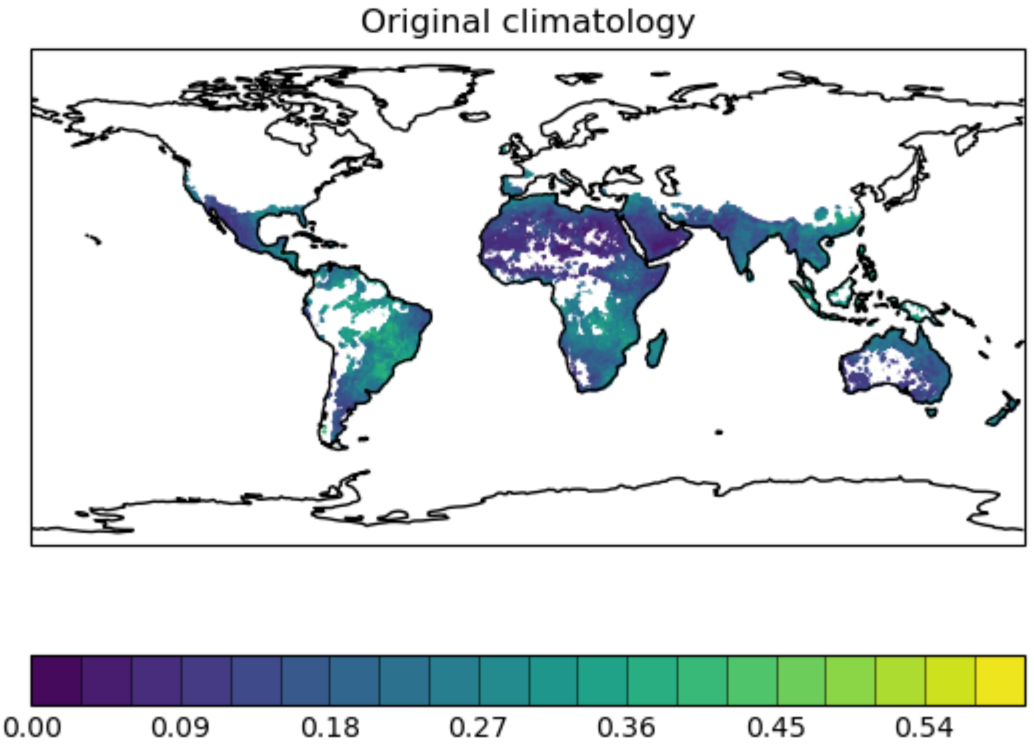
In [6]:
```
# Relative difference
def relative_diff(new, orig):
    rdiff = new.copy()
    rdiff.rename('Relative difference: new vs orig')
    rdiff.data = 2 * (new.data - orig.data) / (new.data + orig.data)

    print('Median absolute of the relative diffs:', np.ma.median(np.ma.ab

    cmap = mpl_cm.get_cmap("brewer_RdBu_11")

    vmin = -0.6
    vmax = 0.6
    levels = mticker.MaxNLocator(nbins=14, symmetric=True).tick_values(vm
    qplt.contourf(rdiff, levels=levels, cmap=cmap)
    plt.gca().coastlines()
    plt.show()
```

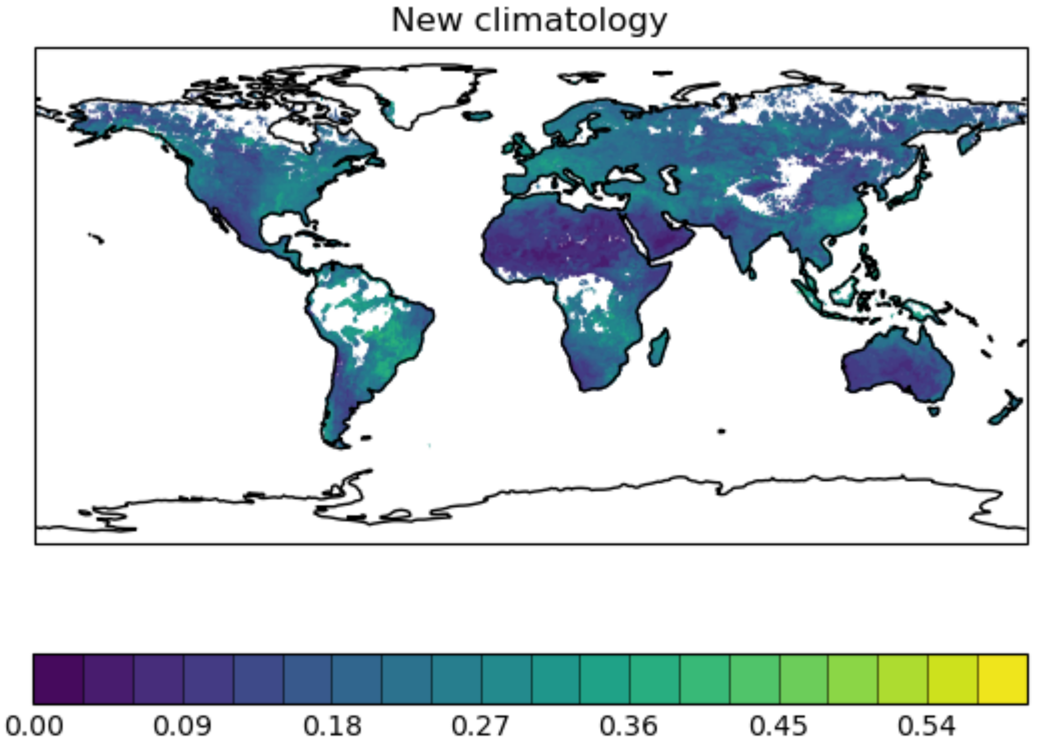
In [7]:
```
# Pattern correlation
def pattern_correlation(new, orig):
    corrcoef = np.ma.corrcoef(orig.data.flatten(), new.data.flatten())[0,
    print('Pattern correlation: r =', corrcoef)

    plt.hexbin(orig.data.flatten(), new.data.flatten())
    plt.title('2D histogram')
    plt.xlabel('Original')
    plt.ylabel('New')
    plt.xlim([0,0.5])
    plt.ylim([0,0.5])
    plt.show()
```

In [8]:
```
seasons = ['djf', 'mam', 'jja', 'son']
```

In [9]:
```
from IPython.display import Markdown, display

for index, season in enumerate(seasons):
    display(Markdown(f"# Season {season}"))

    orig = iris.load_cube(f'/project/cma/clim/ecv_soil_moisture/ecv_soil_
                          callback=reformat)
    new = new_all_seasons.extract(iris.Constraint(season_number=0))

    main_plots(new, orig)

    display(Markdown("### Statistics and absolute difference plot"))
    absolute_diff(new, orig)

    display(Markdown("### Relative difference"))
    relative_diff(new, orig)

    display(Markdown("### Pattern correlation"))
    pattern_correlation(new, orig)
```

