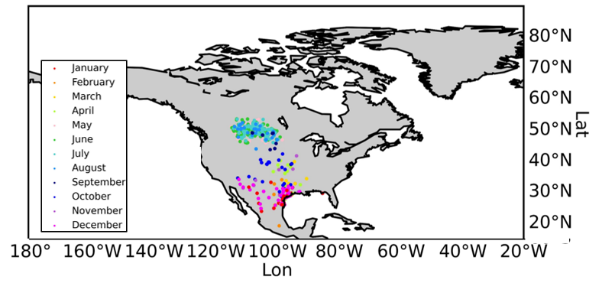
**Figures and the codes for them**

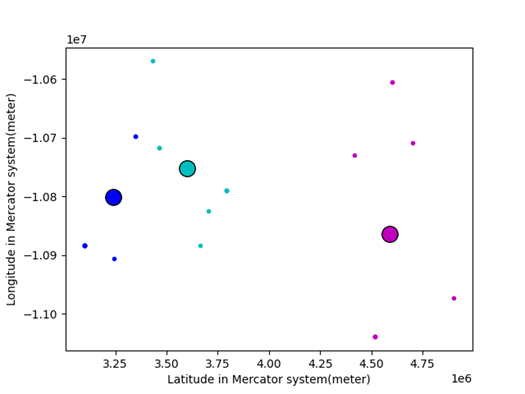
1.Show the raw data



Code:

def map\_2(save\_path,csv\_name):  
 m = Basemap(llcrnrlat=-60, urcrnrlat=90, llcrnrlon=-180, urcrnrlon=-20) # Instantiate a map  
 m.drawcoastlines() # Draw the coastline  
 m.drawmapboundary(fill\_color='white')  
 m.fillcontinents(lake\_color='white') # Draw the continents and fill them in white  
  
 parallels = np.arange(-90., 90., 10.) # Draw latitudes with ranges [-90,90] and intervals of 10  
 m.drawparallels(parallels, labels=[False, True, True, False], color='none')  
 meridians = np.arange(-180., 180., 20.) # Draw the longitude with a range of [-180,180] and an interval of 10  
 m.drawmeridians(meridians, labels=[True, False, False, True], color='none')  
  
 # plt.show()  
  
 datalist = readcsv(save\_path + csv\_name)  
 datalist = datalist[1:]  
  
 LON = []  
 LAT = []  
 for i in range(1):  
 Lat = [[float(data[0]),int(data[2])] for data in datalist]  
 Lon = [[float(data[1]),int(data[2])] for data in datalist]  
 LON.append(Lon)  
 LAT.append(Lat)  
   
 for doc in range(1):  
 colorMap = ['red', 'darkorange', 'gold', 'greenyellow', 'pink', 'limegreen', 'mediumturquoise',  
 'dodgerblue',  
 'navy', 'blue', 'mediumorchid', 'fuchsia']  
 # Show labels  
 label = ['January', 'February', 'March', 'April', 'May', 'June', 'July', 'August', 'September', 'October',  
 'November', 'December']  
  
 marker = ['x', 'o', '.', '+', '<', '\_', '^', 'v', 'H', '|', 's', '\*']  
 j = 0  
 # print(len(lon))  
 flag = True  
  
 col1 = 43101  
 for i in range(13):  
 # print(i)  
 if doc == 0:  
   
 LON1 = [data[0] for data in LON[doc] if data[1]>=col1 and data[1]<col1 + 30]  
 LAT1 = [data[0] for data in LAT[doc] if data[1]>=col1 and data[1]<col1 + 30]  
 col1 = col1 + 30  
 m.scatter(LON1, LAT1, color=colorMap[j],s=1, label=label[j])  
 # plt.show()  
 j += 1  
 if j == 12:  
 j = 0  
 if flag:  
 plt.legend(loc='lower left', shadow=True)  
 flag = False  
 continue  
  
 plt.xlabel('Lon', labelpad=10)  
 plt.ylabel('Lat')  
 plt.savefig(save\_path + '{}.jpg'.format(csv\_name.replace('.csv', '')), dpi=1000)  
 # plt.show()  
 plt.close()

