

In [39]:

```
import pandas as pd
import numpy as np
from matplotlib import pyplot as plt
from tqdm import tqdm

import warnings
warnings.filterwarnings("ignore")

from multiprocessing import cpu_count, Pool
```

executed in 157ms, finished 22:35:09 2018-10-29

In [2]:

```
tr = pd.read_pickle('../data/train.pkl')
tr_log = pd.read_pickle('../data/train_log.pkl')
```

executed in 549ms, finished 21:32:07 2018-10-29

In [3]:

```
tr.head()
```

executed in 22ms, finished 21:32:07 2018-10-29

Out[3]:

	object_id	ra	decl	gal_l	gal_b	ddf	hostgal_specz	ho
0	615	349.046051	-61.943836	320.796530	-51.753706	1	0.0000	
1	713	53.085938	-27.784405	223.525509	-54.460748	1	1.8181	
2	730	33.574219	-6.579593	170.455585	-61.548219	1	0.2320	
3	745	0.189873	-45.586655	328.254458	-68.969298	1	0.3037	
4	1124	352.711273	-63.823658	316.922299	-51.059403	1	0.1934	

In [4]:

```
tr.tail()
```

executed in 22ms, finished 21:32:07 2018-10-29

Out[4]:

	object_id	ra	decl	gal_l	gal_b	ddf	hostgal_specz
7843	130739978	26.718750	-14.940303	172.342697	-72.255675	0	0.0000
7844	130755807	120.101349	-62.696659	275.742955	-16.509746	0	0.1725
7845	130762946	203.108109	-55.682144	308.728904	6.727511	0	0.0000
7846	130772921	79.101562	-35.501846	239.172243	-33.827844	0	0.0000
7847	130779836	301.992188	-17.426323	25.102988	-24.511101	0	0.0000

In [5]:

tr_log.head(20)

executed in 14ms, finished 21:32:07 2018-10-29

Out[5]:

	object_id	mjd	passband	flux	flux_err	detected
0	615	59750.421875	2	-544.810303	3.622952	1
1	615	59750.429688	1	-816.434326	5.553370	1
2	615	59750.437500	3	-471.385529	3.801213	1
3	615	59750.445312	4	-388.984985	11.395031	1
4	615	59752.406250	2	-681.858887	4.041204	1
5	615	59752.414062	1	-1061.457031	6.472994	1
6	615	59752.421875	3	-524.954590	3.552751	1
7	615	59752.433594	4	-393.480225	3.599346	1
8	615	59752.445312	5	-355.886780	10.421921	1
9	615	59767.296875	2	-548.013550	3.462291	1
10	615	59767.304688	1	-815.188599	5.293019	1
11	615	59767.312500	3	-475.516052	3.340643	1
12	615	59767.324219	4	-405.663818	3.496113	1
13	615	59767.335938	5	-421.199066	6.377517	1
14	615	59770.218750	2	-554.903198	3.927843	1
15	615	59770.226562	1	-820.042786	5.875329	1
16	615	59770.234375	3	-477.004730	3.736262	1
17	615	59770.246094	4	-400.270386	3.834955	1
18	615	59770.253906	5	-415.286896	7.435979	1
19	615	59779.320312	2	-630.523682	4.333287	1

In [6]:

df = tr_log[tr_log.object_id==615]

executed in 15ms, finished 21:32:08 2018-10-29

In [7]:

df['date'] = df.mjd.astype(int)

executed in 67ms, finished 21:32:08 2018-10-29

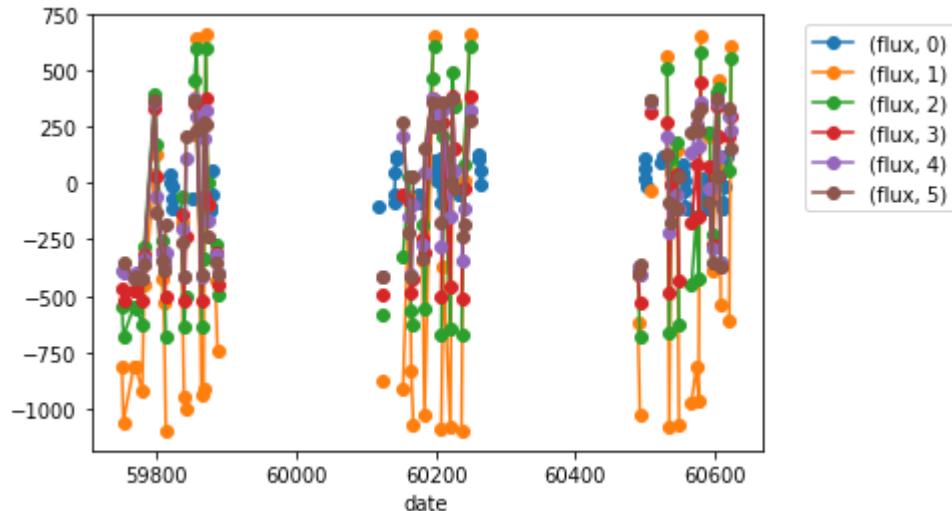
In [8]:

```
pd.pivot_table(df, index=['date'], columns=['passband'], values=['flux']).plot(marker="o",
plt.legend(bbox_to_anchor=(1.05, 1), loc='upper left')
```

executed in 706ms, finished 21:32:09 2018-10-29

Out[8]:

<matplotlib.legend.Legend at 0x7feb6fdcb860>



In [9]:

```
pd.pivot_table(df, index=['date'], columns=['passband'], values=['flux']).reset_index().hea
```

executed in 29ms, finished 21:32:09 2018-10-29

Out[9]:

passband	date	flux					
		0	1	2	3	4	5
0	59750	NaN	-816.434326	-544.810303	-471.385529	-388.984985	
1	59752	NaN	-1061.457031	-681.858887	-524.954590	-393.480225	-355.881
2	59767	NaN	-815.188599	-548.013550	-475.516052	-405.663818	-421.191
3	59770	NaN	-820.042786	-554.903198	-477.004730	-400.270386	-415.281
4	59779	NaN	-921.002502	-630.523682	-518.533997	-422.184509	-422.811