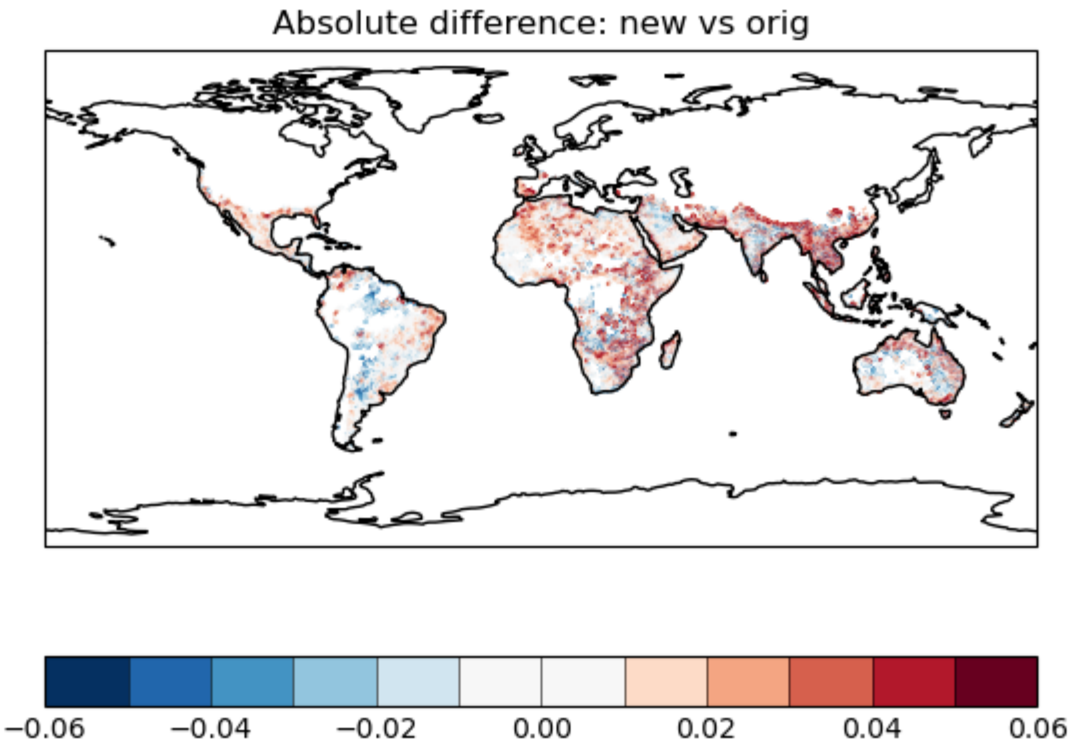
Season djf





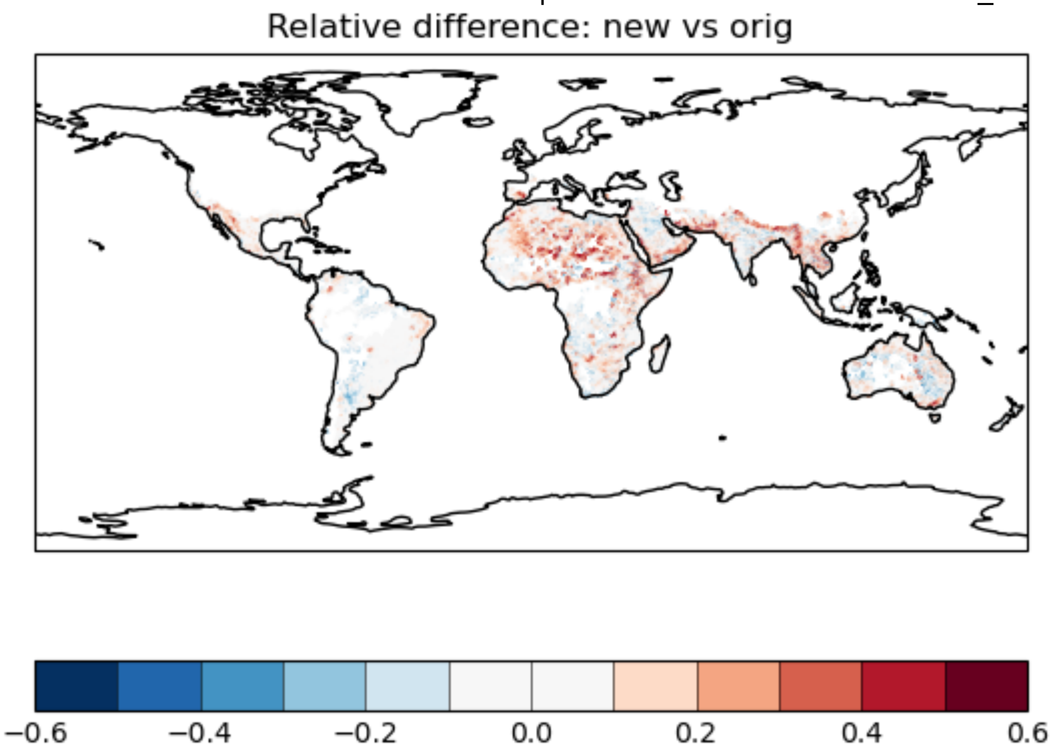
Statistics and absolute difference plot

```
Mean diff: -0.003729965695949966
Median absolute diff: 0.013476461
Standard deviation of diff: 0.026633840159416687
Min diff: -0.18191825
Max diff: 0.2688886
Standard deviation of orig: 0.08146808551925142
Standard deviation of new: 0.06890152488212731
```



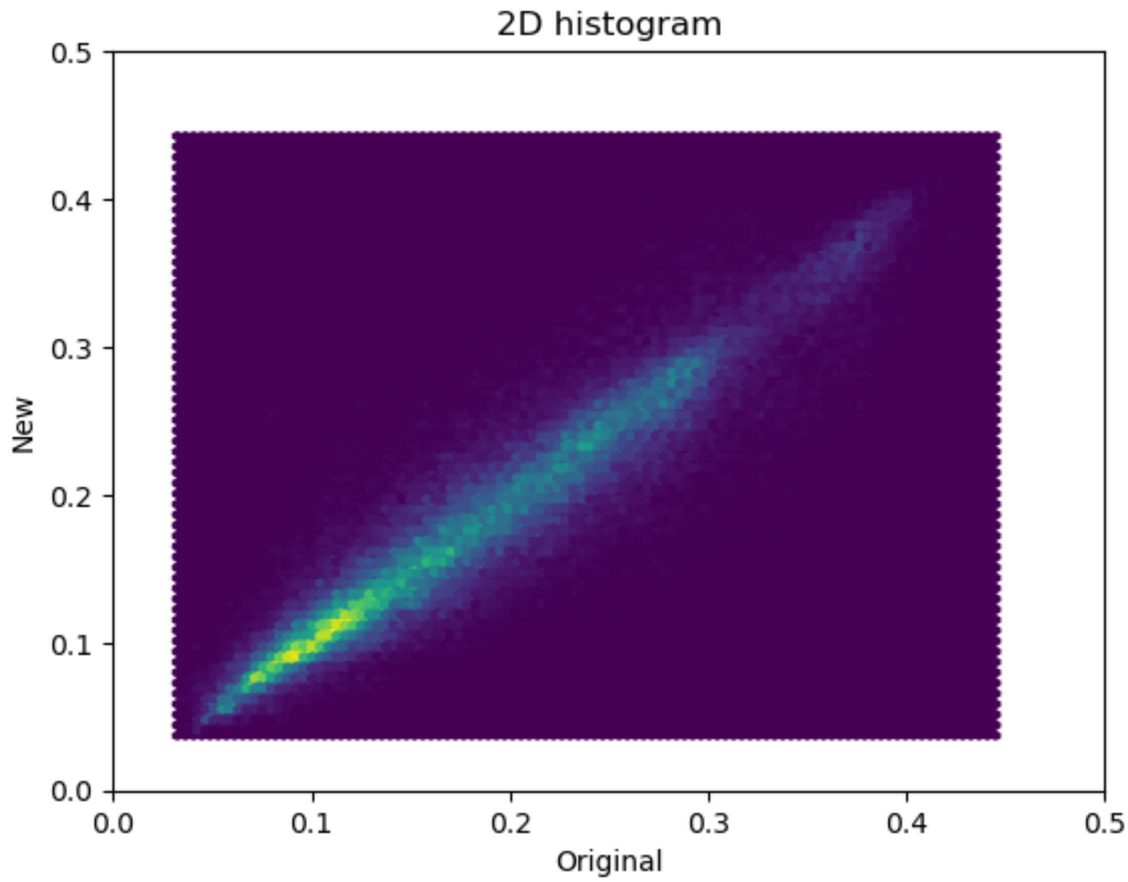
Relative difference

```
Median absolute of the relative diffs: 0.08003883
```



