

In [63]:

```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
import seaborn as sns
%matplotlib inline
%pprint off
#plt.style.use('seaborn')
```

Pretty printing has been turned OFF

In [64]:

```
sns.set_style("ticks")
```

In [65]:

```
df_all = pd.read_csv('data/cleaned.csv')
len(df_all)
```

Out[65]:

606293

In [66]:

```
df_all = df_all.set_index(pd.to_datetime(df_all['localminute']), drop=True)
```

In [67]:

```
df_all = df_all.drop(columns=['Unnamed: 0', 'localminute'])
```

In [68]:

```
display(df_all.head(), len(df_all))
```

	marginal_change	cumul_value	meterid
localminute			
2015-10-01 05:00:00	0.0	93470.0	35
2015-10-01 06:00:00	0.0	93470.0	35
2015-10-01 07:00:00	0.0	93470.0	35
2015-10-01 08:00:00	0.0	93470.0	35
2015-10-01 09:00:00	0.0	93470.0	35

606293

In [69]:

```
groups = df_all.groupby('meterid')
keys = groups.groups.keys() # keys: an iterable of dataids or meter ids
```

In [70]:

```
id_list = list(keys)
display(id_list, len(id_list))
```

```
[35, 44, 77, 94, 114, 187, 222, 252, 370, 483, 484, 661, 739, 744, 871, 10
42, 1086, 1103, 1185, 1283, 1403, 1415, 1507, 1556, 1589, 1619, 1697, 171
4, 1718, 1790, 1791, 1792, 1800, 1801, 2018, 2034, 2072, 2094, 2129, 2233,
2335, 2378, 2449, 2461, 2470, 2575, 2638, 2645, 2755, 2814, 2818, 2945, 29
46, 2965, 2980, 3036, 3039, 3134, 3310, 3367, 3527, 3544, 3577, 3635, 372
3, 3778, 3849, 3893, 3918, 4029, 4031, 4193, 4228, 4296, 4352, 4356, 4373,
4421, 4447, 4514, 4671, 4732, 4767, 4874, 4998, 5129, 5131, 5193, 5275, 53
17, 5395, 5403, 5439, 5484, 5545, 5636, 5658, 5785, 5810, 5814, 5892, 597
2, 6101, 6412, 6505, 6578, 6673, 6685, 6830, 6836, 6863, 6910, 7016, 7017,
7030, 7117, 7287, 7429, 7460, 7566, 7674, 7682, 7739, 7741, 7794, 7900, 79
19, 7965, 7989, 8059, 8084, 8086, 8155, 8156, 8244, 8386, 8467, 8703, 882
9, 8890, 8967, 9052, 9121, 9134, 9160, 9278, 9295, 9474, 9600, 9620, 9631,
9639, 9729, 9766, 9849, 9956, 9982]
```

157

In [81]:

```
for i in id_list:
    df_i = groups.get_group(i)
    #df_i['cumul_value'].plot(figsize=(15,4), title=f'meter {i}, {len(df_i)} samples')
    plt.cla()
    fig = plt.gcf()
    fig.set_size_inches(15,4)
    plt.title(f'meter {i}, {len(df_i)} samples')
    plt.scatter(x=df_i.index, y=df_i['cumul_value'])
    plt.show()
```































