In [25]:

```

def plt_obj(oid=None, save=False, path=None, norm=False, return_df=False):
    if oid is None:
        oid = np.random.choice(tr.object_id)
        df = tr_log[tr_log.object_id==oid]
    if norm:
        df.flux /= df.flux.max()
    target = tr.loc[tr.object_id==oid, 'target'].values[0]
    photoz = tr.loc[tr.object_id==oid, 'hostgal_photoz'].values[0]
    df['date'] = df.mjd.astype(int)

    df = pd.pivot_table(df, index=['date'], columns=['passband'], values=['flux'])

    if return_df:
        return df

    df.plot(marker='o', legend=True)
    plt.legend(bbox_to_anchor=(1.05, 1), loc='upper left')
    plt.title(f'oid:{oid} target:{target} photoz:{photoz}')
    if save==True and path is not None:
        plt.savefig(path)
    return

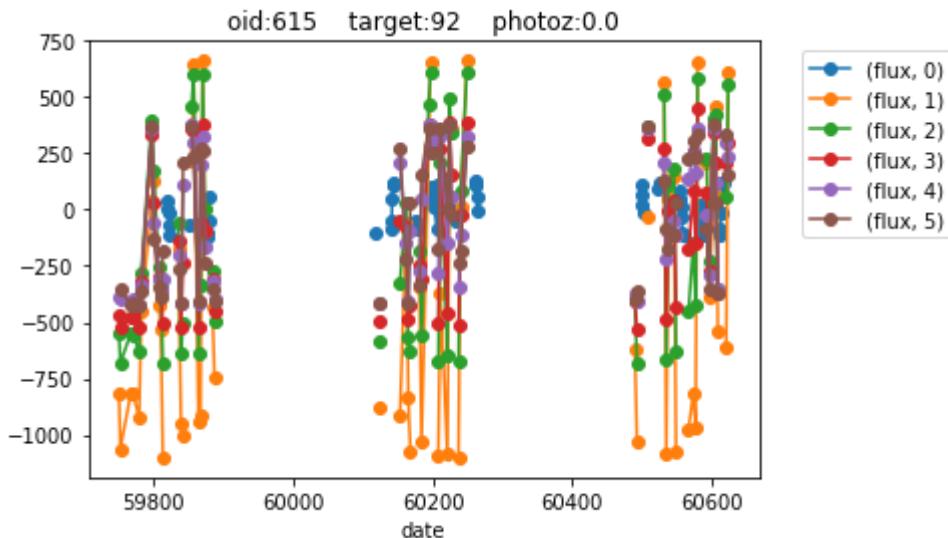
```

executed in 29ms, finished 22:16:37 2018-10-29

In [11]:

plt_obj(615)

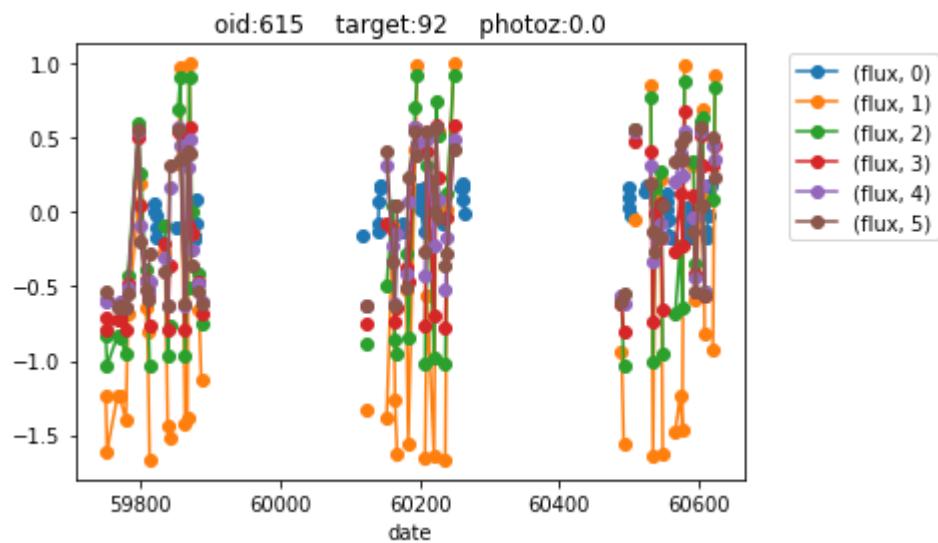
executed in 268ms, finished 21:32:12 2018-10-29



In [18]:

plt_obj(615, norm=True)

executed in 2.76s, finished 21:38:45 2018-10-29



In [19]:

```
classes = [6, 15, 16, 42, 52, 53, 62, 64, 65, 67, 88, 90, 92, 95]
li = []
for c in classes:
    li += tr[tr.target==c].sample(25).object_id.tolist()
```