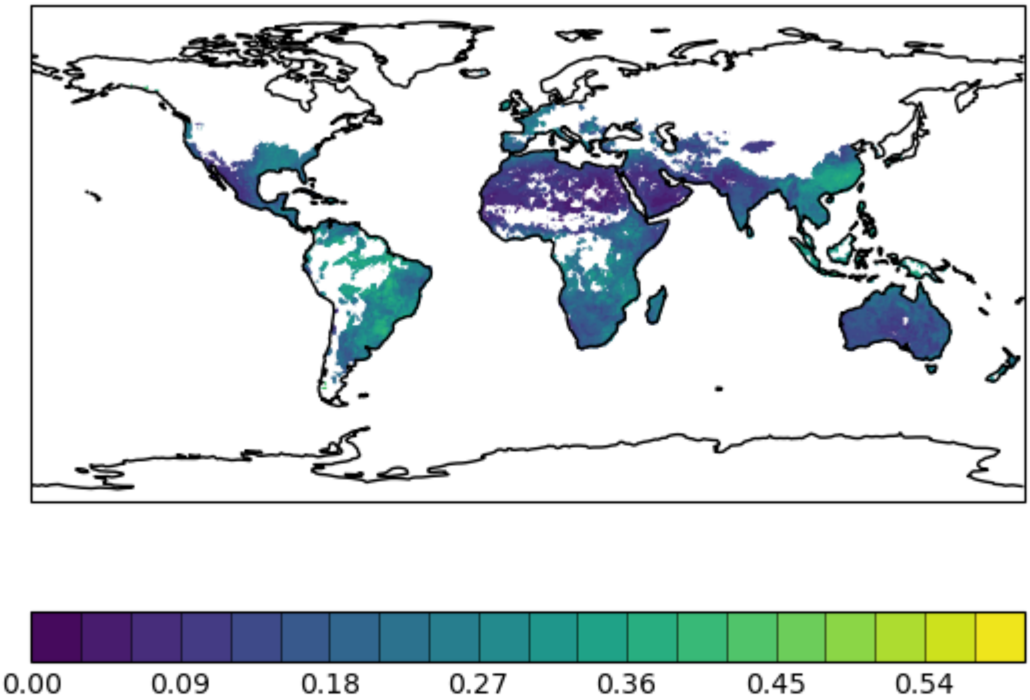
Pattern correlation

```
Pattern correlation: r = 0.9443005466427945
```