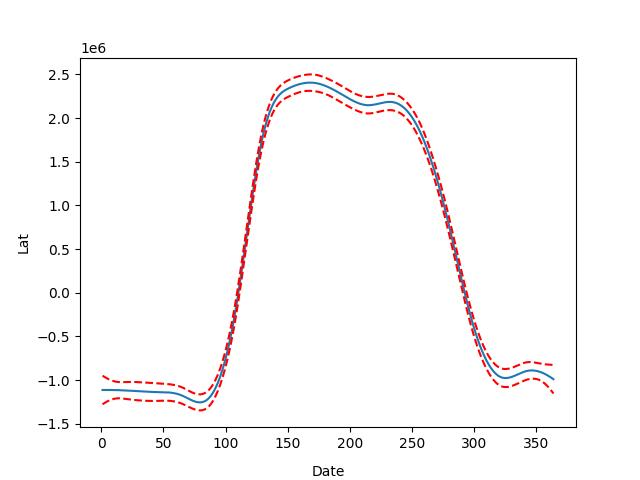
2. The daily centroids from Mean shift



Code:

def mean\_shift(SLDF\_df):  
 # datas = pd.read\_excel('data/clean\_window\_data.xlsx')  
 datas = SLDF\_df.drop(['SLDF'], axis=1)  
 result = []  
 result.append(["LATITUDE", "LONGITUDE", "OBSERVATION DATE"])  
 for date in range(43101, 43466):  
 #for date in range(43101, 43119):  
 print(date)  
 data = datas.loc[date == datas['OBSERVATION DATE']] # .values.tolist()#["answer"]  
 data = data.iloc[:, :2]  
 data = np.array(data)  
 if len(data) == 0:  
 continue  
  
 ms = MeanShift()  
 ms.fit(data)  
 labels = ms.labels\_  
 cluster\_centers = ms.cluster\_centers\_  
  
 labels\_unique = np.unique(labels)  
 n\_clusters = len(labels\_unique)  
  
 for c in cluster\_centers:  
 result.append([float(c[0]), float(c[1]), date])  
  
 colors = cycle('bcmyk')  
 if date % 20 == 0:  
 for k, color in zip(range(n\_clusters), colors):  
 # current\_member indicates true if the label is k and false if not  
 current\_member = labels == k  
 cluster\_center = cluster\_centers[k]  
 # Draw plots  
 plt.plot(data[current\_member, 0], data[current\_member, 1], color + '.')  
 # Draw circles  
 plt.plot(cluster\_center[0], cluster\_center[1], 'o',  
 markerfacecolor=color,  
 markeredgecolor='k',  
 markersize=14)  
  
 plt.xlabel('Latitude in Mercator system(meter)')  
  
 plt.ylabel('Longitude in Mercator system(meter)')  
 plt.show()  
 plt.close()  
 return result

3. Fit the latitude and longitude with date by GAM

图表, 直方图

描述已自动生成

Code:

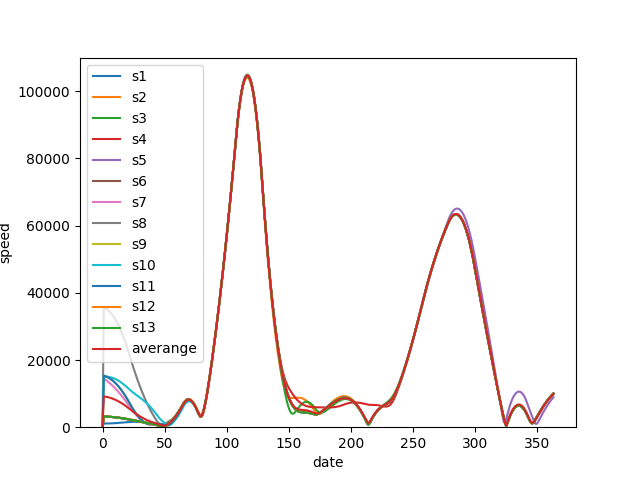
def gam\_pic(gam,save\_path,csv\_name,key,x\_y):  
 XX = gam.generate\_X\_grid(term=0, n=365)  
 # plt.scatter([data for data in range(364)], data, color='b', marker='o')  
 plt.plot(XX[:, 0], gam.partial\_dependence(term=0, X=XX))  
 plt.plot(XX[:, 0], gam.partial\_dependence(term=0, X=XX, width=.95)[1], c='r', ls='--')  
 plt.xlabel('Date', labelpad=10)  
 plt.ylabel(x\_y)  
  
 # plt.show()  
 plt.savefig(save\_path + '\\' + csv\_name.replace('.csv', '') + '\\' + 'gam{}{}.jpg'.format(x\_y,key + 1))  
 plt.close()  
  
