Terrestrial Hydrology Visualization

Project Report

CCNY Spring 2016
CSC 59969 Visualization
Instructor: Professor Michael Grossberg
Group : Lingshan Jiang
Tarekul Islam
Date: May 29, 2016

Table of Contents

Website

https://terrestrial-hydrology-viz.herokuapp.com

Github repository

https://github.com/susanjiang03/Terrestrial_Hydrology_Visualization

Introduction

Terrestrial surface hydrologic state indicators provide unique insight into linkages and feedbacks in terrestrial energy, water and carbon cycles. When considered alone, they are poor indicators of climate change. This project is to use d3 javascript library to make an interactive visualization web application for finding the relation between the total 34 variables and change by seasons and years.

Data Set

http://happy.ccny.cuny.edu/indicators/README.html provided by **Nick Steiner**

There are 4 categories of the indicators:

- **Land Surface Freeze/Thaw State**

  The freeze/thaw (F/T) state of the ground over North America has been characterized using microwave frequency active and passive remote sensing data combined to produce a unified F/T product.

  4 different states: ft_frozen, ft_thawed, ft_trans, ft_itrans

- **Surface Inundation**

  The state of surface inundated area fraction (Fw) for North America is assembled from the global time series of the NASA Inundated Wetlands Earth System Data Record (ESDR)(http://wetlands.jpl.nasa.gov).

5 Histogram of fractional inundation variable (originally in percent): fw*fw*06*swe -3, fw*fw*06*swe -2, fw*fw*06*swe -1, fw*fw*06*swe +1, fw*fw*06_swe +2;

- **Snow Water Equivalent (SWE)**

  The daily estimate of snow water equivalent (SWE), the amount of snow on the ground in the equivalent water depth.

  5 Histogram of average SWE (originally in inches): swe*swe*average-*3, swe*swe*average*-2, swe*swe*average-*1, swe*swe*average*+1, swe*swe*average_+2

- **Energy Flux Datasets**

  Radiative fluxes are associated with downwelling short/longwave solar radiation and subsequent emission of the heat from Earthâ€™s surface at longwave frequencies.

  5 Histogram of short-wave upwelling (originally in W/m^2): energy*sw*up-*3, energy*sw*up*-2, energy*sw*up-*1, energy*sw*up*+1, energy*sw*up_+2:

  5 Histogram of short-wave down-welling (originally in W/m^2): energy*sw*dn-*3, energy*sw*dn*-2, energy*sw*dn-*1, energy*sw*dn*+1, energy*sw*dn_+2:

  5 Histogram of long-wave upwelling (originally in W/m^2): energy*lw*up-*3, energy*lw*up*-2, energy*lw*up-*1, energy*lw*up*+1, energy*lw*up_+2:

  5 Histogram of ling-wave down-welling (originally in W/m^2). energy*lw*dn-*3,energy*lw*dn*-2,energy*lw*dn-*1,energy*lw*dn*+1,energy*lw*dn_+2:

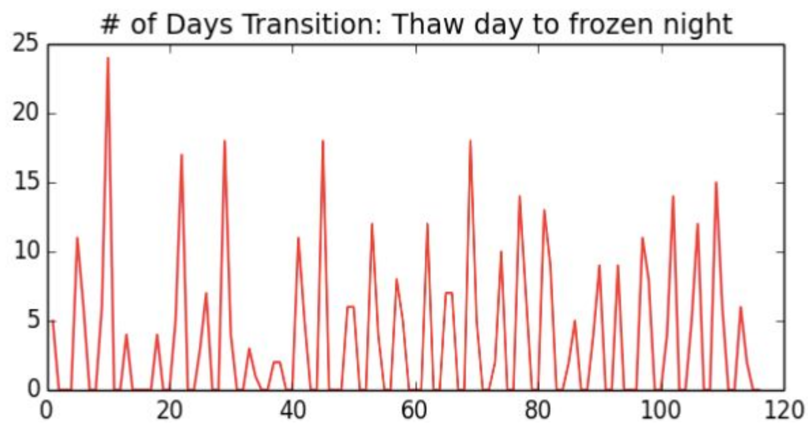
In the given data csv file, each row contains the index location, start of 3-month period(YYYY-MM-DD) (date), latitude, longitude, and data for above 34 variables for the location and the 3 month period. For periods where data is not available, the value 255 is used. Data should range from 0 to ~90 (+- several days depending on year and months). For non-binary data sources (all except Freeze/Thaw) the data is expressed as a count of zscores (histogram) calculated per-location. There are total 154,427 different locations data for each 115 three-month period. The dates are starting from 1978-12-01 to 2007-09-01. Not all 34 variables data are available in each date. To study the relations among 34 variables, and we filter out the 10% of 154,427 locations data for dates starting from 1991-12-01 to 2007-09-01.