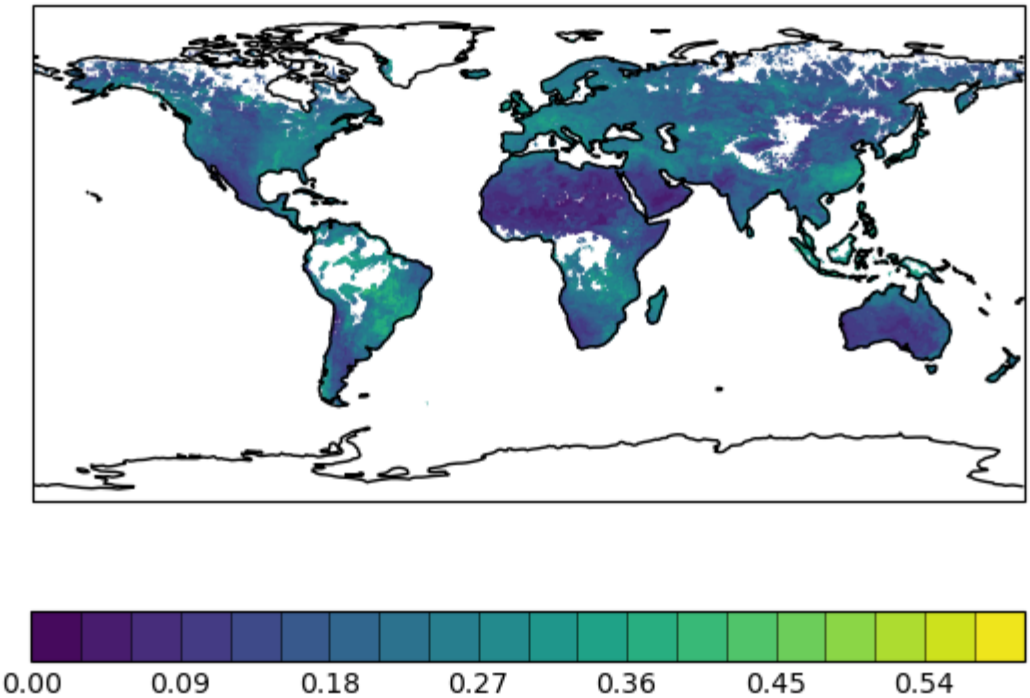


Season mam

Original climatology

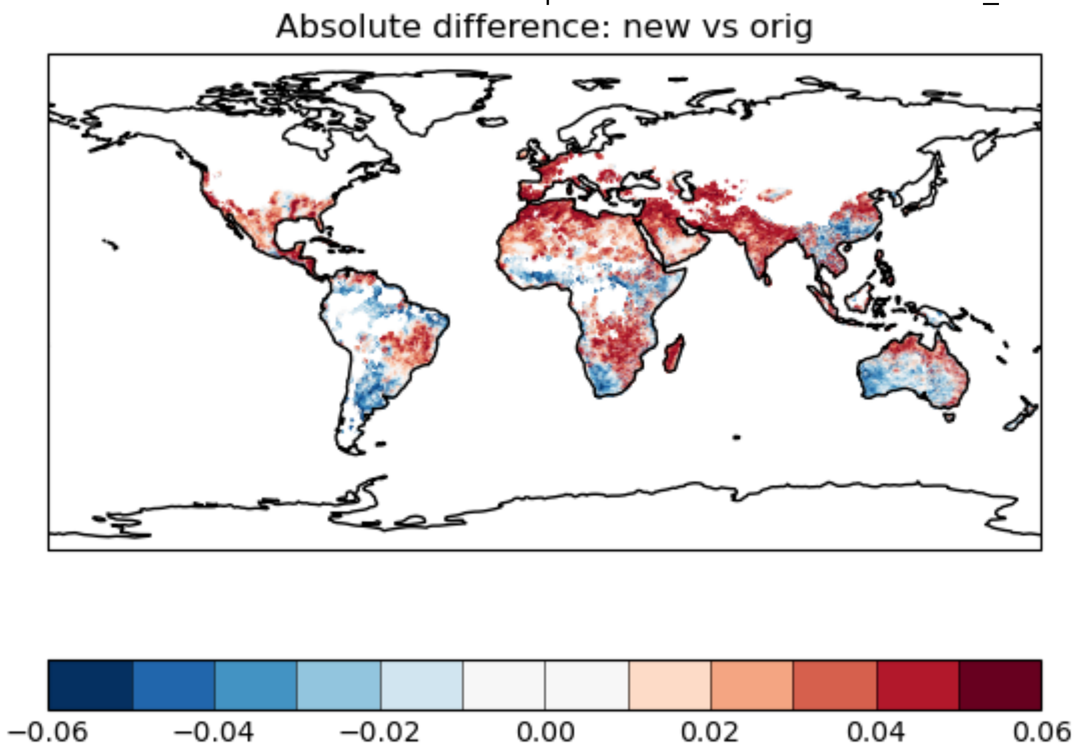


New climatology



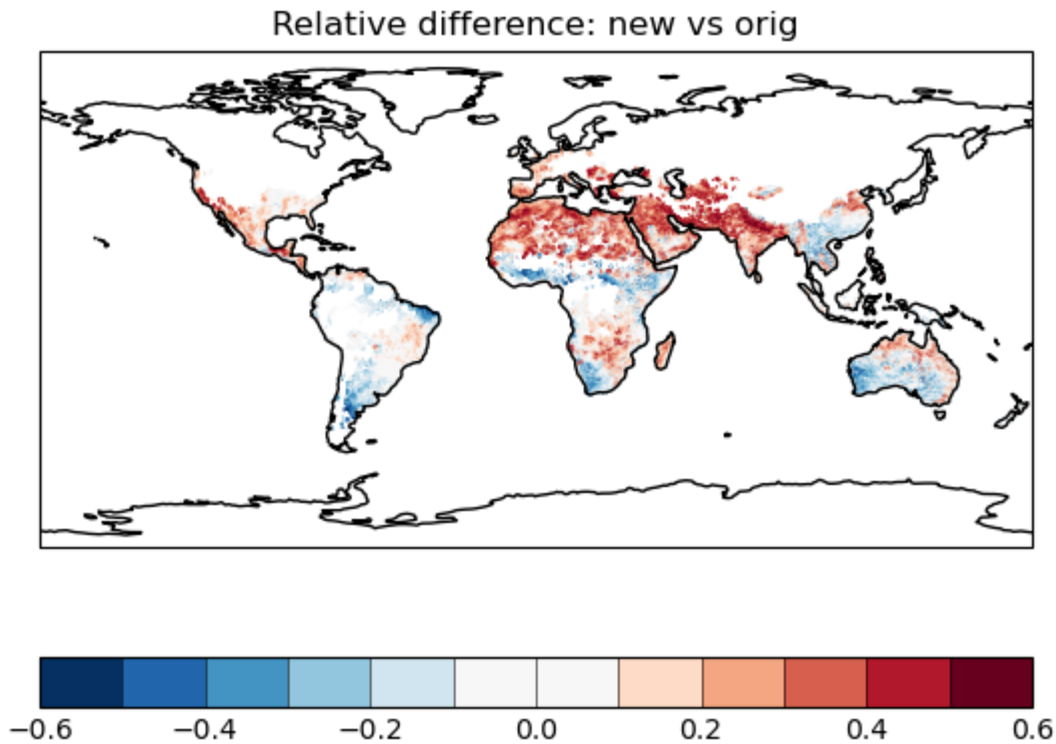
Statistics and absolute difference plot

```
Mean diff: 0.0020726852053120234
Median absolute diff: 0.026901260018348694
Standard deviation of diff: 0.041680076020456515
Min diff: -0.23808417
Max diff: 0.27176502
Standard deviation of orig: 0.08151050866879034
Standard deviation of new: 0.06890152488212731
```



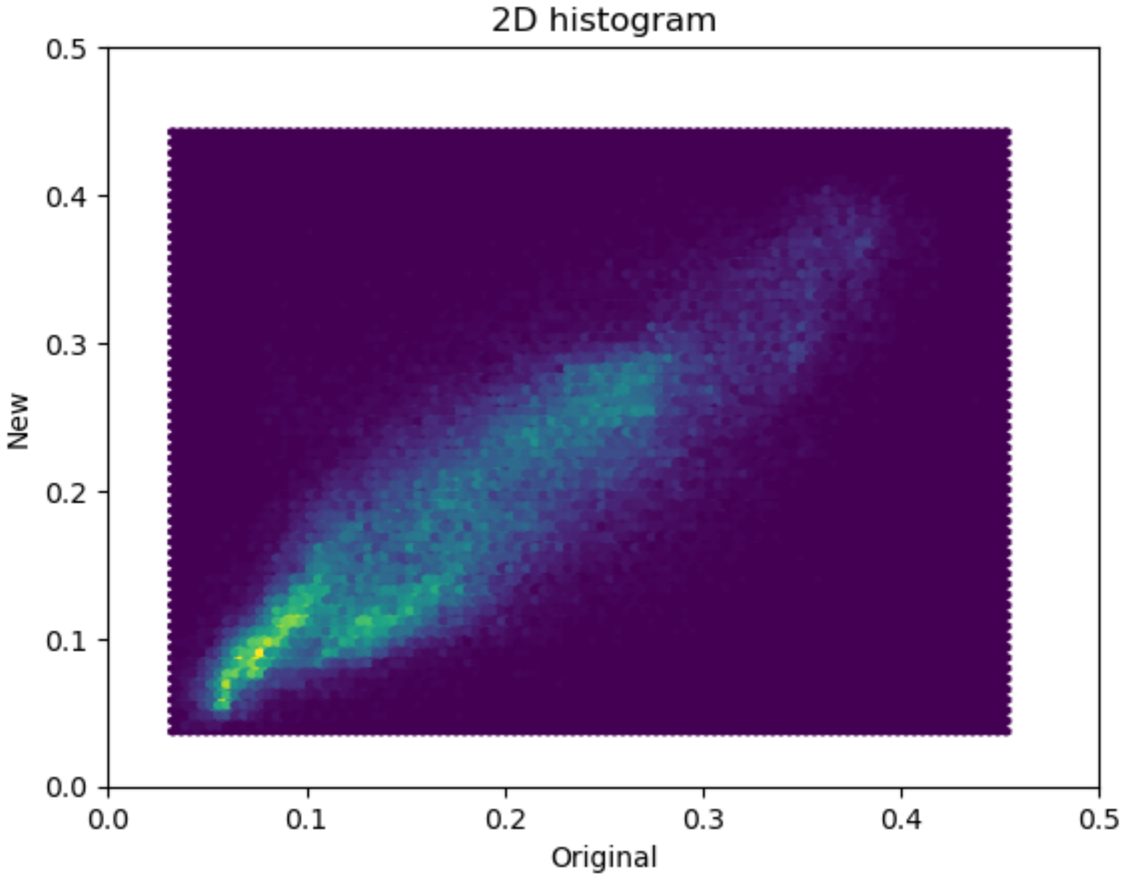
Relative difference

```
Median absolute of the relative diffs: 0.1609104573726654
```