executed in 22ms, finished 22:00:30 2018-10-29

In [40]:

```

classes = [6, 15, 16, 42, 52, 53, 62, 64, 65, 67, 88, 90, 92, 95]

for c in tqdm(classes):
    li = tr[tr.target==c].sample(25).object_id.tolist()
    fig, axes = plt.subplots(ncols=5, nrows=5, figsize=(28, 25), sharex=True)

    # プロット
    for i,(ax, oid) in enumerate(zip(axes.ravel(), li)):
        df=plt_obj(oid, return_df=True)

        target = tr.loc[tr.object_id==oid, 'target'].values[0]
        photoz = tr.loc[tr.object_id==oid, 'hostgal_photoz'].values[0]

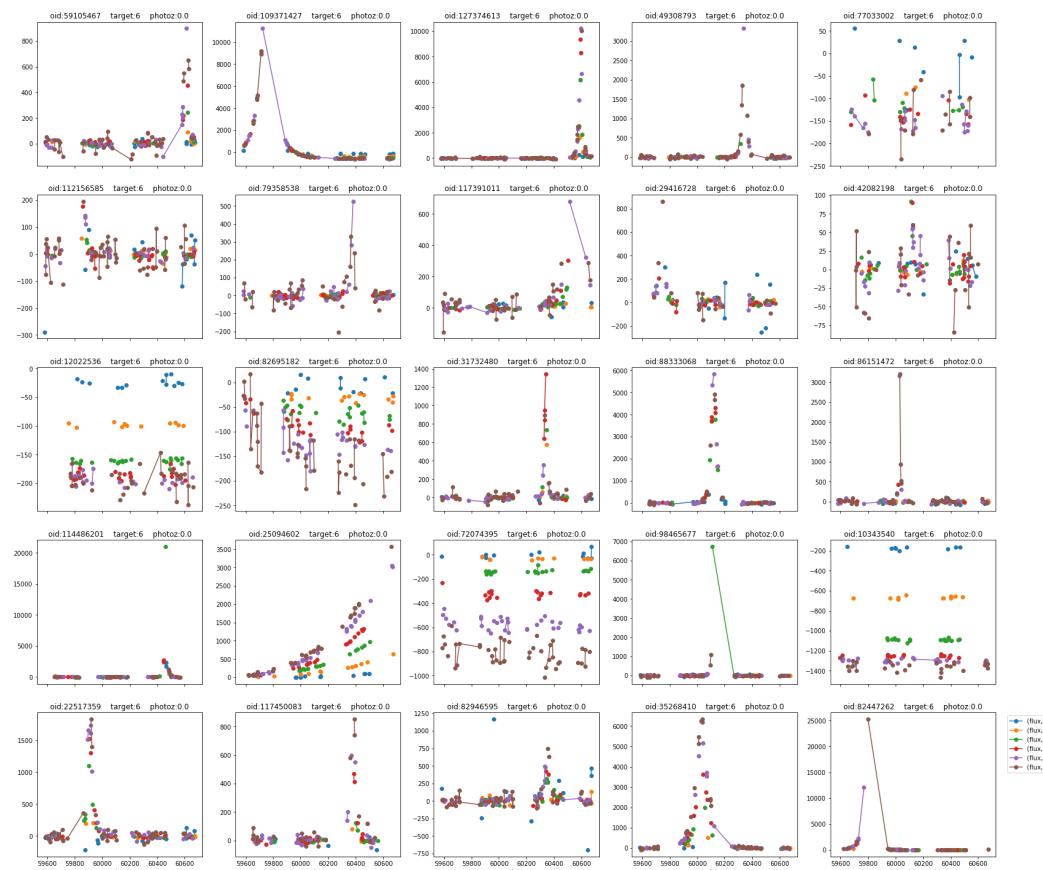
        if i==24:
            legend=True
        else:
            legend = False
        df.plot(ax=ax, marker="o", legend=legend, title=f'oid:{oid} target:{target} photoz:{photoz}')

        plt.legend(bbox_to_anchor=(1.05, 1), loc='upper left')

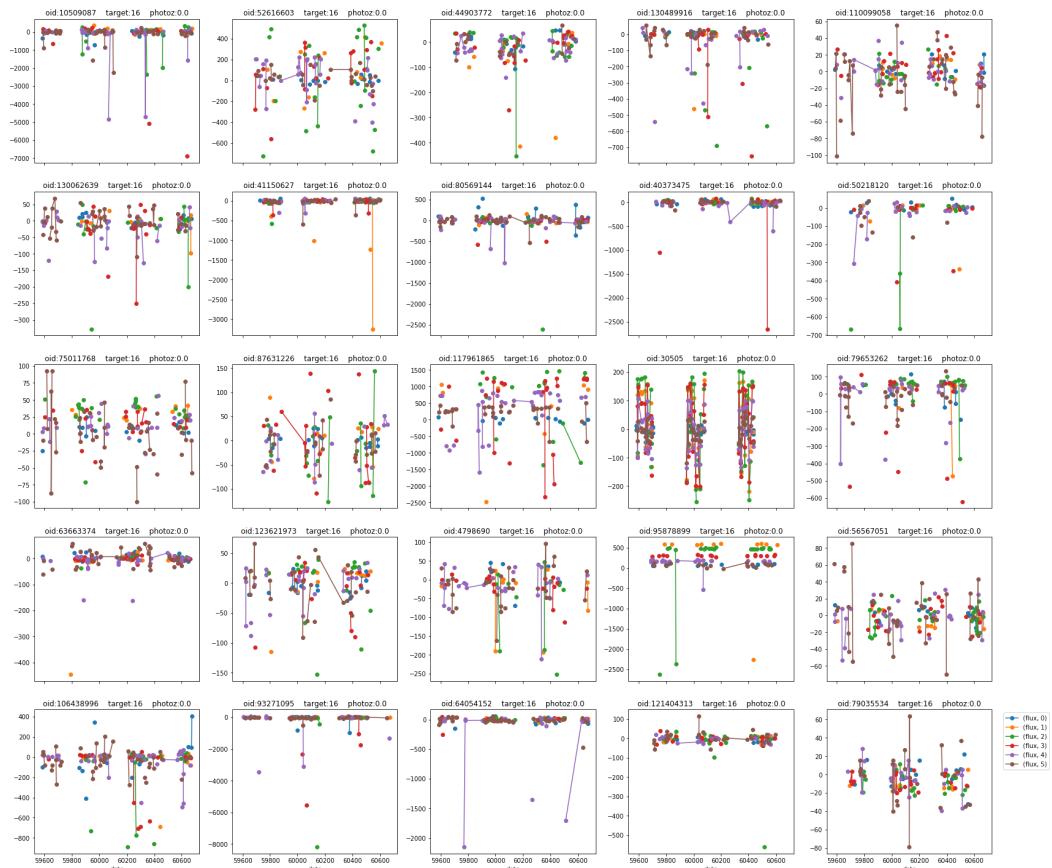
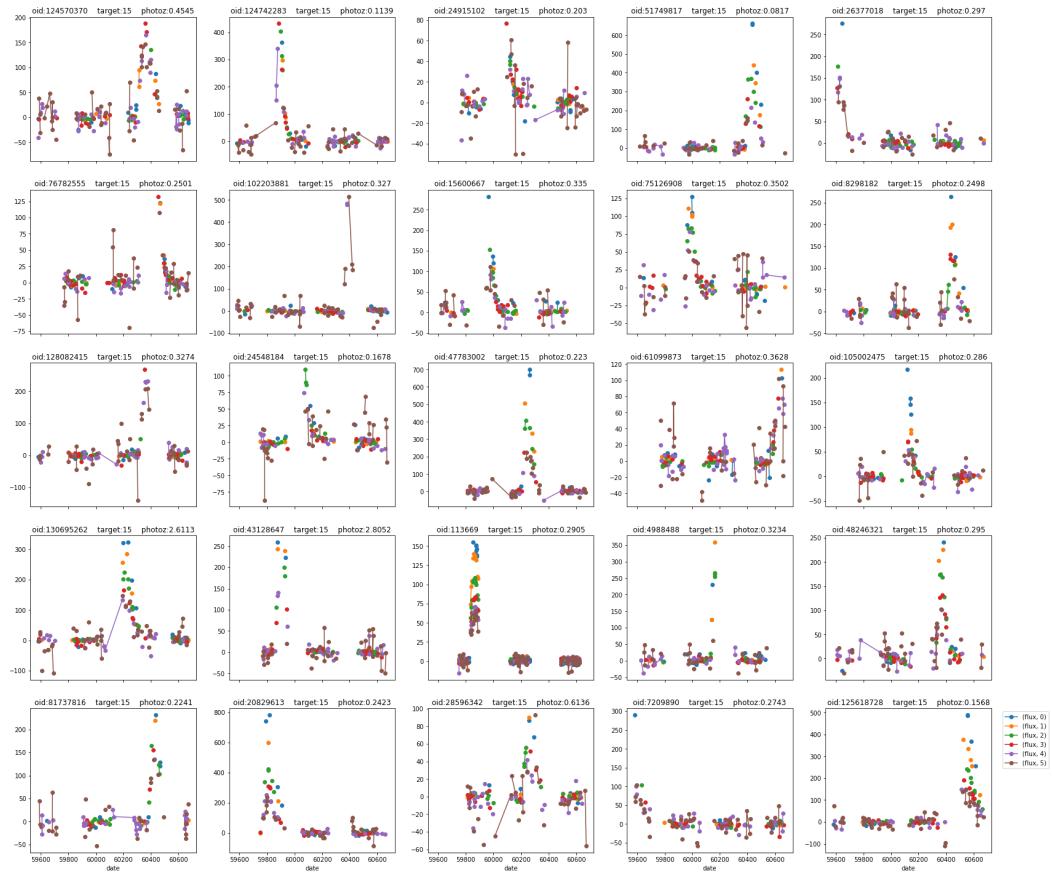
```