#gam() for fitting longitude and latitude with time  
def gam(save\_path, csv\_name,key):  
 df = pd.read\_excel(save\_path + '\\' + csv\_name.replace('.csv','') + '\\' + 'ni\_traj{}.xls'.format(key + 1),  
 sheet\_name='Sheet1')  
 date = df["date"]  
 x = df["X"]  
 y = df["Y"]  
 gam\_model = LinearGAM().fit(date, x)  
 gam\_pic(gam\_model,save\_path,csv\_name,key,'Lat')  
 predictions\_x = gam\_model.predict(date)  
 gam\_model = LinearGAM().fit(date, y)  
 gam\_pic(gam\_model, save\_path,csv\_name, key, 'Lon')  
 predictions\_y = gam\_model.predict(date)  
 datas = [["X\*", "Y\*","date"]]  
 # datas = []  
 for i in range(len(x)):  
 datas.append([predictions\_x[i], predictions\_y[i],date[i]])  
 data\_write(save\_path + '\\' + csv\_name.replace('.csv','') + '\\' + 'result\_{}.xls'.format(key + 1), datas)  
  
 Lat = []  
 Lon = []  
 for i in range(len(x)):  
 res = webMercator2wgs84(predictions\_y[i], predictions\_x[i])  
 Lon.append(res[0])  
 Lat.append(res[1])  
 return Lon, Lat

4. Show the estimation results

图示, 地图

描述已自动生成Code:

ef map\_1(save\_path,csv\_name):  
  
 # plt.rcParams['figure.figsize'] = (28, 8)  
 # plt.show()  
  
 excel\_list = os.listdir(save\_path + '\\' + csv\_name.replace('.csv', '') + '\\')  
 excel\_list1 = []  
 for csv\_excel in excel\_list:  
 if 'ni\_traj' in csv\_excel:  
 excel\_list1.append(csv\_excel)  
  
 LON = []  
 LAT = []  
 for i in range(len(excel\_list1)):  
 Lon, Lat = Gam(save\_path, csv\_name, i)  
 LON.append(Lon)  
 LAT.append(Lat)  
  
 m = Basemap(llcrnrlat=-60, urcrnrlat=90, llcrnrlon=-180, urcrnrlon=-20) # Instantiate a map  
 m.drawcoastlines() # Draw the coastline  
 m.drawmapboundary(fill\_color='white')  
 m.fillcontinents(lake\_color='white') # Draw the continents and fill them in white  
  
 parallels = np.arange(-90., 90., 10.) # Draw latitudes with ranges [-90,90] and intervals of 10  
 m.drawparallels(parallels, labels=[False, True, True, False], color='none')  
 meridians = np.arange(-180., 180., 20.) # Draw the longitude with a range of [-180,180] and an interval of 10  
 m.drawmeridians(meridians, labels=[True, False, False, True], color='none')  
 for doc in range(0, len(LON)):  
 colorMap = ['red', 'darkorange', 'gold', 'greenyellow', 'pink', 'limegreen', 'mediumturquoise',  
 'dodgerblue',  
 'navy', 'blue', 'mediumorchid', 'fuchsia']  
 # Show labels  
 label = ['January', 'February', 'March', 'April', 'May', 'June', 'July', 'August', 'September', 'October',  
 'November', 'December']  
  
 marker = ['x', 'o', '.', '+', '<', '\_', '^', 'v', 'H', '|', 's', '\*','.','x', 'o', '.', '+', '<', '\_', '^']  
 j = 0  
 # print(len(lon))  
 flag = True  
  
  
  
 for i in range(0, len(LON[doc]) - 30, 30):  
 # print(i)  
 if doc == 0:  
 m.plot(LON[doc][i:i + 30], LAT[doc][i:i + 30], marker=marker[doc], linewidth=0.4,  
 color=colorMap[j],  
 markersize=0.5, label=label[  
 j])  
 # plt.show()  
 j += 1  
 if j == 12:  
 j = 0  
 if flag:  
 plt.legend(loc='lower left', shadow=True)  
 flag = False  
 continue  
 else:  
 m.plot(LON[doc][i:i + 30], LAT[doc][i:i + 30], marker=marker[doc], linewidth=0.4,  
 color=colorMap[j],  
 markersize=0.5)  
 # plt.show()  
 j += 1  
 if j == 12:  
 j = 0  
 if flag:  
 plt.legend(loc='lower left', shadow=True)  
 flag = False  
 continue  
  
 plt.xlabel('Lon', labelpad=10)  
 plt.ylabel('Lat')  
 plt.savefig(save\_path + '\\' + csv\_name.replace('.csv', '') + '\\' + 'map.jpg', dpi=1000)  
 # plt.show()  
 plt.close()