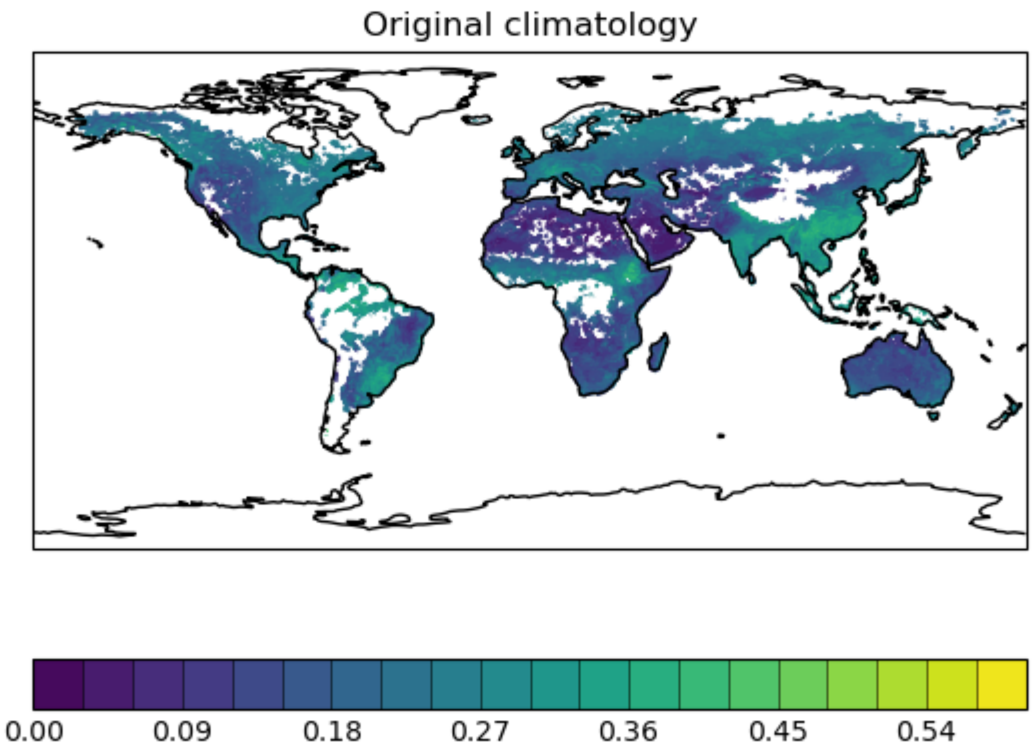


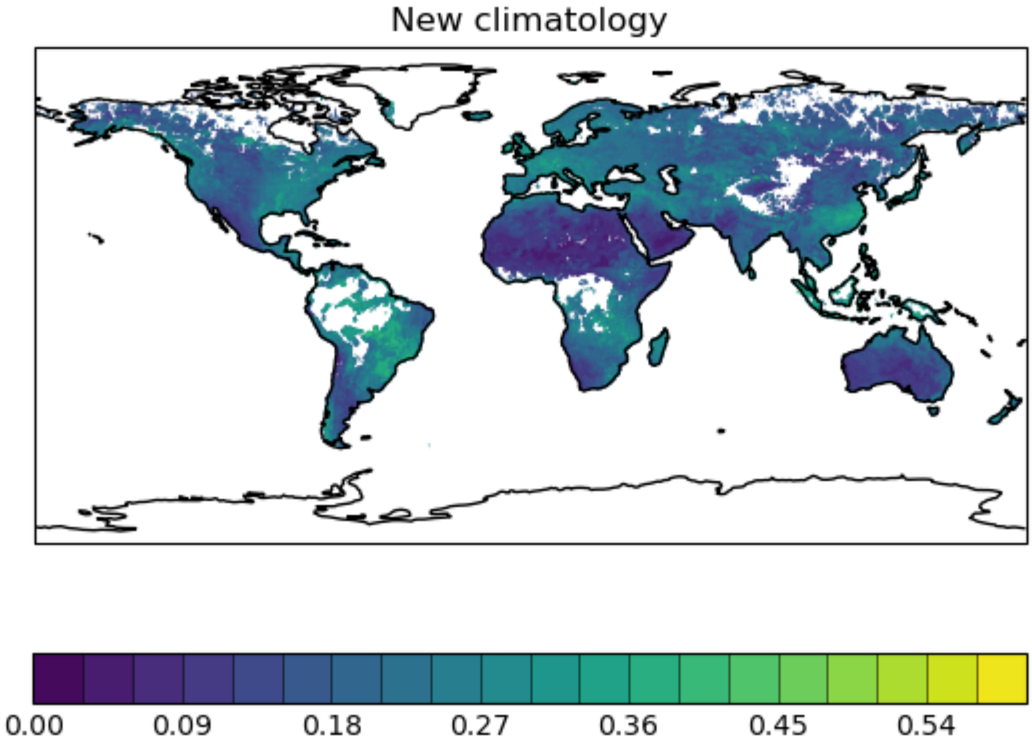
Pattern correlation

```
Pattern correlation: r = 0.8627160613908036
```