executed in 5m 48s, finished 22:41:06 2018-10-29

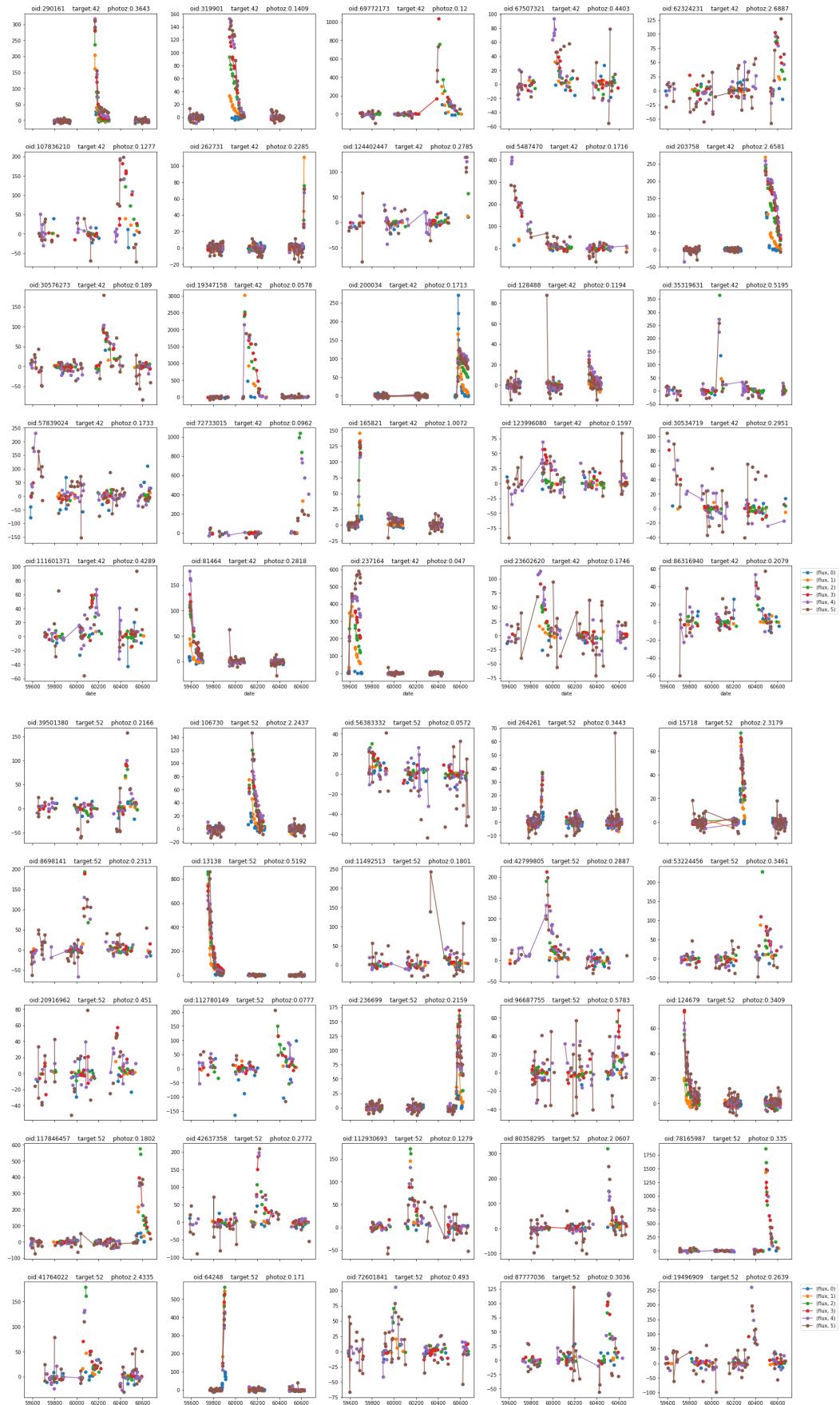
100% | 14/14 [05:20<00:00, 22.92s/it]