Task

Carefully tracking the combination of these state variables with the location they influence supports how both are linked. Ask questions: How does the location affect the climate? How does the climate affect the locations. Are similar locations around the world linked to the same climate patterns? To find answers and more importantly create even more questions from the data is the main goal.

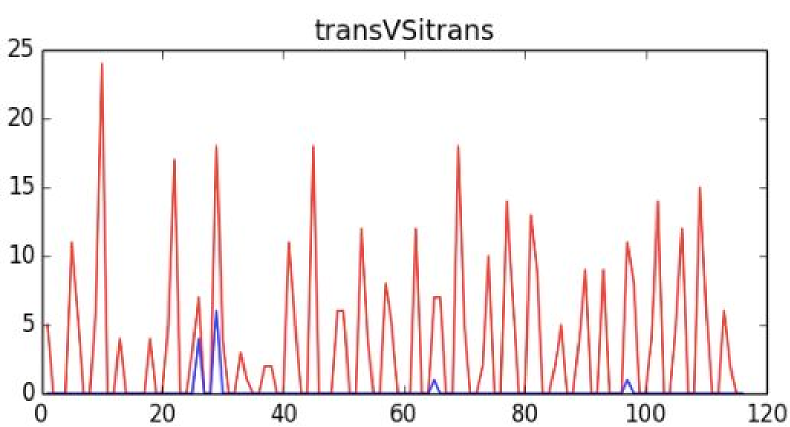
How

In the beginning we created scatter plots for one variable for a certain location.



_: Lat: 37.98058 Lon: 140.0957

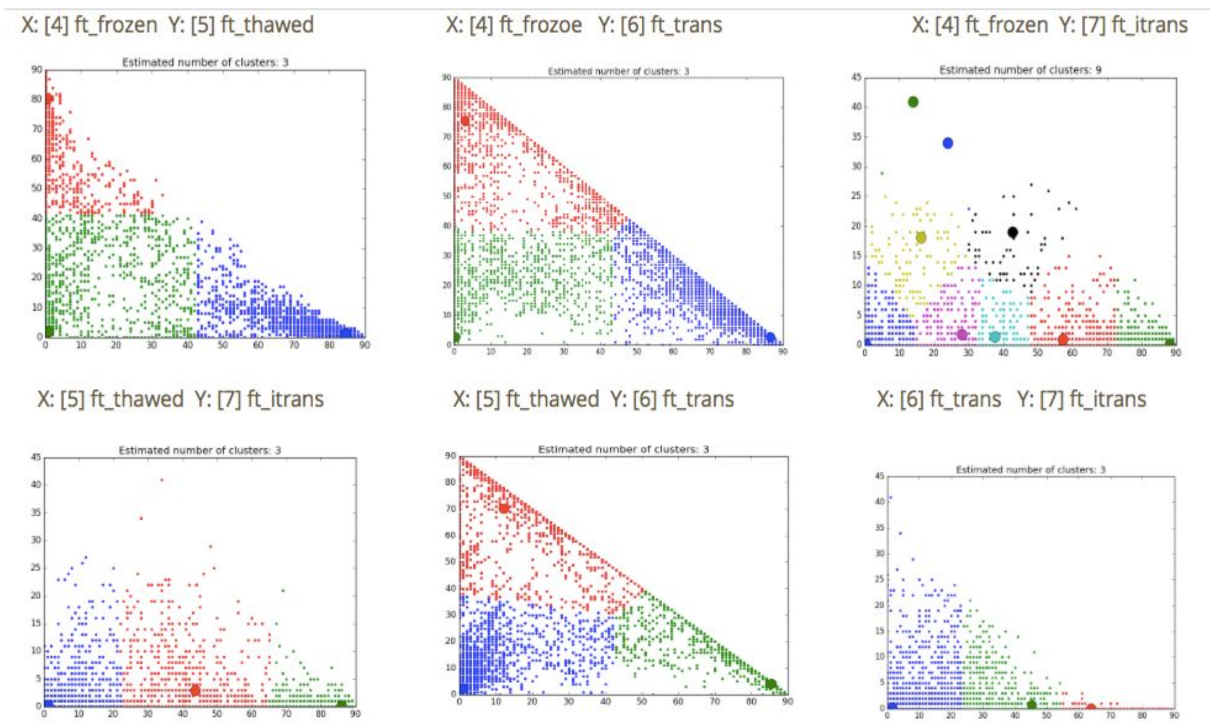
We plotted two variables together in same graph