2D histogram
New
Original
0.0
0.1
0.2
0.3
0.4
0.5

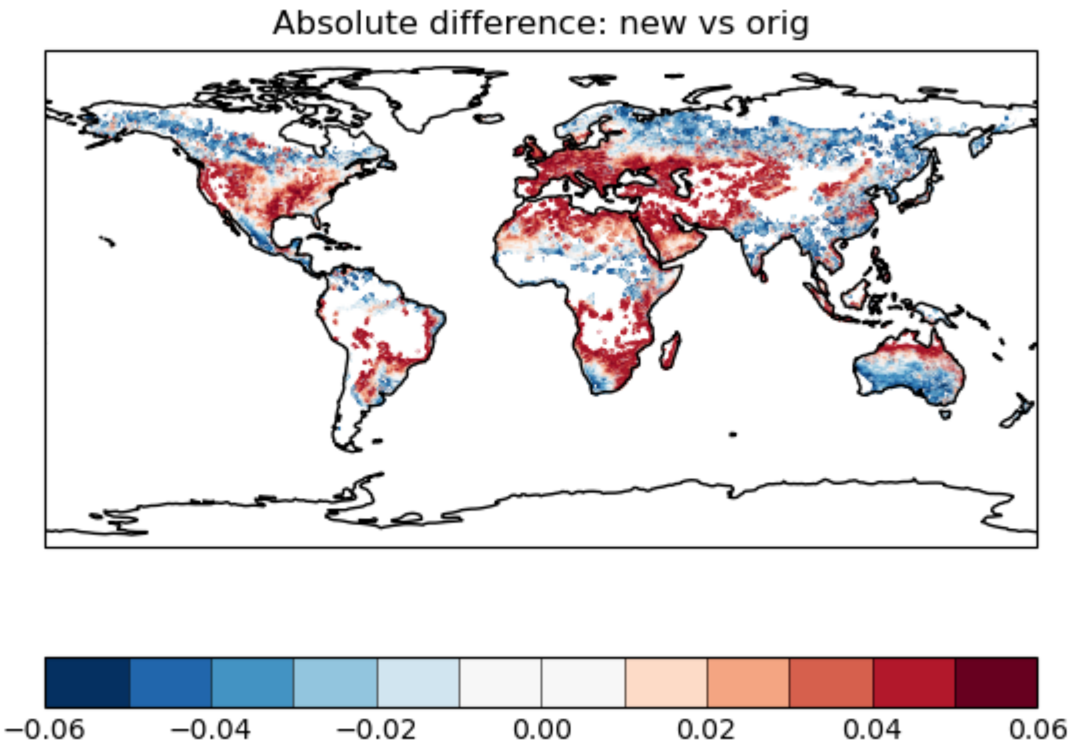
Season jja


Original climatology
0.00
0.09
0.18
0.27
0.36
0.45
0.54



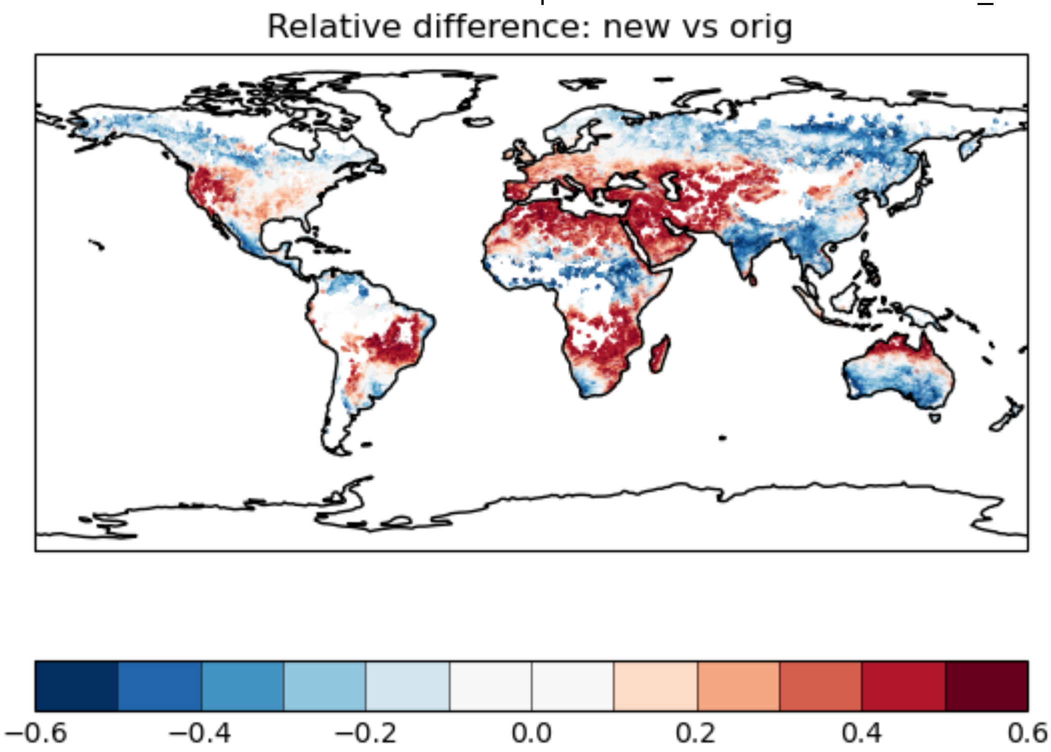
Statistics and absolute difference plot

```
Mean diff: -0.002945320519262027
Median absolute diff: 0.04604482
Standard deviation of diff: 0.07214059211623246
Min diff: -0.3155495
Max diff: 0.29487613
Standard deviation of orig: 0.0743900420058299
Standard deviation of new: 0.06890152488212731
```



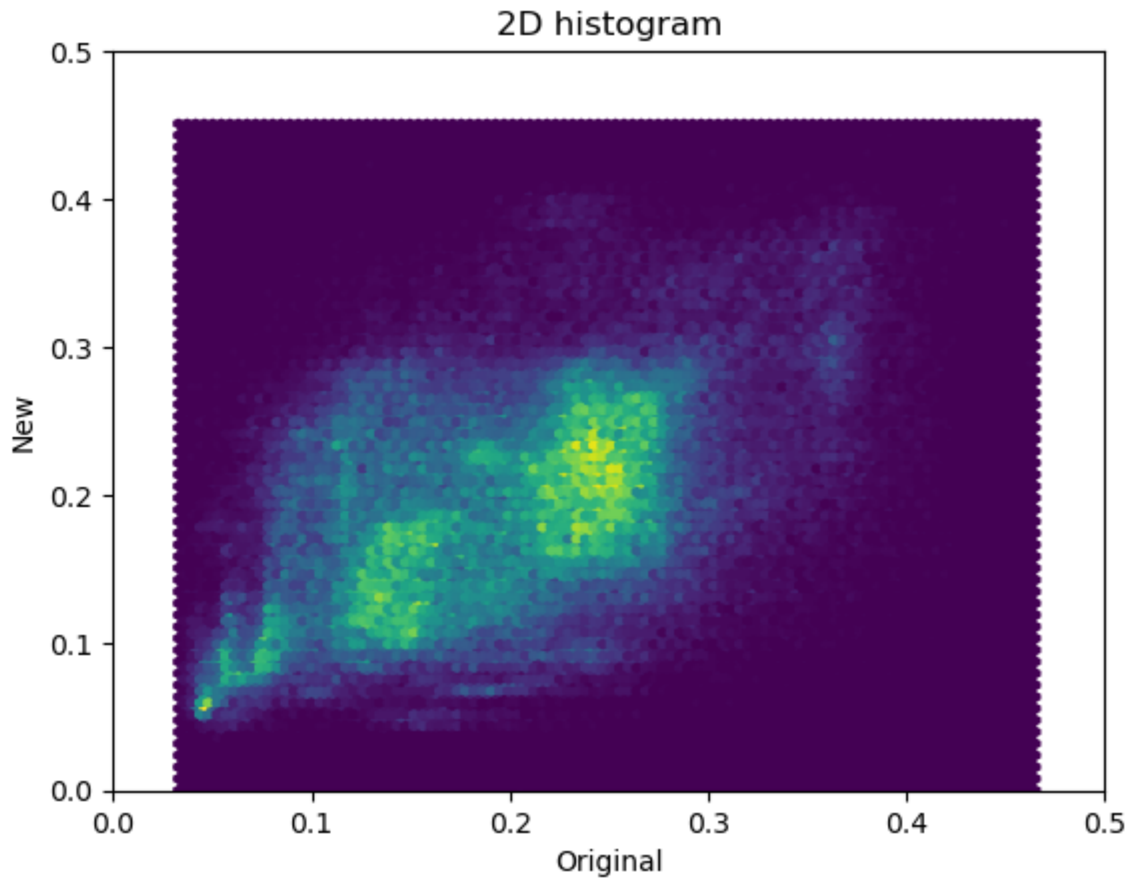
Relative difference

```
Median absolute of the relative diffs: 0.2530849
```