5. Speed

Code:

def calculate\_speed(path):  
  
 df = pd.read\_excel(path, header=None)  
 list\_speed = [0]  
 for i in range(1,364):  
 # x1 = float(df.iloc[i + 1, 0])  
 x = float(df.iloc[i + 1, 0]) - float(df.iloc[i, 0])  
 y = float(df.iloc[i + 1, 1]) - float(df.iloc[i, 1])  
 distance = (x \*\* 2 + y \*\* 2) \*\* 0.5  
 list\_speed.append(distance)  
 return list\_speed  
  
def speed\_main(path):  
 path1 = get\_file\_list(path)  
 list\_averange = []  
 index = 1  
 for i in path1:  
 name1 = i.split('\\')[-1]  
 if name1[0:2] == 're':  
 #print(name1)  
 result = calculate\_speed(i)  
 result1 = [[datas] for datas in result ]  
 savecsvs(path + '\\speed\\' + 's{}.csv'.format(str(index)),result1)  
 list\_averange.append(result)  
  
 #plt.xlim(0,600)  
 plt.plot(result,label='s{}'.format(str(index)))  
 plt.xlabel("date")  
 plt.ylabel("speed")  
 #plt.title('%s'%i)  
 index = index + 1  
  
 result2 = []  
 for m in range(len(list\_averange[0])):  
 sum = 0  
 for n in range(len(list\_averange)):  
 sum = sum + float(list\_averange[n][m])  
 savecsv(path + '\\speed\\' + 'sa.csv',[sum/len(list\_averange)])  
 result2.append(sum/len(list\_averange))

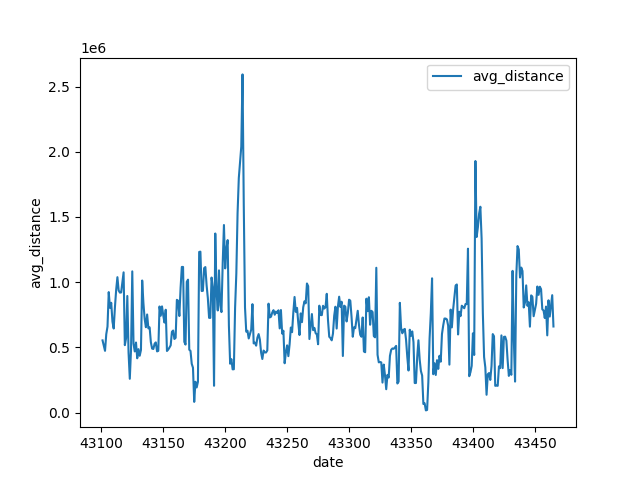
图表, 折线图, 直方图

描述已自动生成7. Offset distance

Code:

def get\_distance\_point2line(point, line):  
  
 line\_point1, line\_point2 = np.array(line[0:2]), np.array(line[2:])  
 vec1 = line\_point1 - point  
 vec2 = line\_point2 - point  
 distance = np.abs(np.cross(vec1, vec2)) / np.linalg.norm(line\_point1 - line\_point2)  
 return distance  
  
#Index calculation of offset distance  
def calculate\_distance(path\_list):  
  
 list\_distance = []  
 for l in path\_list:  
 name1 = l.split('\\')[-1]  
 if name1[0] == 'r':  
 tmp = pd.read\_excel(l, header=None)  
 list\_distance.append(tmp)  
  
 mean\_x = []  
 mean\_y = []  
 for q in list\_distance:  
 mean\_x.append(float(q.iloc[1, 0]))  
 mean\_y.append(float(q.iloc[1, 1]))  
  
 mean\_x1 = np.mean(mean\_x)  
 mean\_y1 = np.mean(mean\_y)  
 print(mean\_x1, mean\_y1)  
  
  
  
 dict1 = {}  
 list1 = []  
 max\_list\_x = []  
 max\_list\_y = []  
 for r in path\_list:  
 name1 = r.split('\\')[-1]  
 if name1[0:2] == 're':  
 tmp = pd.read\_excel(r, header=None)  
 for l in range(1,365):  
 x1 = float(tmp.iloc[1, 0])  
 y1 = float(tmp.iloc[1, 1])  
  
 x = float(tmp.iloc[l, 0])  
 y = float(tmp.iloc[l, 1])  
 distance1 = ((x - x1) \*\* 2 + (y - y1) \*\* 2) \*\* 0.5  
 dict1[distance1] = l  
 list1.append(distance1)  
 max\_distance = max(list1)  
 max\_x\_1 = float(tmp.iloc[dict1[max\_distance], 0])  
 max\_y\_1 = float(tmp.iloc[dict1[max\_distance], 1])  
 max\_list\_x.append(max\_x\_1)  
 max\_list\_y.append(max\_y\_1)  
 list1 = []  
 dict1 = {}  
  
 mean\_x2 = np.mean(max\_list\_x)  
 mean\_y2 = np.mean(max\_list\_y)  
 print(mean\_x2, mean\_y2)  
  
  
 line = [mean\_x1, mean\_y1, mean\_x2, mean\_y2]  
  
 list11 = []  
 for r in path\_list:  
 list\_tmp = []  
 name1 = r.split('\\')[-1]  
 if name1[0:2] == 're':  
 tmp = pd.read\_excel(r, header=None)  
 for p in range(1,365):  
 x = float(tmp.iloc[p, 0])  
 y = float(tmp.iloc[p, 1])  
 distance2 = get\_distance\_point2line((x, y), line)  
 list\_tmp.append(distance2)  
 list11.append(list\_tmp)  
 return list11  
  
def off\_distance\_main(path):  
 path1 = get\_file\_list(path)  
 result = calculate\_distance(path1)  
 count1 = 0  
 tmp = 0  
 index = 1  
 for j in result:  
 list\_averange = []  
 result2 = []  
 items = [[datas] for datas in j]  
 savecsvs(path + '\\off\_distance\\' + 'd{}.csv'.format(str(index)),items)  
 index = index + 1  
 for b in range(len(j)):  
 for l in result:  
 tmp += l[b]  
 averange = int(tmp/len(result))  
 tmp = 0  
 result2.append(averange)  
 items = [[datas] for datas in result2]  
 savecsvs(path + '\\off\_distance\\' + 'da.csv', items)  
  
 count = 1  
 for l in result:  
 name = path1[count].split('\\')[-1]  
 #plt.xlim(0,360)  
 plt.plot(l,label='d{}'.format(count))  
 plt.xlabel("date")  
 plt.ylabel("distance")  
 #plt.title('%s'%l)  
 count += 1  
 #plt.title('%s'%name)  
  
 #plt.legend()  
 #plt.savefig('q2/DistancePreDay.png')  
 #plt.show()  
  
 #plt.plot(result2,label='aveVar')  
 #plt.xlim(0,420)  
 #plt.ylim()  
 plt.xlabel("date")  
 plt.ylabel("distance")  
 #plt.title('averangeVar')  
 plt.legend()  
 plt.savefig(path + '\\off\_distance\\DistancePreDayAverange.png')  
 plt.show()

7. Centroid average distances

Code:

def get\_avg(datalist):#Obtain the Euclidean distance for each day  
 data1 = [datalist[0][0],datalist[0][1]]  
 avg\_datalist = []  
 for i in range(1,len(datalist)):  
 datas = [datalist[i][0],datalist[i][1]]  
 avg\_datalist.append(distEclud(data1,datas))  
 if avg\_datalist:  
 avg = sum(avg\_datalist)/(len(datalist) - 1)  
 else:  
 avg = 0  
 return avg  
  
def avg\_distance(save\_path):  
  
 df=pd.read\_excel(save\_path + '\shift.xls')  
 train\_data = np.array(df)  
 excel\_list = train\_data.tolist() # list  
 # excel\_list = excel\_list[1:]  
  
 datelist = get\_date\_list(excel\_list)  
  
 avg\_list = []  
 for date in datelist:  
 same\_date = get\_same\_date\_list(date,excel\_list)  
 avg\_date = get\_avg(same\_date)  
 avg\_list.append(avg\_date)  
 savecsv(save\_path + r'\avg.csv',[date,avg\_date])  
  
 plt.plot(datelist, avg\_list, label="avg\_distance")  
 plt.xlabel("date")  
 plt.ylabel("avg\_distance")  
 plt.legend(loc="best")  
 plt.gca().ticklabel\_format(axis="x", useOffset=False)  
 plt.savefig(save\_path + r'\avg.png')  
 plt.show()