Red: Thawed during day and frozen during night Blue: Thawed during night and frozen during day

Mean Shift Clustering Algorithm

Mean shift is a procedure for locating the maxima of a density function given discrete data sampled from that function.[1]
We use mean shift algorithm in python to find the clutters between variables in each date.. The examples of the clusters between 2 variables as following:



We give fit X as data matrix of 34 columns and about 15,000 rows to run the mean shift clustering for each date, find the cluster labels for each location, and append label data to the last column in the origin csv data file.

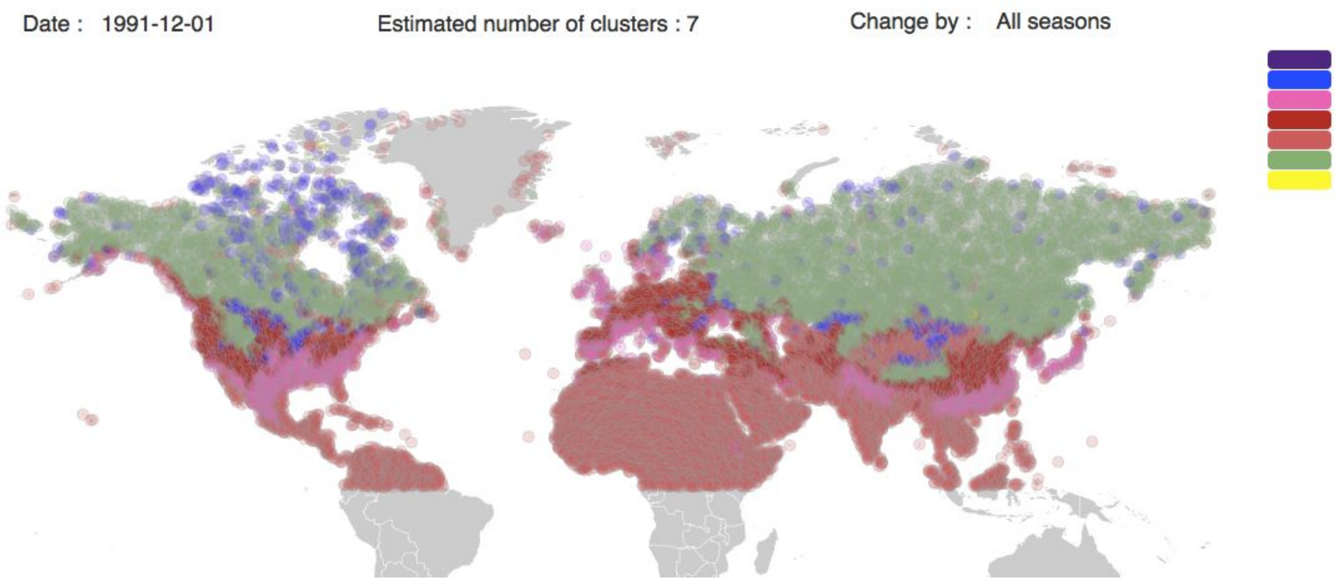
1 Cheng, Yizong (August 1995). "Mean Shift, Mode Seeking, and Clustering". *IEEE Transactions on Pattern Analysis and Machine Intelligence* (IEEE) **17** (8): 790–799. doi:10.1109/34.400568.

```
def meanShift_clustering(data_as_list, quantile):

    X = np.array(data_as_list)
    ##########################################################################################
    # Compute clustering with MeanShift
    # The following bandwidth can be automatically detected using
    print "\n++++++++  Running Mean shift clustering algorithm for %d variables: ++++++++" %len(data_as_list[0])
    print "quantile: %r" %quantile
    bandwidth = estimate_bandwidth(X, quantile = quantile )
    ms = MeanShift(bandwidth = bandwidth, bin_seeding = True)
    #ms = MeanShift#(bin_seeding=True)
    ms.fit(X)

    labels = ms.labels_
    cluster_centers = ms.cluster_centers_

    #print cluster_centers

    labels_unique = np.unique(labels)
    n_clusters_ = len(labels_unique)

    print "Finish running Mean shift clustering algorithm."
    print("\nnumber of estimated clusters : %d" % n_clusters_)
    return labels,n_clusters_, cluster_centers
```

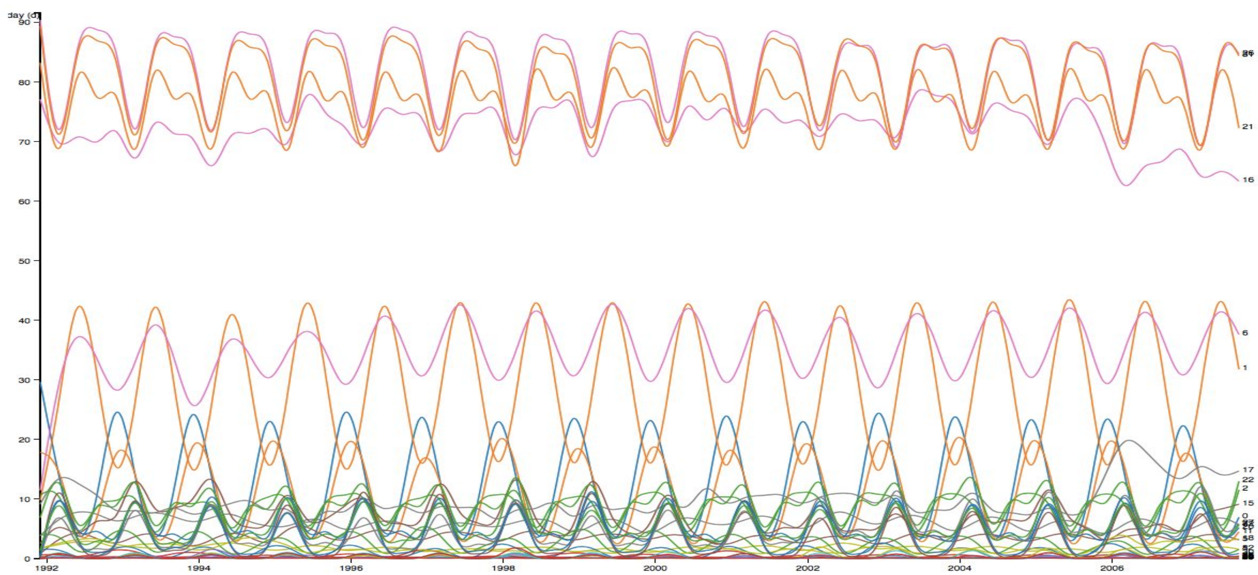
```
# def meanShift_clustering_writeCSV_plotMap(in_fileName,list_of_index,quantile):
def meanShift_clustering_writeCSV(the_dir,date,list_of_index,quantile):
    in_fileName = ""
    file_name = ""
    for fileName in os.listdir(the_dir):
        if fileName.endswith(date + ".csv"):
            file_name = fileName
            in_fileName = os.path.join(the_dir, fileName)
    if in_fileName == "":
        print "Error, no csv file for %s in %s"%(date,the_dir)
        return
    ####### (1) MEANSHIFT #############################################
    print "\n\nMeanShift for %s" %in_fileName
    data_as_list = get_data_as_list(in_fileName, list_of_index)
    labels, n_clusters_, cluster_centers = meanShift_clustering(data_as_list,quantile)


    ####### (2) Change the label by find out the location whose values are closet to cluster centers.
    #change order the label by average lat of same clusetering row
    all_data_as_list = get_data_as_list(in_fileName,range(0,38))
    # new_labels, new_cluster_centers, list_of_location_center =
    new_labels, new_center_cluster_centers = change_cluster_label_order_by_center_lat_lon(date, all_data_as_list,labels,cluster_centers)
    # ######## (3) Append LABELS TO CSV , Append to each row in in_fileName ########################################

    append_labelsClusters_to_csv_file(in_fileName, 'clusterLabel',new_labels)

    # ##########(4) WRITE CLUSTER CENTERS TO CSV in CLUSTERCENTERCSV folder ##########################

    append_clusterCenters_to_csv_file(new_center_cluster_centers)
```

The web app

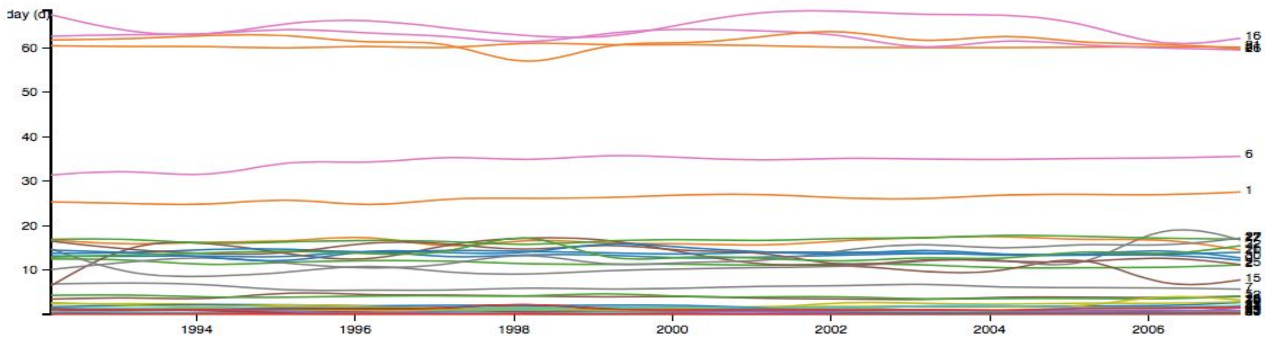
Use the data file with cluster labels obtained after running mean shift clustering function, we plot the different clusters on the world map using d3. First choose to view the change by years or which season. Then it will start from the starting date. It may take some time to load all the points on the page. By clicking the arrow or moving the slider, the map will show color clusters for different dates. If it does not show the points, refresh the page by change to other dates and come back to the date. We can also see the line chart below the map, which show the averages of 34 variables changed by the dates.



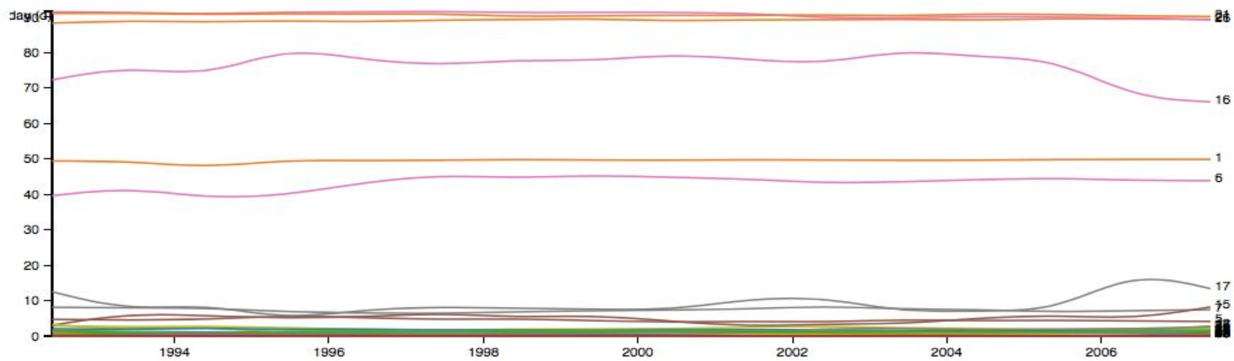
Average of 34 variables change by all season:



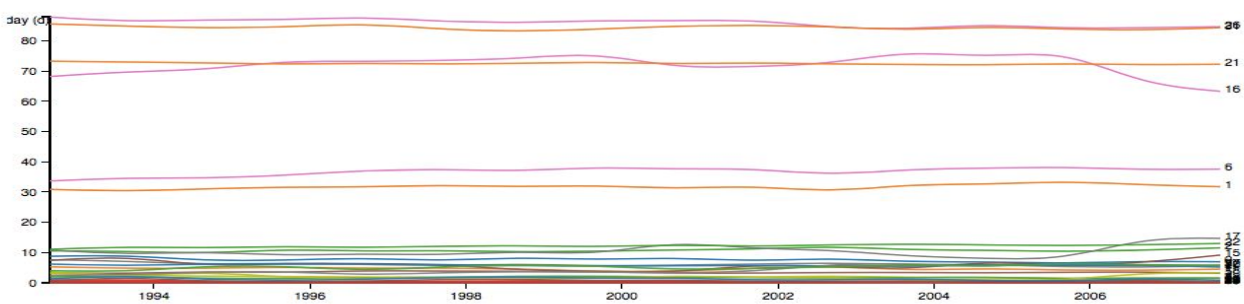
Change by springs:



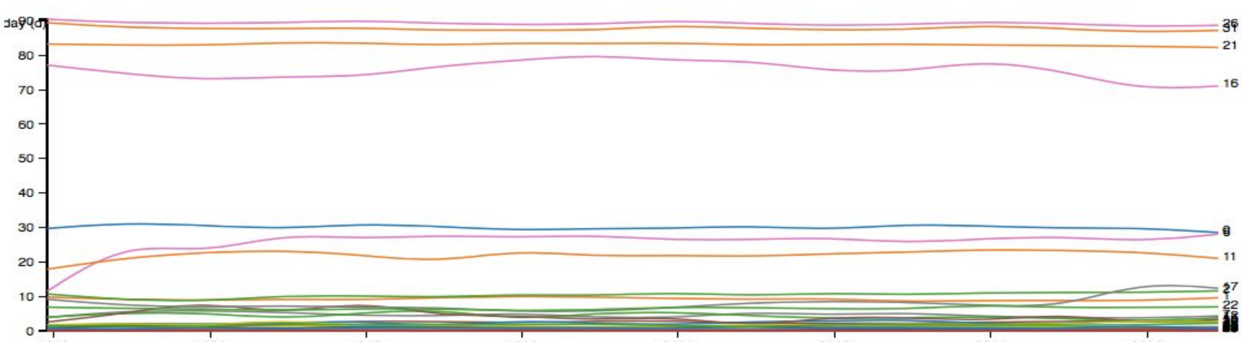
Summer:



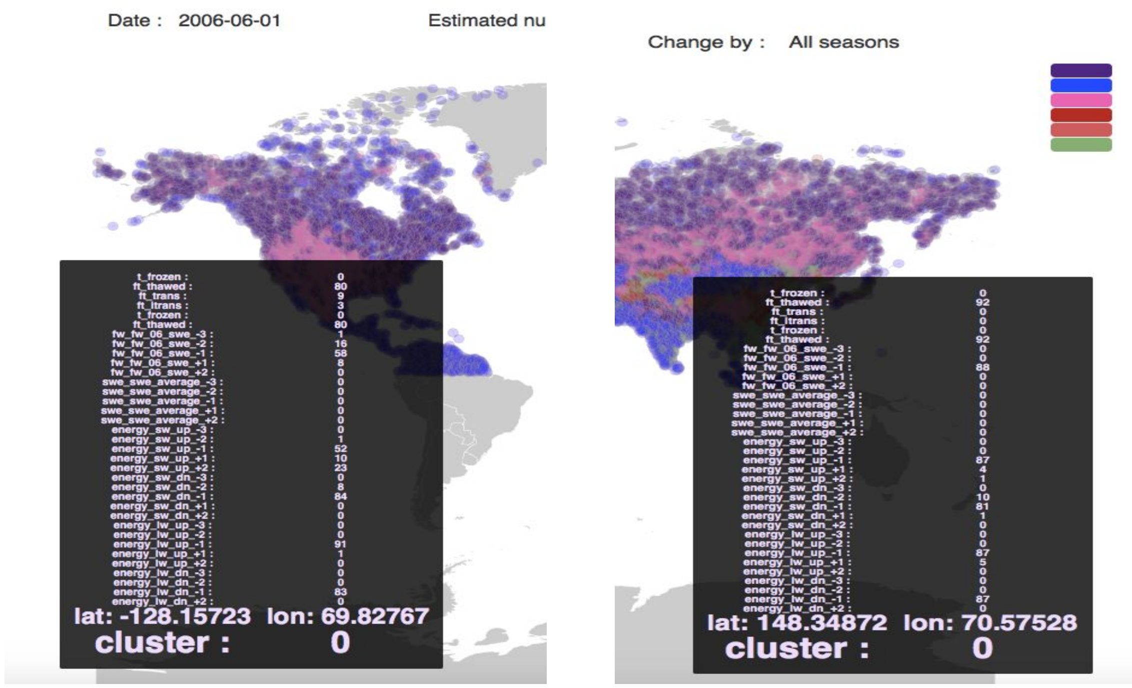
Fall:



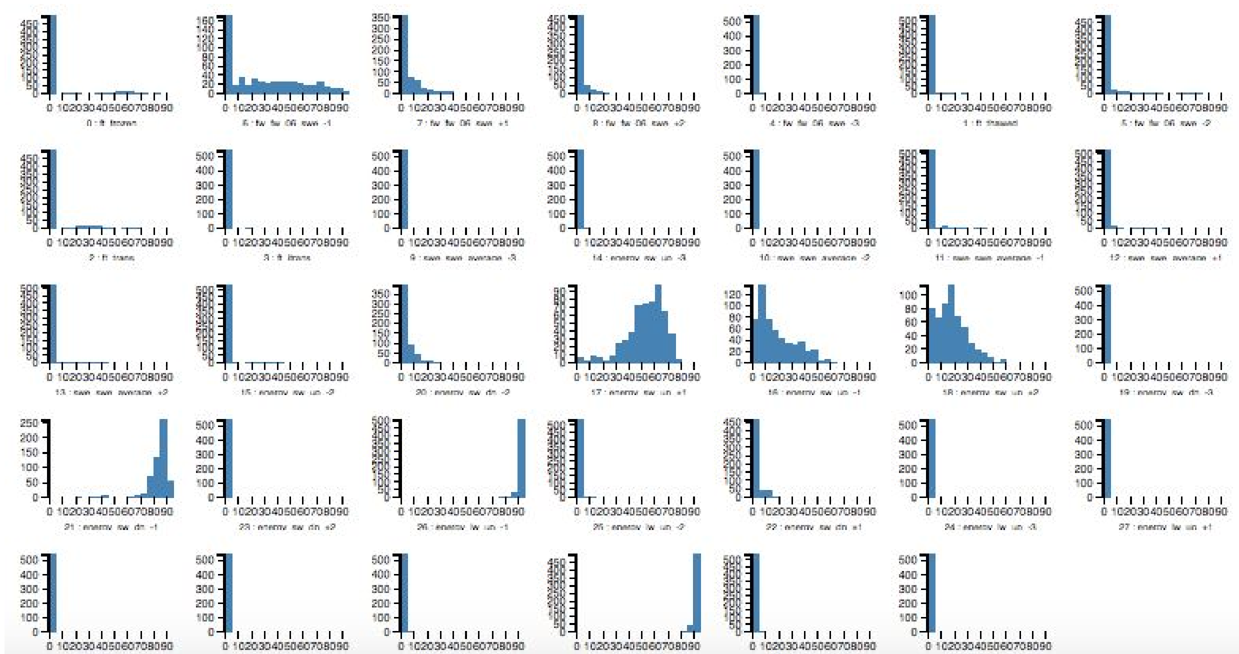
Winter:



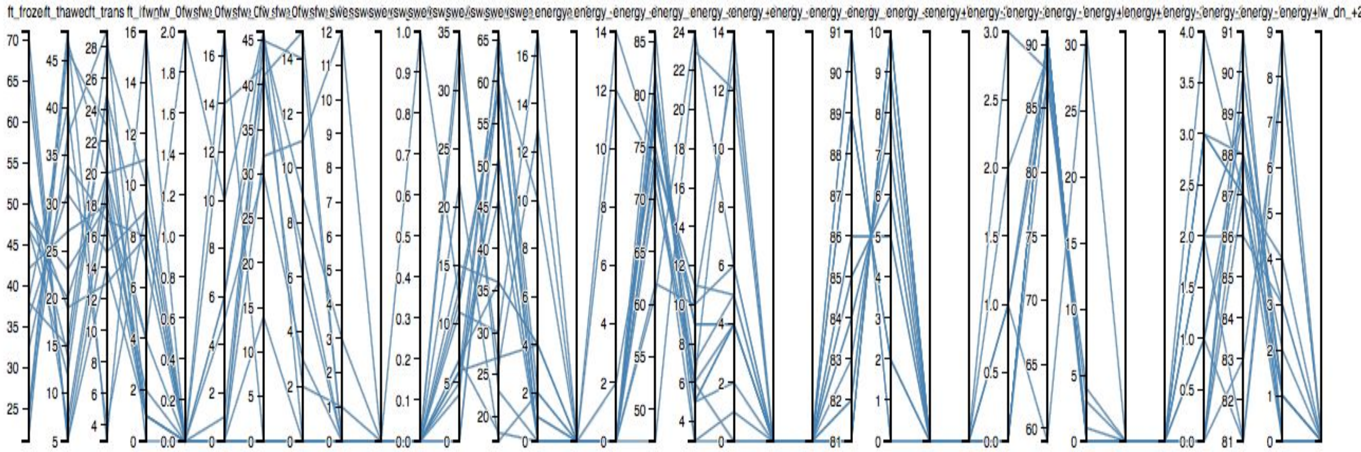
Hover the points on the map, we can view the detail data for the location.

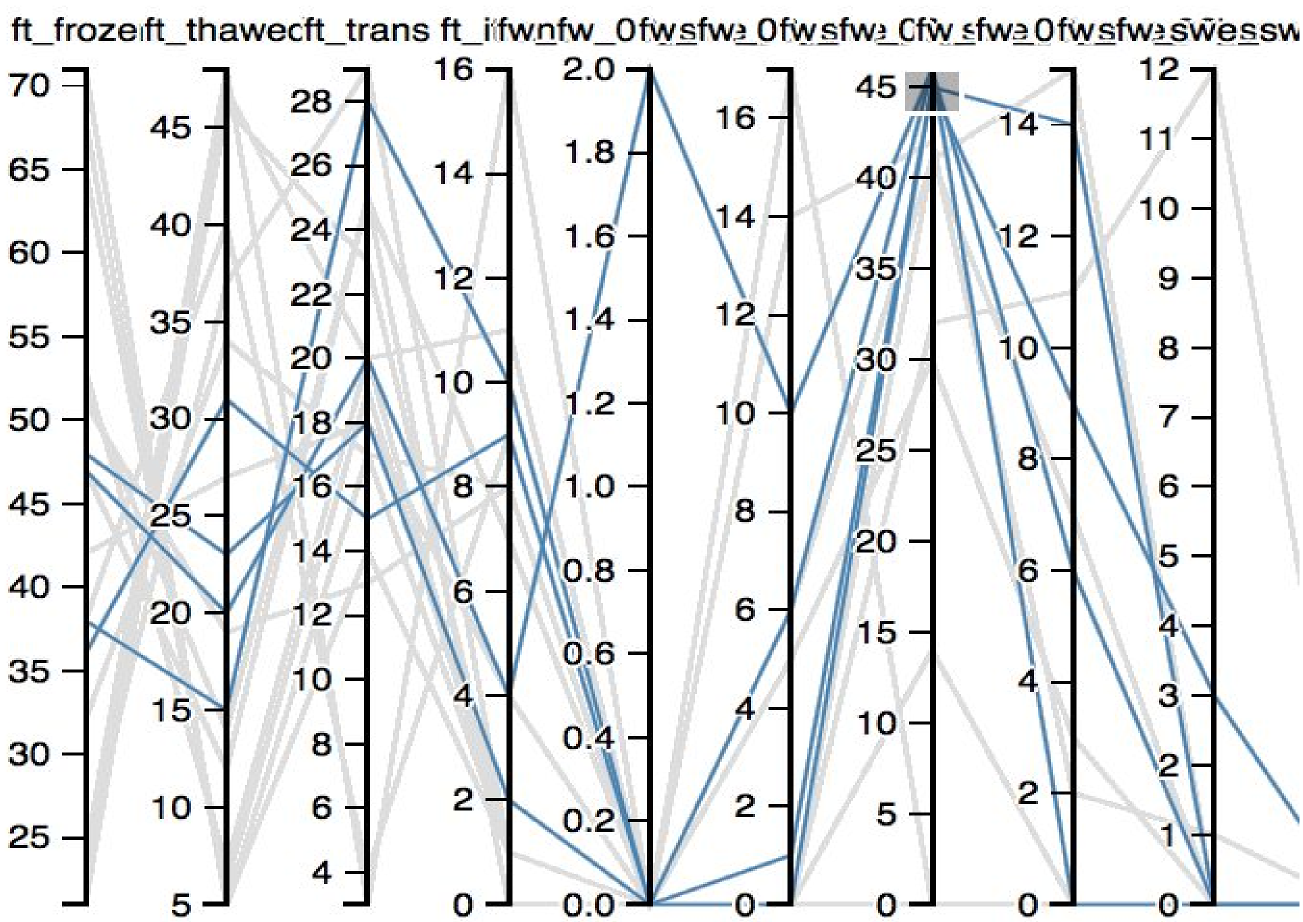


By clicking a cluster button in the button group on the upper right of the map, it will show the histogram for all 34 variables of this cluster.



We also use the parallel coordinate graph in d3 to show the relation among the variables.





Conclusion

As you can see, we find similar patterns in each map. The areas above the equator have same colors. The areas near the equator share similar colors. The areas below the equator have similar colors also. For instance when we looked at the variables to see what was similar between two areas across the ocean but above the equator we found that the ft_frozen, ft_thawed, ft_trans, ft_itrans, energy_sw_up_-1 ,energy_sw_dn_-1 , energy_sw_up_-1 , and energy_lw_dn_-1 variables were similar. We can say that the **Land Surface Freeze/Thaw State, Energy Flux Datasets** shown above are a good indicator of which regions are similar.