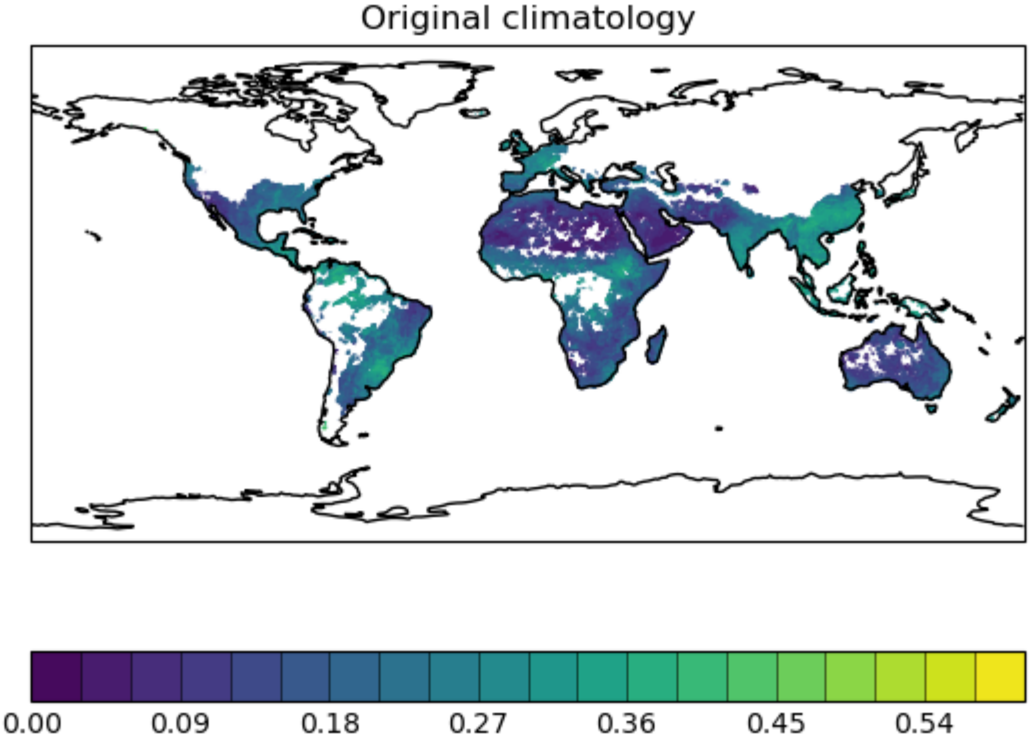


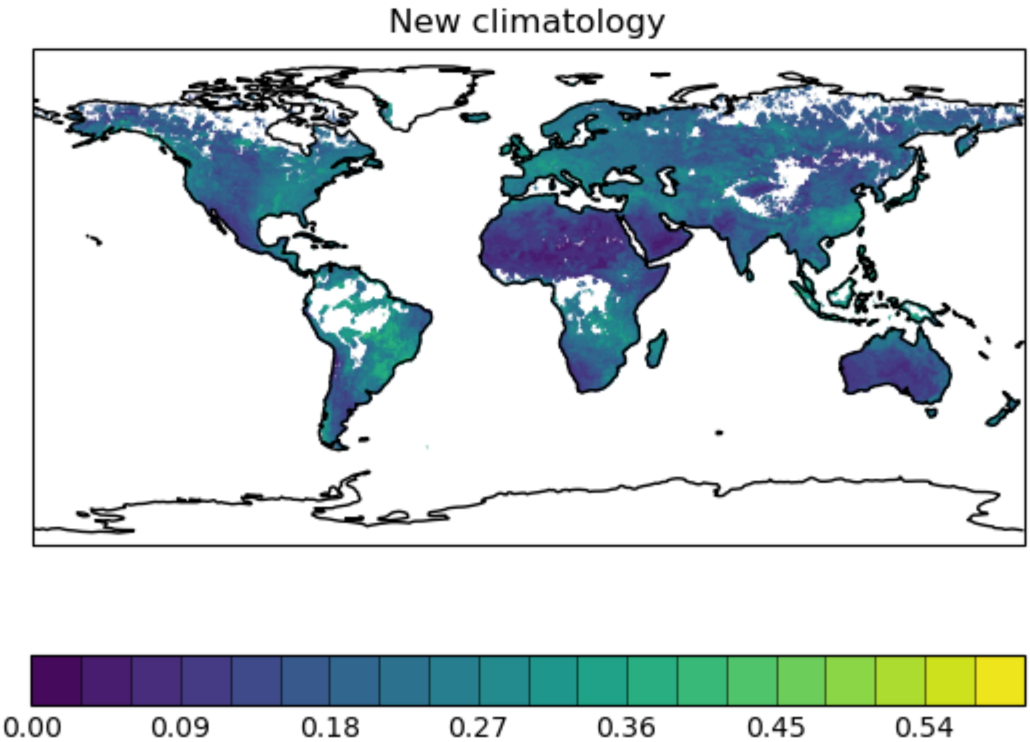
Pattern correlation

```
Pattern correlation: r = 0.5043109961430711
```