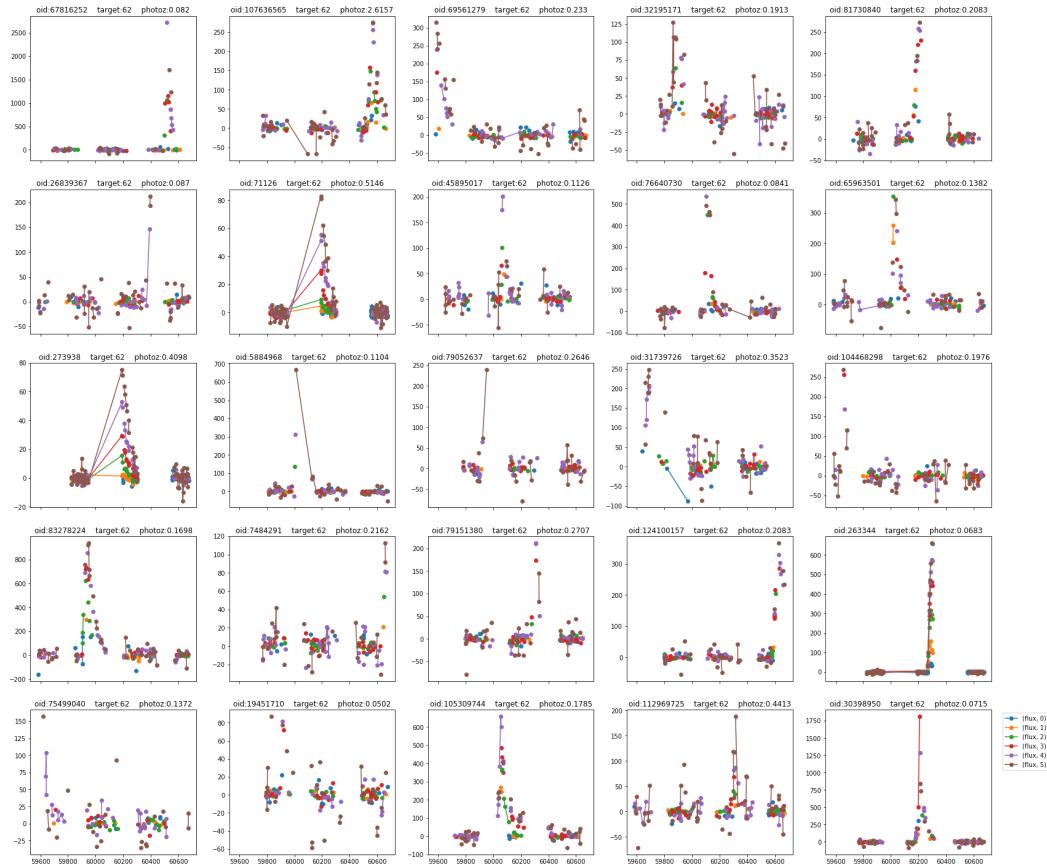
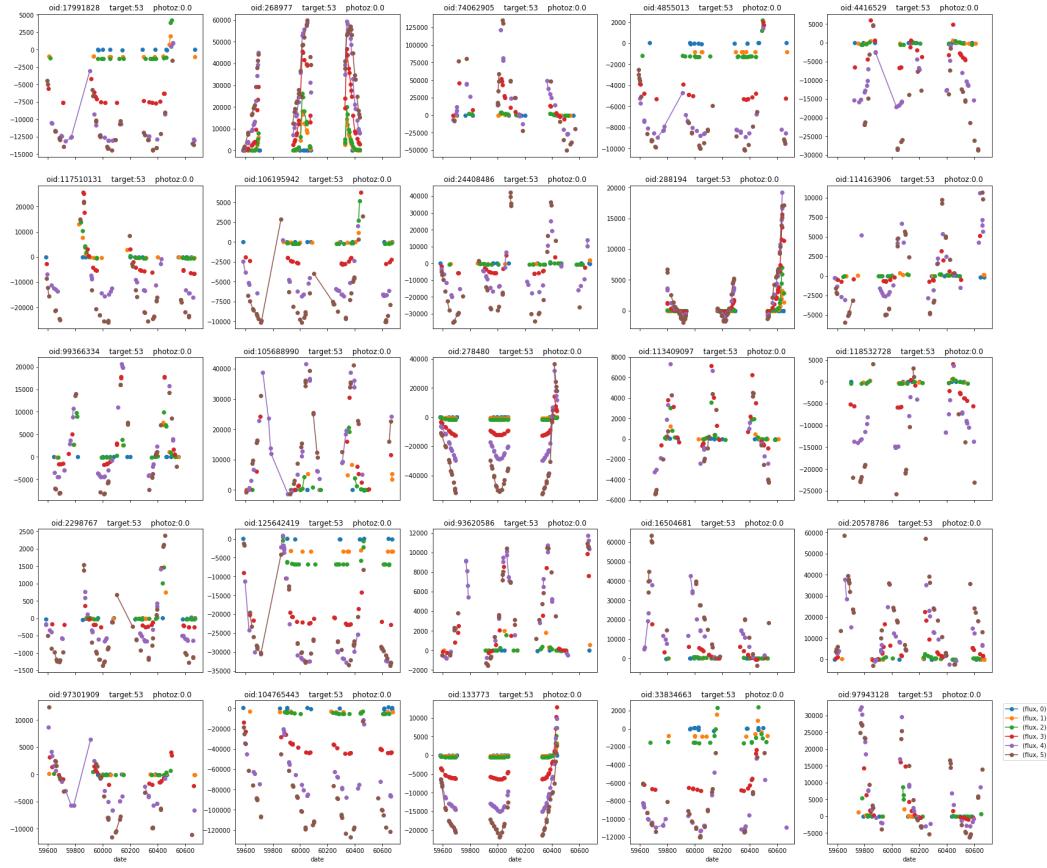
EDA_002