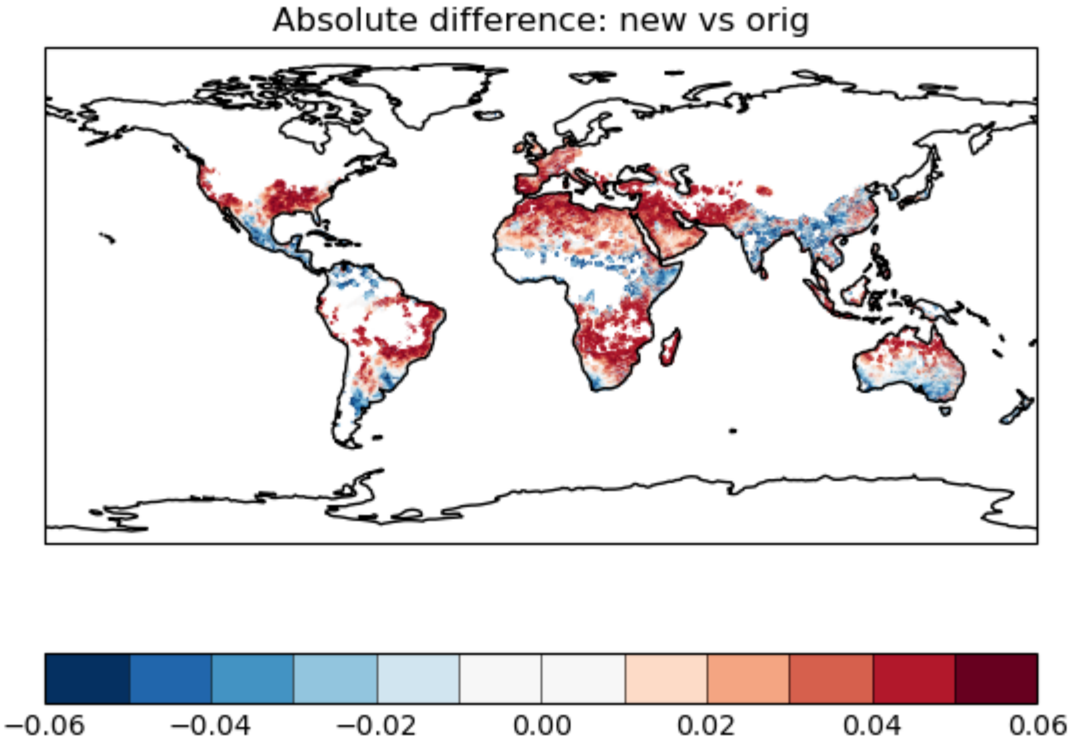
Season son





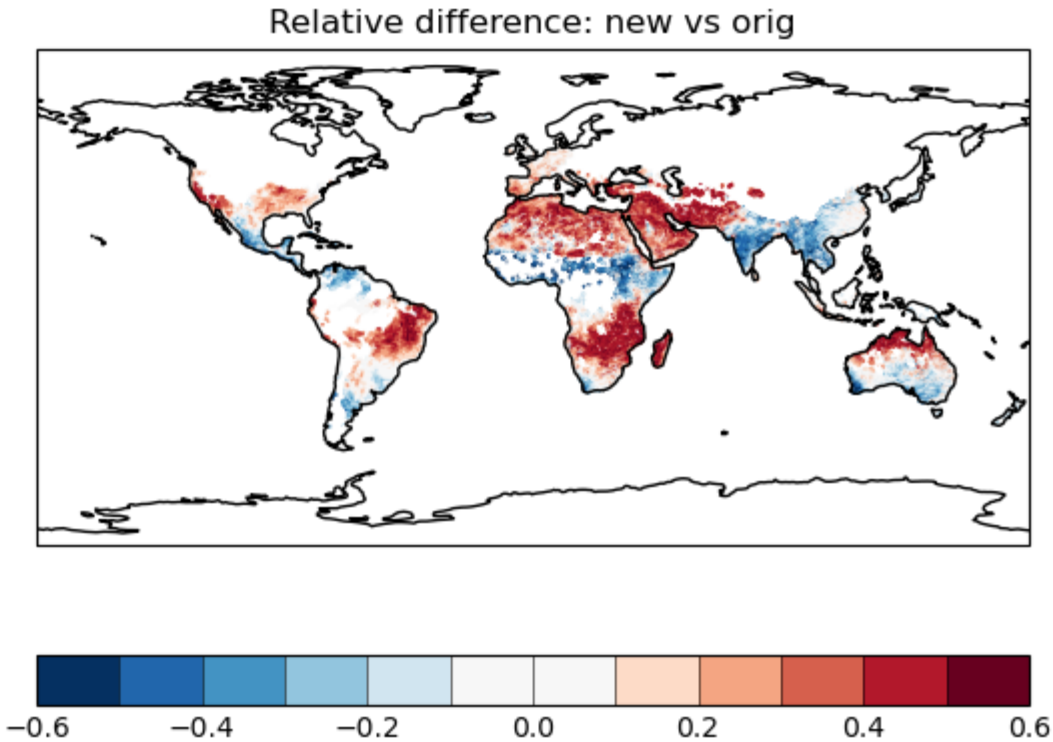
Statistics and absolute difference plot

```
Mean diff: 0.002089745758122653
Median absolute diff: 0.040813446
Standard deviation of diff: 0.0671858411866439
Min diff: -0.27769002
Max diff: 0.26776916
Standard deviation of orig: 0.08323328903613864
Standard deviation of new: 0.06890152488212731
```



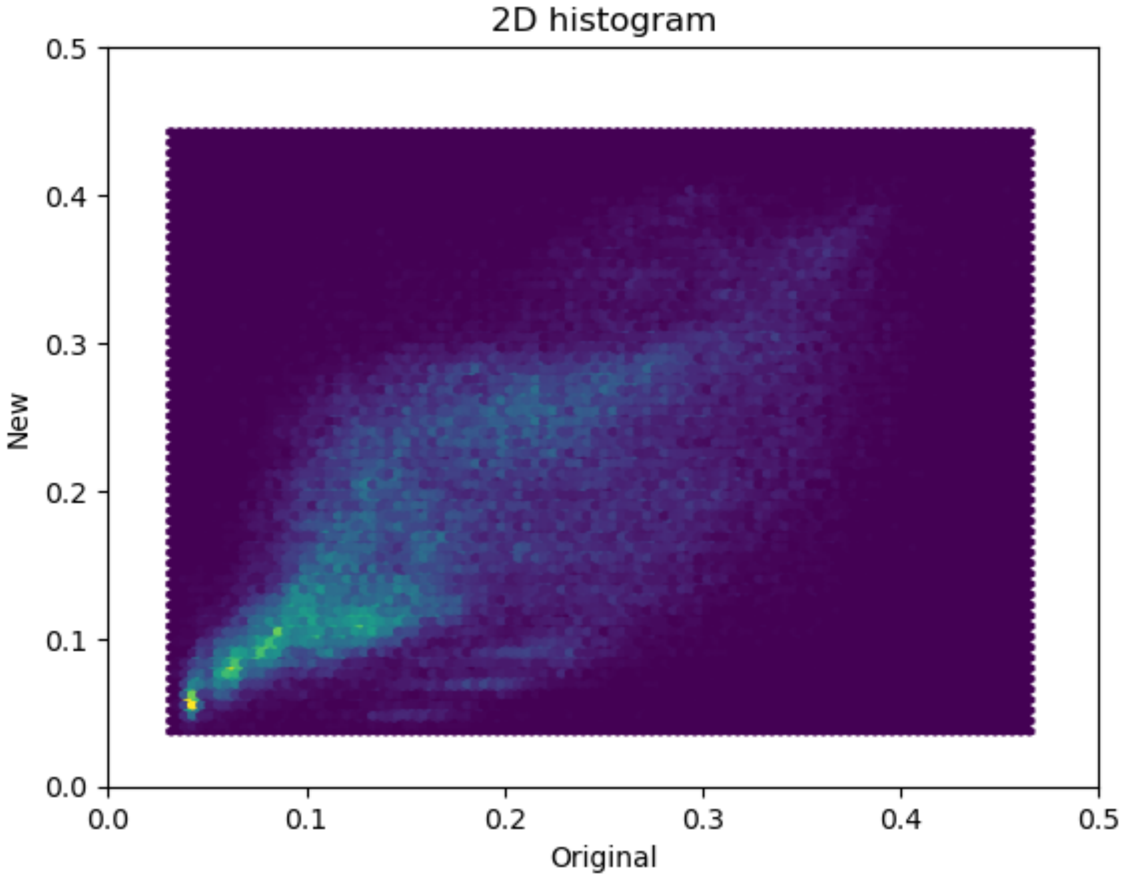
Relative difference

```
Median absolute of the relative diffs: 0.24918813
```



Pattern correlation

```
Pattern correlation: r = 0.6560917796458514
```



Conclusion

At points where they both have data, typical differences are small: the median absolute difference is more than an order of magnitude smaller than the typical values in the dataset. Also an order of magnitude smaller than typical model-observation differences and therefore will not have a large effect on the metrics produced by this recipe, which are computed as a median-absolute difference between model and observation.

The area-weighted mean difference between datasets is an order of magnitude smaller again, indicating no systematic bias between the old and new. Pattern correlation is high for DJF and MAM (>0.85); it is lower for SON (0.65) and lowest in JJA (0.5). This suggests that the original climatology's higher coverage in JJA of N high latitudes (relative to other seasons) includes data of significantly lower accuracy. The differences at these high latitudes are spatially noisy, which will reduce the pattern correlation.

The new data is from V4.3 of the CCI soil moisture dataset, while the original appears to use v0.1 according to the filenames referenced in the original climatology script (e.g. `ESACCI-L3S_SOILMOISTURE-SSMV-MERGED-*-fv00.1.nc`). Gruber et al (2019, https://doi.org/10.5194/essd-11-717-2019) document many improvements between these versions, which result in the dataset gaining greater coverage and accuracy. Thus the above differences are likely to arise from improvements to the CCI datasets, rather than e.g. differences in how the climatologies have been processed. The processing of climatologies used Iris to calculate time means, with default missing data tolerance.