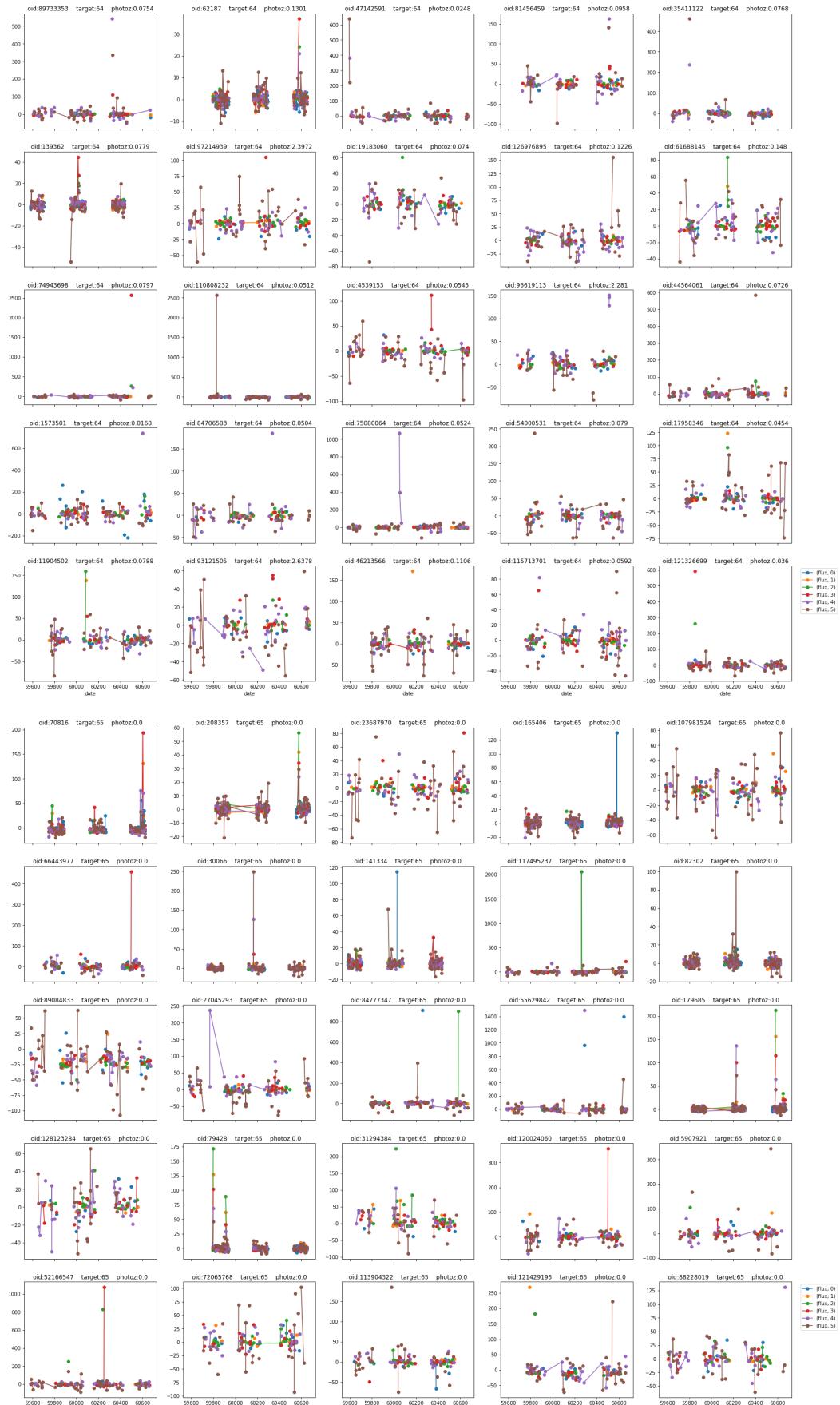
EDA_002

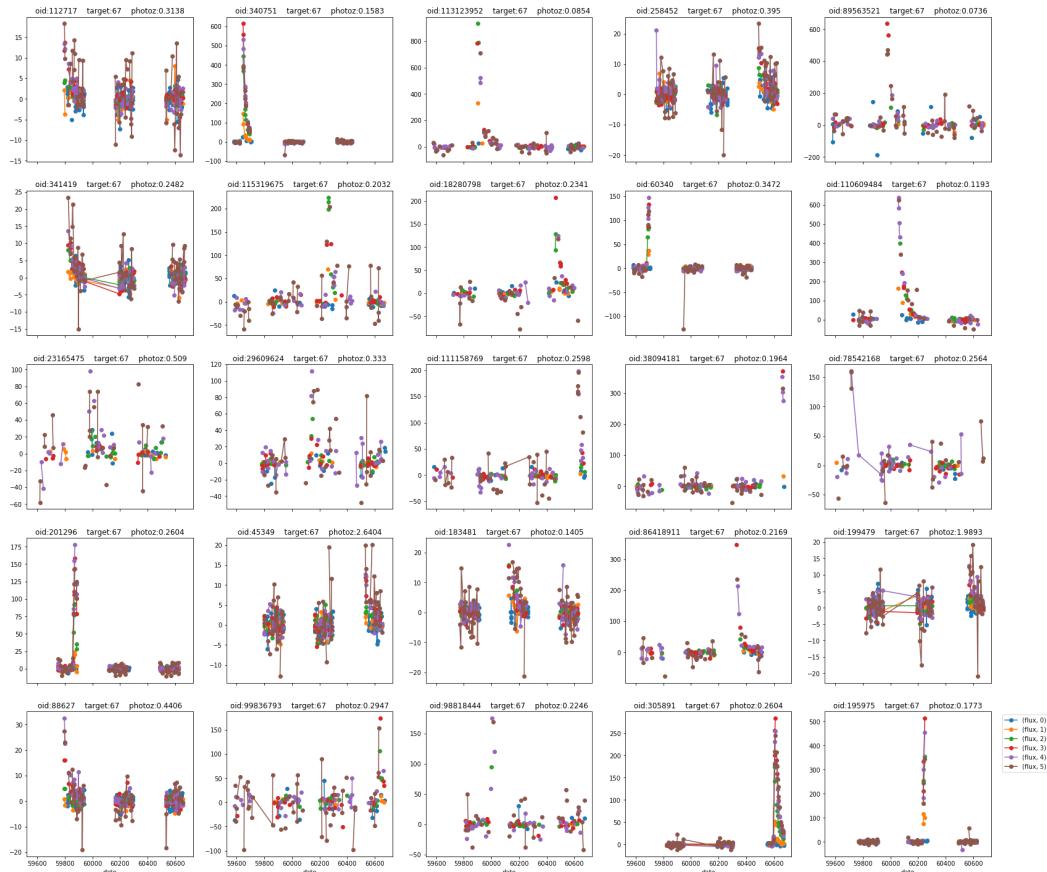


EDA_002

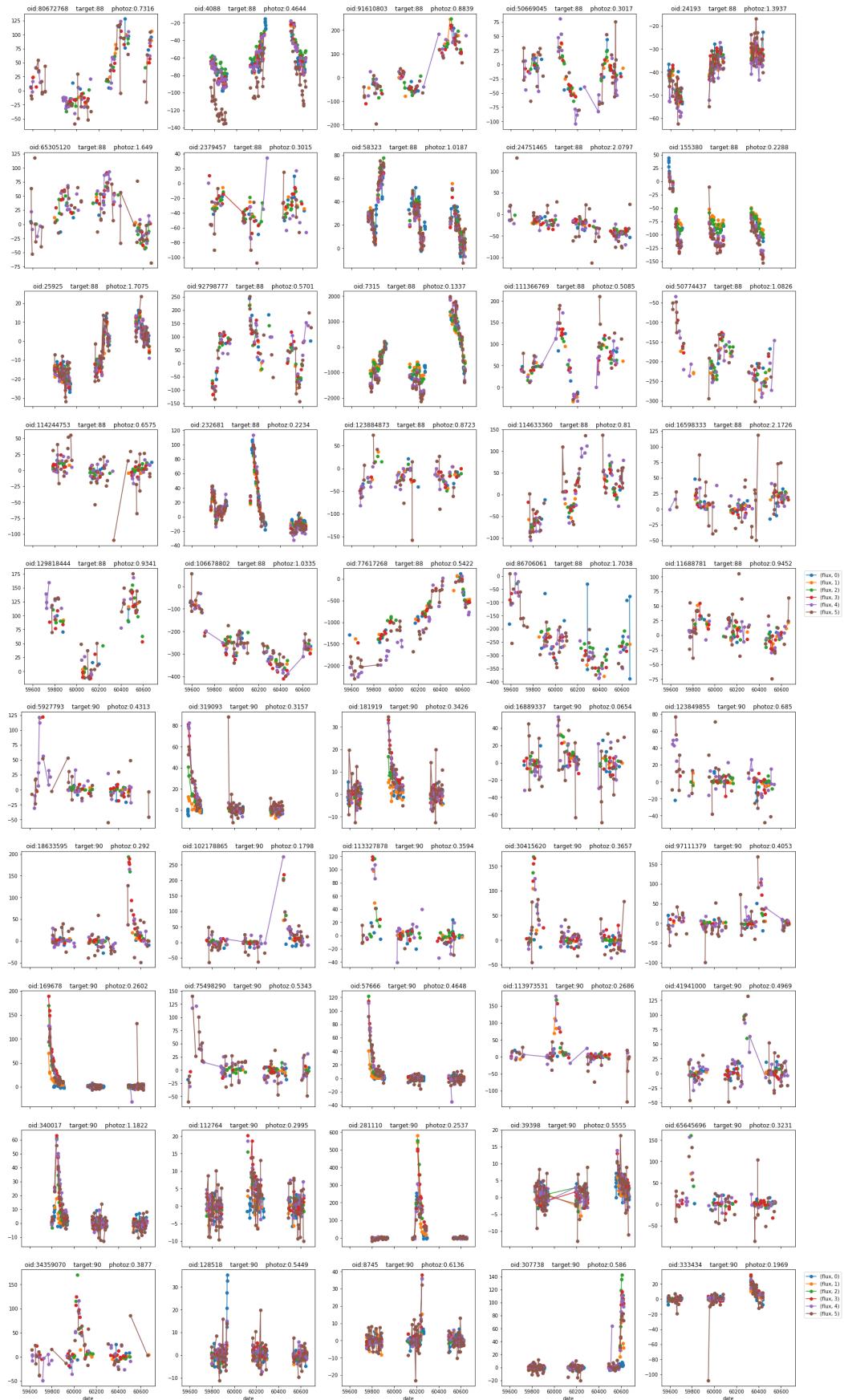


EDA_002

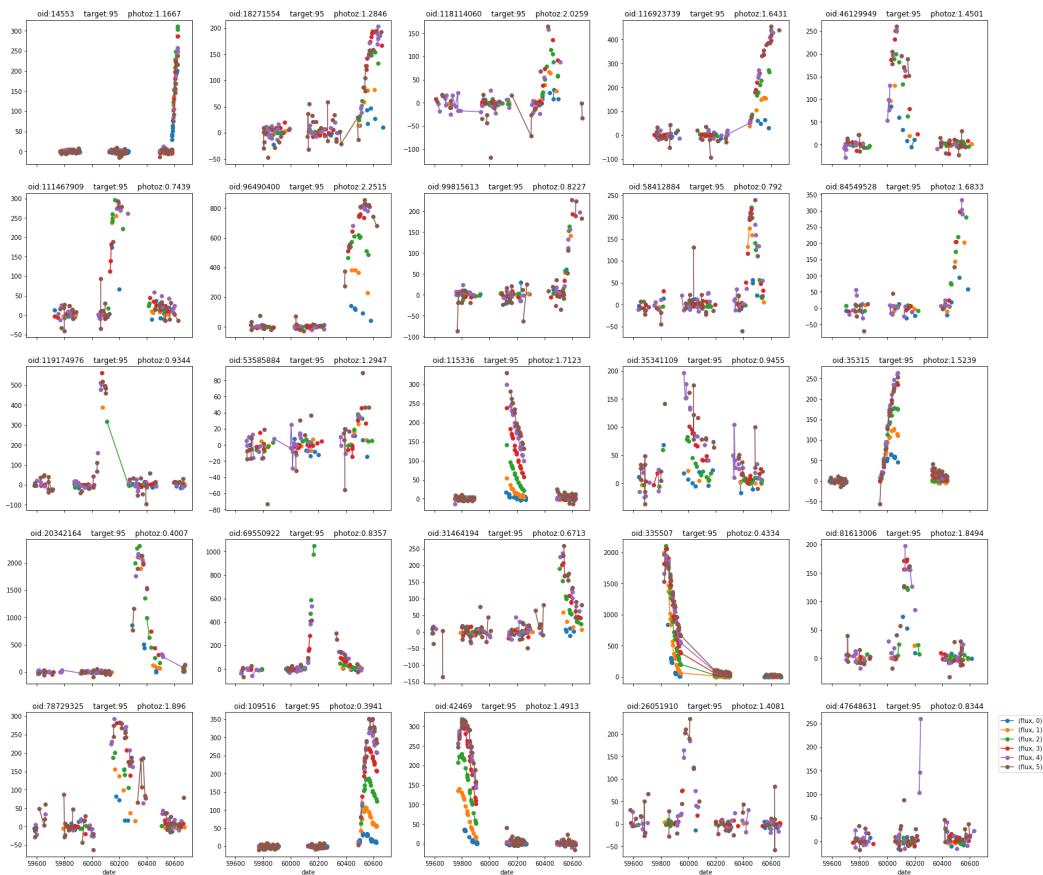
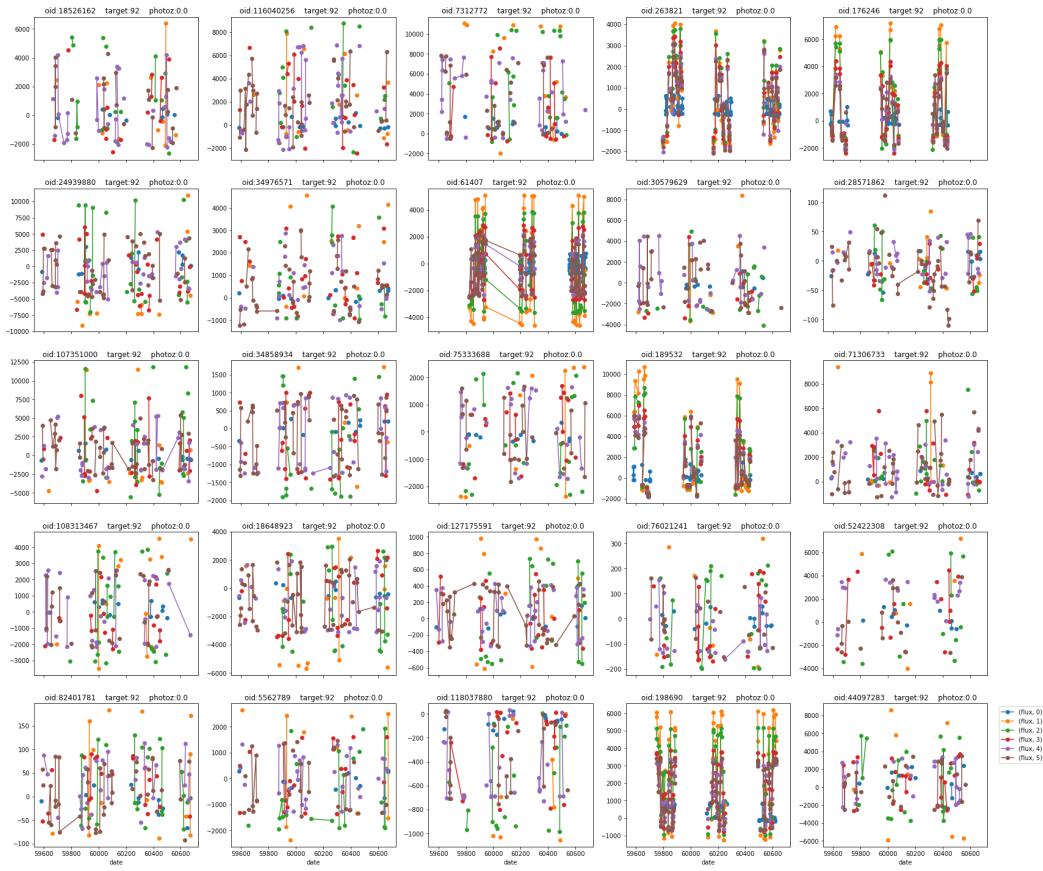




EDA_002



EDA_002



In []: