

In [3]:

```
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 ON

In [4]:

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

Out[4]:

593956

In [5]:

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

In [6]:

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

In [7]:

```
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

593956

In [8]:

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

In [9]:

```
id_list = list(keys)
id_list.remove(8703) # meterid 8703 was found to have data for only 2 weeks in Oct.
display(id_list, len(id_list))
```

```
[35,  
 44,  
 77,  
 94,  
 114,  
 187,  
 222,  
 252,  
 370,  
 483,  
 484,  
 661,  
 739,  
 744,  
 871,  
 1042,  
 1086,  
 1103,  
 1185,  
 1283,  
 1403,  
 1415,  
 1507,  
 1556,  
 1589,  
 1619,  
 1697,  
 1714,  
 1718,  
 1790,  
 1791,  
 1792,  
 1800,  
 1801,  
 2018,  
 2034,  
 2072,  
 2094,  
 2129,  
 2233,  
 2335,  
 2378,  
 2449,  
 2461,  
 2470,  
 2575,  
 2638,  
 2755,  
 2814,  
 2818,  
 2945,  
 2946,  
 2965,  
 2980,  
 3036,  
 3039,  
 3134,  
 3310,  
 3367,  
 3527,  
 3544,
```

3577,
3635,
3723,
3778,
3849,
3893,
3918,
4029,
4031,
4193,
4228,
4296,
4352,
4356,
4373,
4421,
4447,
4514,
4732,
4767,
4998,
5129,
5131,
5193,
5275,
5317,
5395,
5403,
5439,
5484,
5636,
5658,
5785,
5810,
5814,
5892,
5972,
6412,
6505,
6578,
6673,
6685,
6830,
6836,
6863,
6910,
7016,
7017,
7030,
7117,
7287,
7429,
7460,
7674,
7682,
7739,
7741,
7794,
7900,
7919,
7965,

```
7989,  
8059,  
8084,  
8086,  
8155,  
8156,  
8244,  
8386,  
8467,  
8829,  
8890,  
8967,  
9052,  
9121,  
9134,  
9160,  
9278,  
9295,  
9474,  
9600,  
9631,  
9639,  
9729,  
9766,  
9849,  
9956,  
9982]
```

149

Visualize meter data.

In [10]:

```
# 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()
```

In [11]:

```
# get list of meterids whose first datapoint is at least as early as pre-specified 'origin_date'.
# and whose last datapoint is after some specified 'end_date'?
# the reason is to get a set of data that can be normalized and then fed as regression
# input.

origin_date = '2015-11-01'

valid_list = []
for meterid in id_list:
    df_i = groups.get_group(meterid)
    if df_i.index[0] <= pd.to_datetime(origin_date):
        valid_list.append(meterid)

display(valid_list, len(valid_list))
```

```
[35,  
 44,  
 77,  
 94,  
 114,  
 187,  
 222,  
 252,  
 370,  
 483,  
 484,  
 661,  
 739,  
 744,  
 871,  
 1042,  
 1086,  
 1103,  
 1185,  
 1283,  
 1415,  
 1507,  
 1556,  
 1589,  
 1619,  
 1697,  
 1714,  
 1718,  
 1790,  
 1791,  
 1792,  
 1800,  
 1801,  
 2018,  
 2034,  
 2072,  
 2094,  
 2129,  
 2233,  
 2335,  
 2378,  
 2449,  
 2461,  
 2470,  
 2575,  
 2638,  
 2818,  
 2945,  
 2965,  
 2980,  
 3039,  
 3134,  
 3310,  
 3367,  
 3527,  
 3544,  
 3577,  
 3635,  
 3723,  
 3778,  
 3849,
```

3893,
3918,
4029,
4031,
4193,
4228,
4296,
4352,
4356,
4373,
4421,
4447,
4514,
4732,
4767,
4998,
5129,
5131,
5193,
5275,
5395,
5403,
5439,
5484,
5636,
5785,
5810,
5814,
5892,
5972,
6412,
6505,
6578,
6685,
6830,
6836,
6863,
6910,
7016,
7017,
7030,
7117,
7287,
7429,
7460,
7674,
7682,
7739,
7741,
7794,
7900,
7919,
7965,
7989,
8059,
8084,
8086,
8155,
8156,
8386,
8467,

```
8829,  
8890,  
8967,  
9052,  
9121,  
9134,  
9278,  
9295,  
9474,  
9631,  
9639,  
9729,  
9766,  
9849,  
9956,  
9982]
```

138

In [12]:

```
# normalize the cumulative values before fitting the model over the values!!

def normalize_cumul(df, start_date):
    """
    normalize the cumul values towards a common start_date.

    df: dataframe. dataframe representing one meterid's data. df must have datetimeindex,
    and 'cumul_value' column
    """
    df1 = df.loc[df.index >= pd.to_datetime(start_date)]
    const = df1['cumul_value'].iloc[0] # get first cumul value.
    ser_norm = df1['cumul_value'] - const
    df_new = df1.assign(norm_cumul_value=ser_norm).drop(columns=['cumul_value'])

    return df_new

df_eg = normalize_cumul(df_i, origin_date)
display(df_eg.head(10), df_eg.tail(10))
```

	marginal_change	meterid	norm_cumul_value
localminute			
2015-11-01 00:00:00	4.0	9982	0.0
2015-11-01 01:00:00	0.0	9982	0.0
2015-11-01 02:00:00	0.0	9982	0.0
2015-11-01 03:00:00	0.0	9982	0.0
2015-11-01 04:00:00	8.0	9982	8.0
2015-11-01 05:00:00	0.0	9982	8.0
2015-11-01 06:00:00	0.0	9982	8.0
2015-11-01 07:00:00	0.0	9982	8.0
2015-11-01 08:00:00	0.0	9982	8.0
2015-11-01 09:00:00	0.0	9982	8.0

	marginal_change	meterid	norm_cumul_value
localminute			
2016-03-31 07:00:00	0.0	9982	12574.0
2016-03-31 08:00:00	0.0	9982	12574.0
2016-03-31 09:00:00	0.0	9982	12574.0
2016-03-31 10:00:00	0.0	9982	12574.0
2016-03-31 11:00:00	0.0	9982	12574.0
2016-03-31 12:00:00	0.0	9982	12574.0
2016-03-31 13:00:00	0.0	9982	12574.0
2016-03-31 14:00:00	0.0	9982	12574.0
2016-03-31 15:00:00	0.0	9982	12574.0
2016-03-31 16:00:00	72.0	9982	12646.0

Aggregate/combined Linear Regression model (stuck here).

In [13]:

```
# combine all meterids' data.
# maybe create new columns for 1 hot encoding indicating which meterid the datapoint belongs to
# it is possible to extend this model by using one-hot encoding to indicate a particular cluster of meterids instead.

def normalize_and_combine(origin_date):
    """
    finds all meterids that have a datapoint at origin_date and
    normalizes all cumulative readings based on that origin_date by
    subtracting the first cumul value from all cumul values for each meterid.
    effectively setting the first cumul value to 0 at origin_date, and all subsequent
    values are based off that.

    ...
    valid_list = []
    for meterid in id_list:
        df_i = groups.get_group(meterid)
        if df_i.index[0] <= pd.to_datetime(origin_date):
            valid_list.append(meterid)

    df_combinedxy = pd.DataFrame()
    for meterid in valid_list:
        df_i = groups.get_group(meterid)
        df_i = normalize_cumul(df_i, origin_date)
        df_i = df_i.reset_index()
        df_combinedxy = df_combinedxy.append(df_i, ignore_index=True)

    return df_combinedxy

df_normalized = normalize_and_combine(origin_date)
display(df_normalized.head(), df_normalized.tail(), df_normalized.describe())
```

	localminute	marginal_change	meterid	norm_cumul_value
0	2015-11-01 00:00:00	0.0	35	0.0
1	2015-11-01 01:00:00	12.0	35	12.0
2	2015-11-01 02:00:00	0.0	35	12.0
3	2015-11-01 03:00:00	2.0	35	14.0
4	2015-11-01 04:00:00	0.0	35	14.0

	localminute	marginal_change	meterid	norm_cumul_value
490048	2016-03-31 12:00:00	0.0	9982	12574.0
490049	2016-03-31 13:00:00	0.0	9982	12574.0
490050	2016-03-31 14:00:00	0.0	9982	12574.0
490051	2016-03-31 15:00:00	0.0	9982	12574.0
490052	2016-03-31 16:00:00	72.0	9982	12646.0

	marginal_change	meterid	norm_cumul_value
count	490053.000000	490053.000000	490053.000000
mean	5.109978	4633.380975	9437.259582
std	17.900346	2930.912281	8410.615437
min	0.000000	35.000000	0.000000
25%	0.000000	2034.000000	2548.000000
50%	0.000000	4352.000000	7658.000000
75%	2.000000	7287.000000	14500.000000
max	1640.000000	9982.000000	55700.000000

In [14]:

```
# Label encoding
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
labels = le.fit_transform(df_normalized['meterid'])
display(labels, len(labels))

df_normalized = df_normalized.assign(label=labels)
df_normalized.head()
```

```
array([ 0,  0,  0, ..., 137, 137, 137], dtype=int64)
```

```
490053
```

Out[14]:

	localminute	marginal_change	meterid	norm_cumul_value	label
0	2015-11-01 00:00:00	0.0	35	0.0	0
1	2015-11-01 01:00:00	12.0	35	12.0	0
2	2015-11-01 02:00:00	0.0	35	12.0	0
3	2015-11-01 03:00:00	2.0	35	14.0	0
4	2015-11-01 04:00:00	0.0	35	14.0	0

In [15]:

```
from sklearn.linear_model import LinearRegression
```

In [16]:

```
# split dataset into train, val, test (how?)
# normalize cumul values?
# Lr_model.fit(x, y)
# score on train set
# score on test set
# plot timeseries of actual value and timeseries of predicted value on same plot
# is this for each meter? or how do we generate a set of actual values representing all
# meters?
# plot predicted value vs actual value. plot predicted value = actual value. analyse va
riance and bias.

"""
if predicting an average consumption pattern, then x and y axis must be normalized.
maybe normalize each meterid's y data by subtracting all cumul_values by the first cumu
l_value.
the model then predicts the change from the first available datapoint.
!UNCLEAR how to normalize x axis.
"""

"""
since different households (HH) have different consumption patterns, with some HH's con
sumption patterns being quite
correlated with other HHs, it might be good to fit different models for each of these
'clusters'.
That way, we can get higher prediction accuracy for specific HHs in each cluster.
Can we clearly identify different clusters from the corr analysis done in Part 1?
"""

# repeat above using SVR, tune parameters with validation dataset.
```

Out[16]:

"\nsince different households (HH) have different consumption patterns, wi
th some HH's consumption patterns being quite \ncorrelated with other HHs,
it might be good to fit different models for each of these 'clusters'.\\nTh
at way, we can get higher prediction accuracy for specific HHs in each clu
ster.\\nCan we clearly identify different clusters from the corr analysis d
one in Part 1?\n"

Linear Regression model for one meterid.

In [17]:

```

meterid = 44
df_i = groups.get_group(meterid)
df_i.head()

```

Out[17]:

	marginal_change	cumul_value	meterid
localminute			
2015-10-12 23:00:00	0.0	165674.0	44
2015-10-13 00:00:00	0.0	165674.0	44
2015-10-13 01:00:00	0.0	165674.0	44
2015-10-13 02:00:00	10.0	165684.0	44
2015-10-13 03:00:00	0.0	165684.0	44

In [18]:

```

lr_model = LinearRegression()

# convert DateTimeIndex to numerical values for regression.
# each hour is now denoted by an integer.
X = np.arange(start=0, stop=len(df_i.index), step=1).reshape(-1, 1)

y = df_i['cumul_value']
# should we shift y axis down by first cumul value? more meaningful? can more easily compare across meterids?

```

In [19]:

```

# split dataset

valid_start, test_start = int(len(df_i)*0.7), int(len(df_i)*0.8)
x_train, x_valid, x_test = X[:valid_start,:], X[valid_start:test_start,:], X[test_start:,:]
y_train, y_valid, y_test = y[:valid_start], y[valid_start:test_start], y[test_start:]

```

In [20]:

```
lr_model.fit(x_train, y_train)
```

Out[20]:

```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)
```

In [21]:

```
# run predictions and get scores

y_train_pred = lr_model.predict(x_train)
r2_train = lr_model.score(x_train, y_train)

y_valid_pred = lr_model.predict(x_valid)
r2_valid = lr_model.score(x_valid, y_valid)

y_test_pred = lr_model.predict(x_test)
r2_test = lr_model.score(x_test, y_test)
```

In [22]:

```
# plot results

plt.title(f'meterid {meterid}; r2_train: {r2_train:.3f}; r2_val: {r2_valid:.3f}; r2_test: {r2_test:.3f}')
plt.xlabel(f'hours elapsed since {df_i.index[0]}')
plt.plot(x_train, y_train, label='actual train')
plt.plot(x_train, y_train_pred, label='prediction on train')

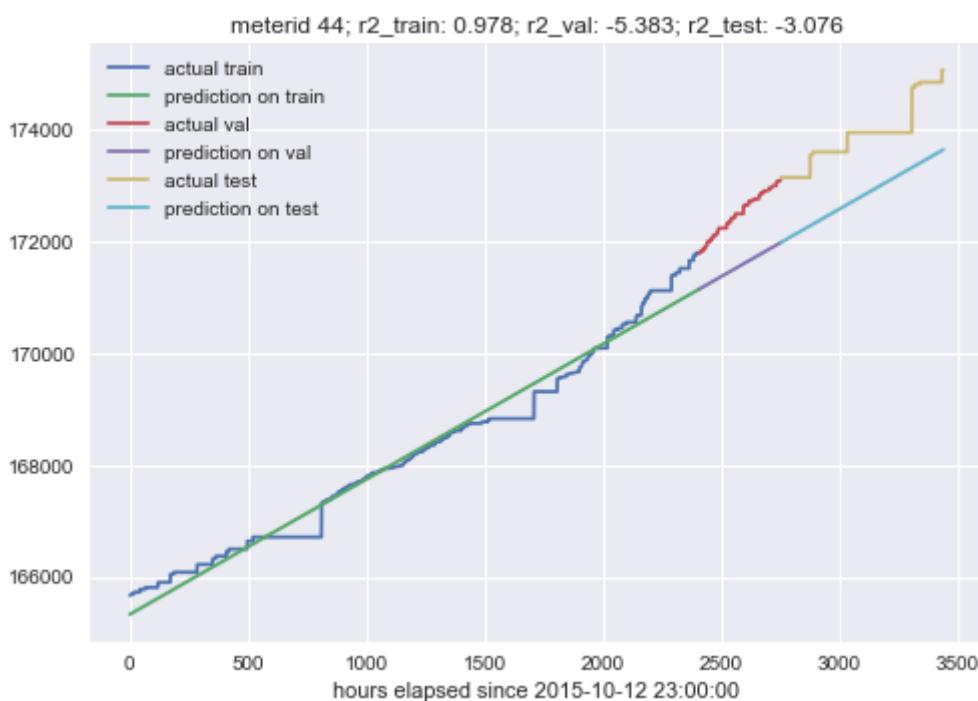
plt.plot(x_valid, y_valid, label='actual val')
plt.plot(x_valid, y_valid_pred, label='prediction on val')

plt.plot(x_test, y_test, label='actual test')
plt.plot(x_test, y_test_pred, label='prediction on test')

plt.legend()
```

Out[22]:

```
<matplotlib.legend.Legend at 0x22349af8080>
```



In [23]:

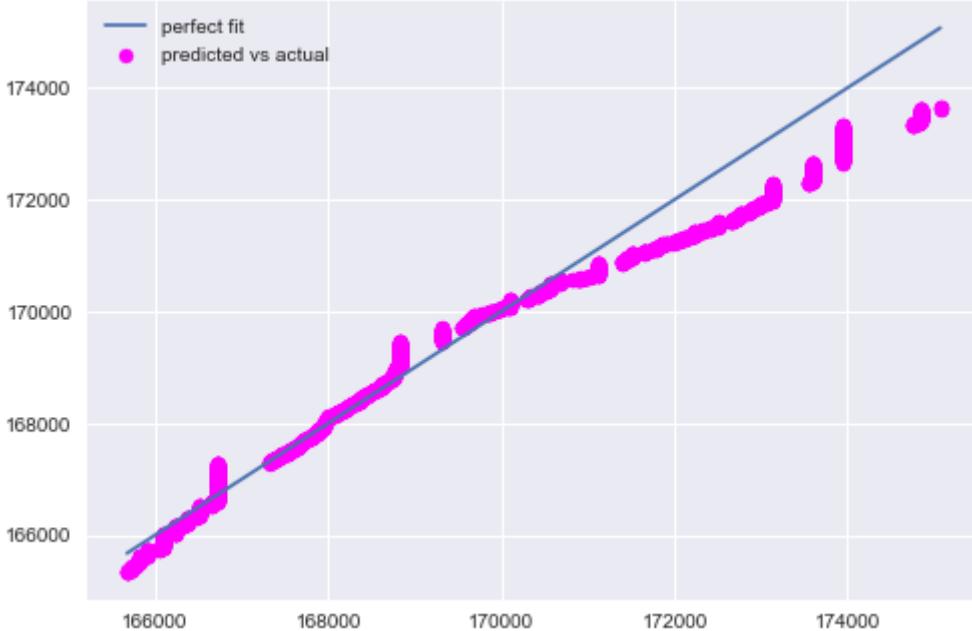
```
# plot actual vs predicted value
actual = []
actual.extend(y_train)
actual.extend(y_valid)
actual.extend(y_test)
len(actual)

predicted = []
predicted.extend(y_train_pred)
predicted.extend(y_valid_pred)
predicted.extend(y_test_pred)
len(predicted)

plt.scatter(actual, predicted, color='magenta', label='predicted vs actual')
plt.plot(actual, actual, label='perfect fit')
plt.legend()
```

Out[23]:

<matplotlib.legend.Legend at 0x22349bde128>



Combined code

In [51]:

```
from joblib import dump, load

origin_date = '2015-11-01' # starting date to normalize x axis (time in hours) to.

df_normalized = normalize_and_combine(origin_date)
display(df_normalized.head(), df_normalized.tail(), df_normalized.describe())

le = LabelEncoder()
labels = le.fit_transform(df_normalized['meterid'])
df_normalized = df_normalized.assign(label=labels)
display(df_normalized.head())

new_groups = df_normalized.groupby('meterid')
new_keys = new_groups.groups.keys() # keys: an iterable of dataids or meter ids

new_id_list = list(new_keys)
display(len(new_id_list))

for meterid in new_id_list:
    df_i = new_groups.get_group(meterid)

    # convert DateTimeIndex to numerical values (ints) for regression.
    X = np.arange(start=0, stop=len(df_i.index), step=1).reshape(-1, 1)
    y = df_i['norm_cumul_value']

    # split dataset
    valid_start, test_start = int(len(df_i)*0.8), int(len(df_i)*0.9)
    x_train, x_valid, x_test = X[:valid_start,:], X[valid_start:test_start,:], X[test_start:,:]
    y_train, y_valid, y_test = y[:valid_start], y[valid_start:test_start], y[test_start:,:]

    lr_model.fit(x_train, y_train)
    dump(lr_model, f'models/lr_{meterid}.joblib') # save model to hard disk.

    # run predictions and get scores
    y_train_pred = lr_model.predict(x_train)
    r2_train = lr_model.score(x_train, y_train)

    y_valid_pred = lr_model.predict(x_valid)
    r2_valid = lr_model.score(x_valid, y_valid)

    y_test_pred = lr_model.predict(x_test)
    r2_test = lr_model.score(x_test, y_test)

    # plot line demarcating end of training data.
    test_mark = X[valid_start][0]
    plt.plot([test_mark, test_mark], [y_train_pred.min(), y_test_pred.max()], 'k-')

    # plot trendline and actual line.
    plt.title(f'meterid {meterid}; r2_train: {r2_train:.3f}; r2_val: {r2_valid:.3f}; r2_test: {r2_test:.3f}')
    plt.xlabel(f'hours elapsed since {df_i.localminute.iloc[0]}')
    plt.plot(x_train, y_train, label='actual train')
    plt.plot(x_train, y_train_pred, label='prediction on train')

    plt.plot(x_valid, y_valid, label='actual val')
    plt.plot(x_valid, y_valid_pred, label='prediction on val')
```

```
plt.plot(x_test, y_test, label='actual test')
plt.plot(x_test, y_test_pred, label='prediction on test')

plt.legend()
plt.show()
plt.cla()

# plot prediction error plot (actual vs predicted value)
actual = []
actual.extend(y_train)
actual.extend(y_valid)
actual.extend(y_test)
len(actual)

predicted = []
predicted.extend(y_train_pred)
predicted.extend(y_valid_pred)
predicted.extend(y_test_pred)
len(predicted)

plt.scatter(actual, predicted, color='magenta', label='predicted vs actual')
plt.plot(actual, actual, label='perfect fit')
plt.legend()

plt.show()
plt.cla()
```

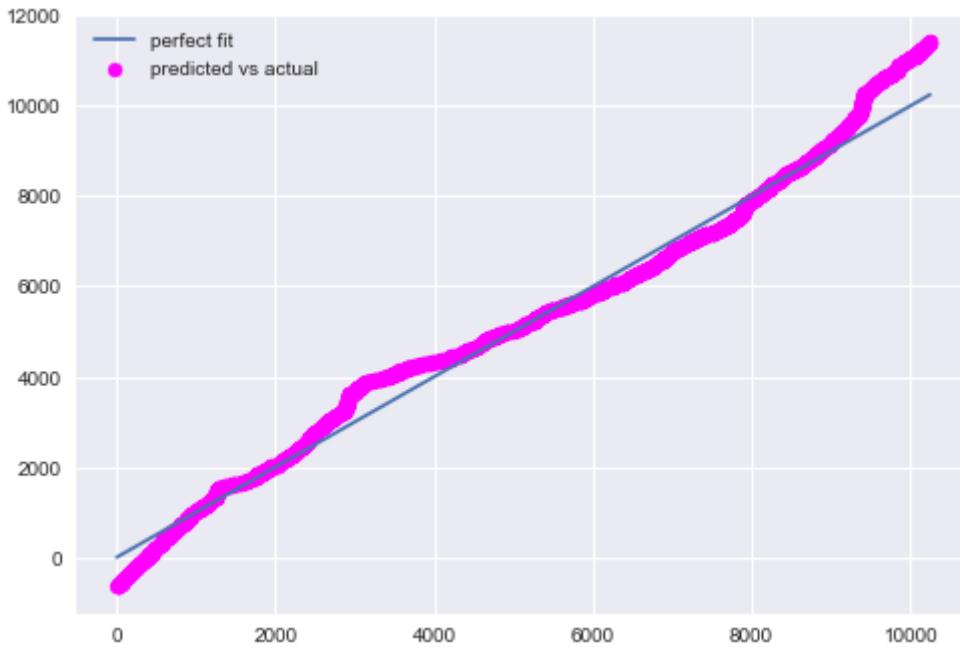
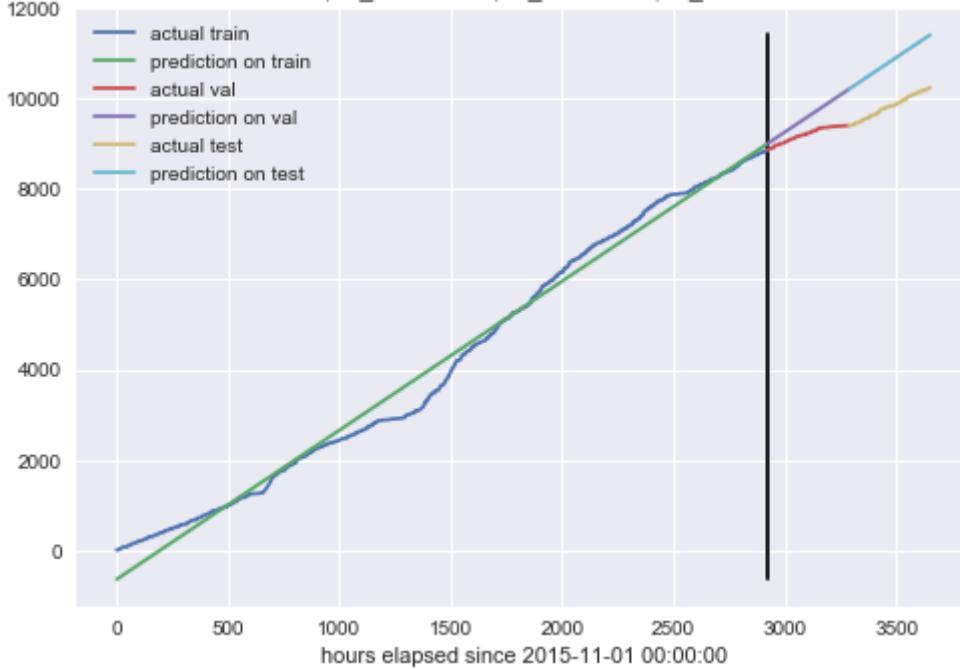
	localminute	marginal_change	meterid	norm_cumul_value
0	2015-11-01 00:00:00	0.0	35	0.0
1	2015-11-01 01:00:00	12.0	35	12.0
2	2015-11-01 02:00:00	0.0	35	12.0
3	2015-11-01 03:00:00	2.0	35	14.0
4	2015-11-01 04:00:00	0.0	35	14.0

	localminute	marginal_change	meterid	norm_cumul_value
490048	2016-03-31 12:00:00	0.0	9982	12574.0
490049	2016-03-31 13:00:00	0.0	9982	12574.0
490050	2016-03-31 14:00:00	0.0	9982	12574.0
490051	2016-03-31 15:00:00	0.0	9982	12574.0
490052	2016-03-31 16:00:00	72.0	9982	12646.0

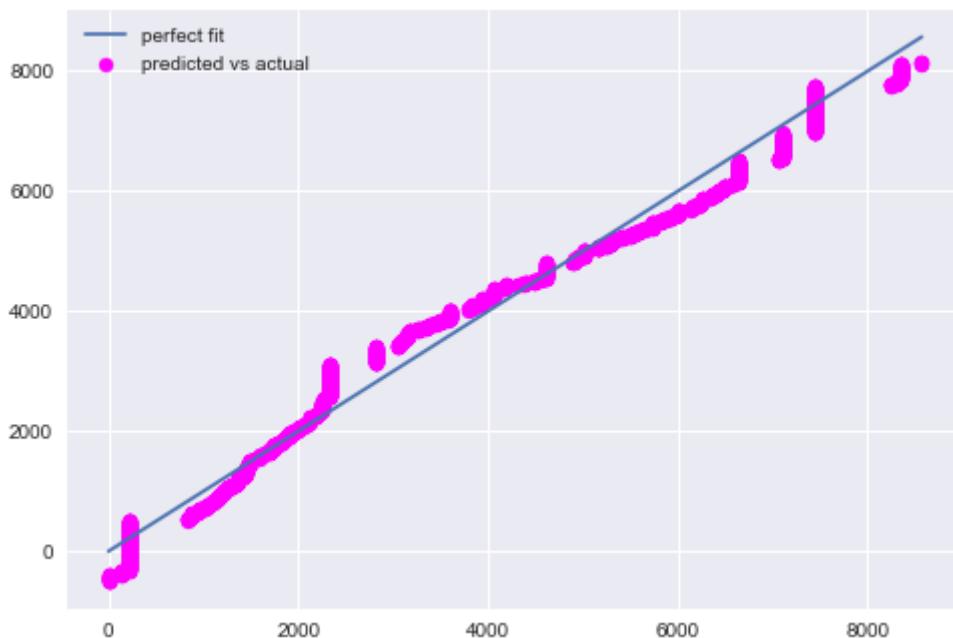
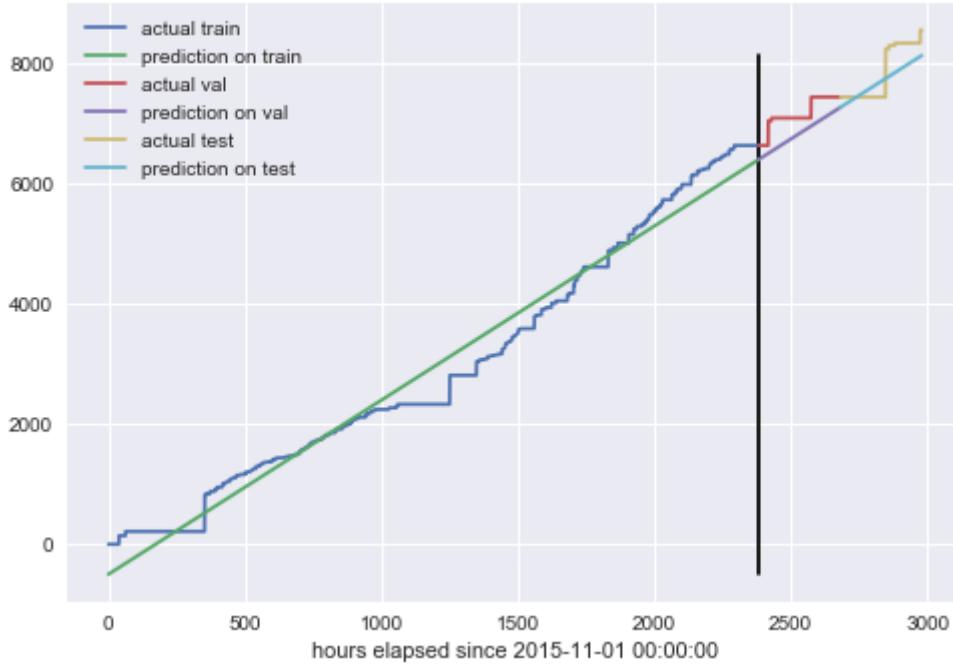
	marginal_change	meterid	norm_cumul_value
count	490053.000000	490053.000000	490053.000000
mean	5.109978	4633.380975	9437.259582
std	17.900346	2930.912281	8410.615437
min	0.000000	35.000000	0.000000
25%	0.000000	2034.000000	2548.000000
50%	0.000000	4352.000000	7658.000000
75%	2.000000	7287.000000	14500.000000
max	1640.000000	9982.000000	55700.000000

	localminute	marginal_change	meterid	norm_cumul_value	label
0	2015-11-01 00:00:00	0.0	35	0.0	0
1	2015-11-01 01:00:00	12.0	35	12.0	0
2	2015-11-01 02:00:00	0.0	35	12.0	0
3	2015-11-01 03:00:00	2.0	35	14.0	0
4	2015-11-01 04:00:00	0.0	35	14.0	0

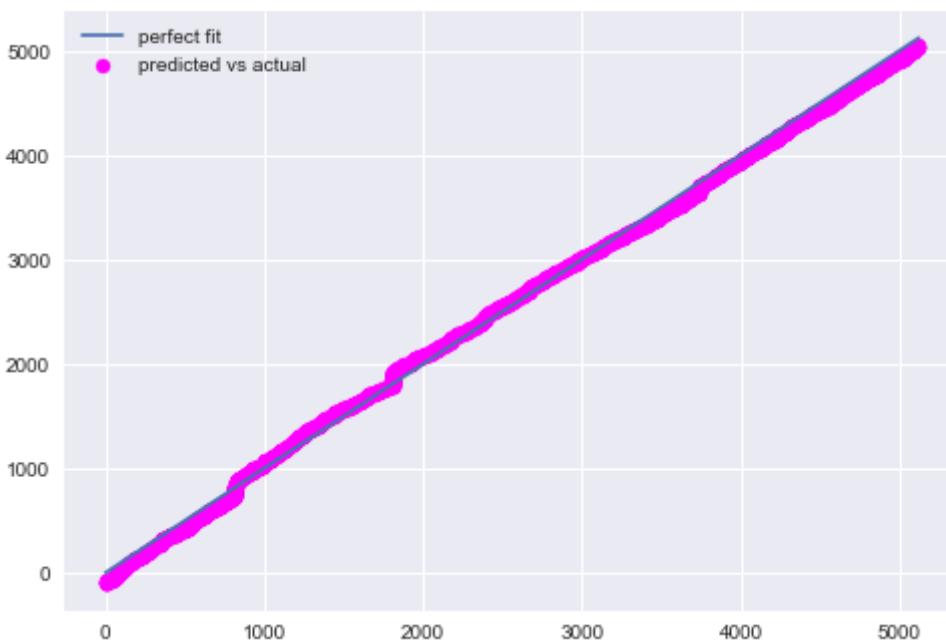
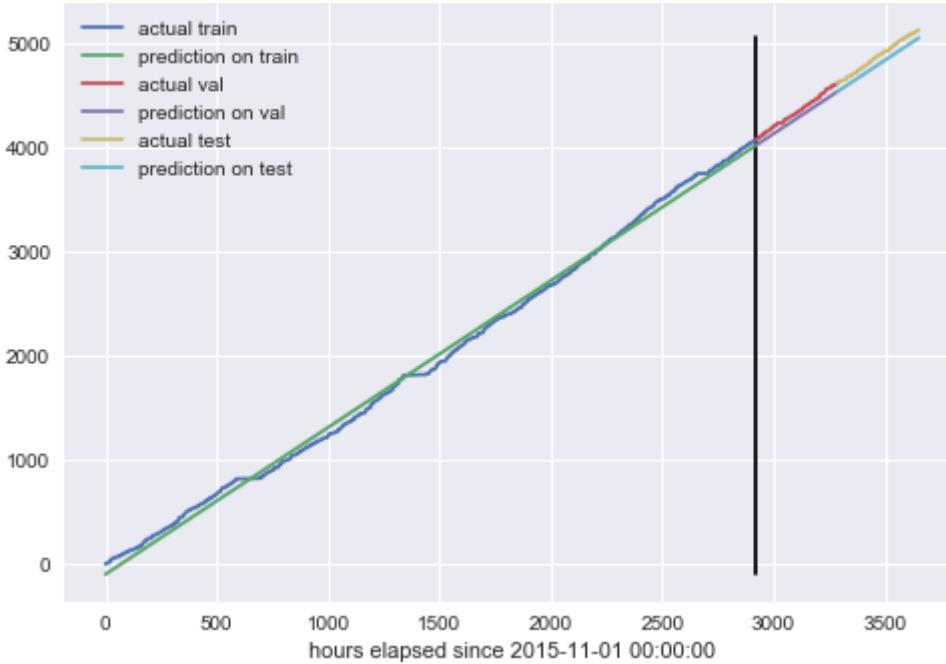
meterid 35; r2_train: 0.989; r2_val: -5.704; r2_test: -14.132



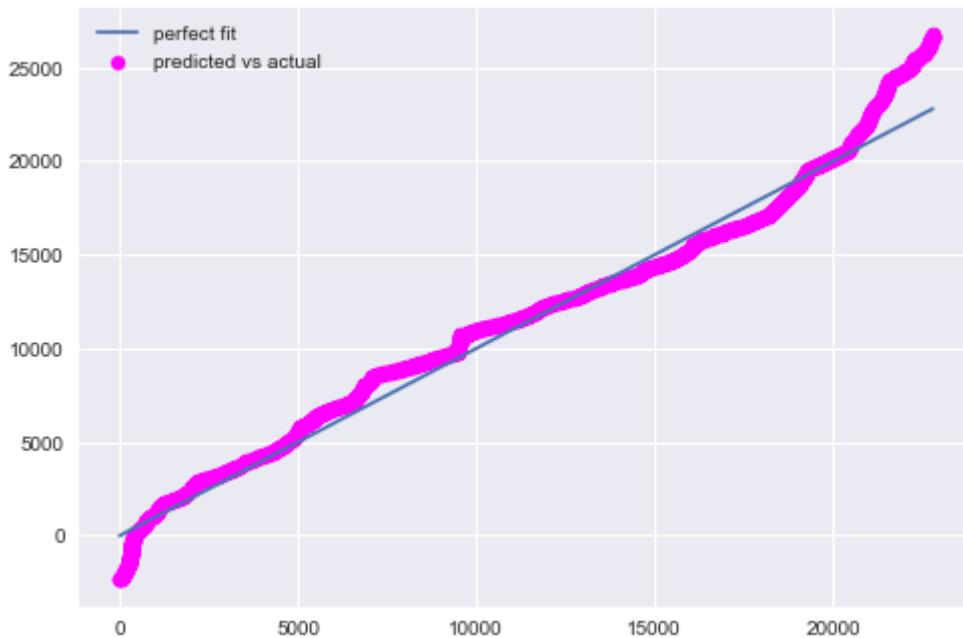
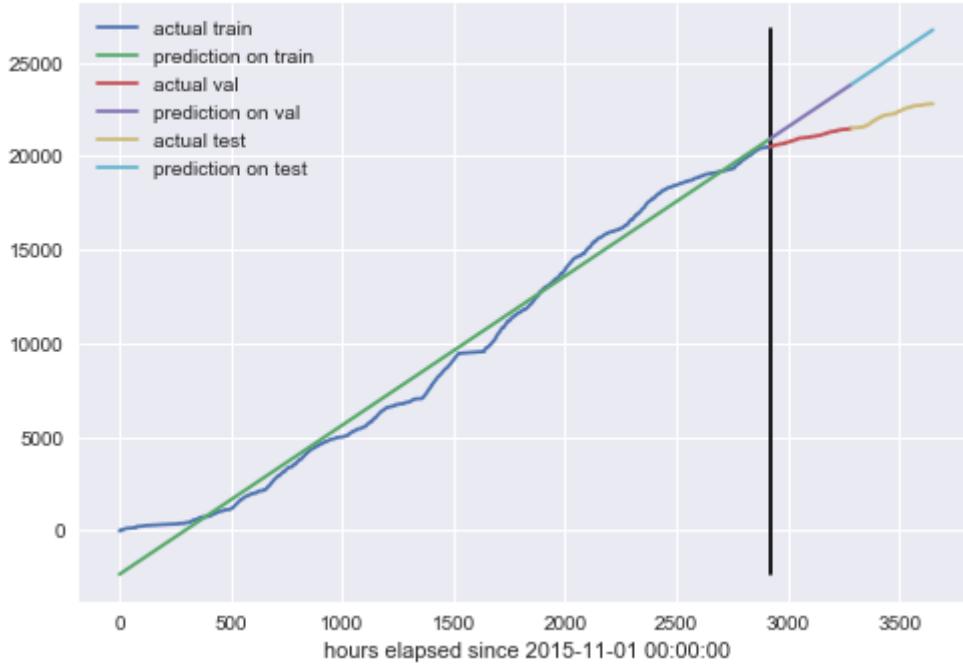
meterid 44; r2_train: 0.976; r2_val: -0.932; r2_test: 0.572



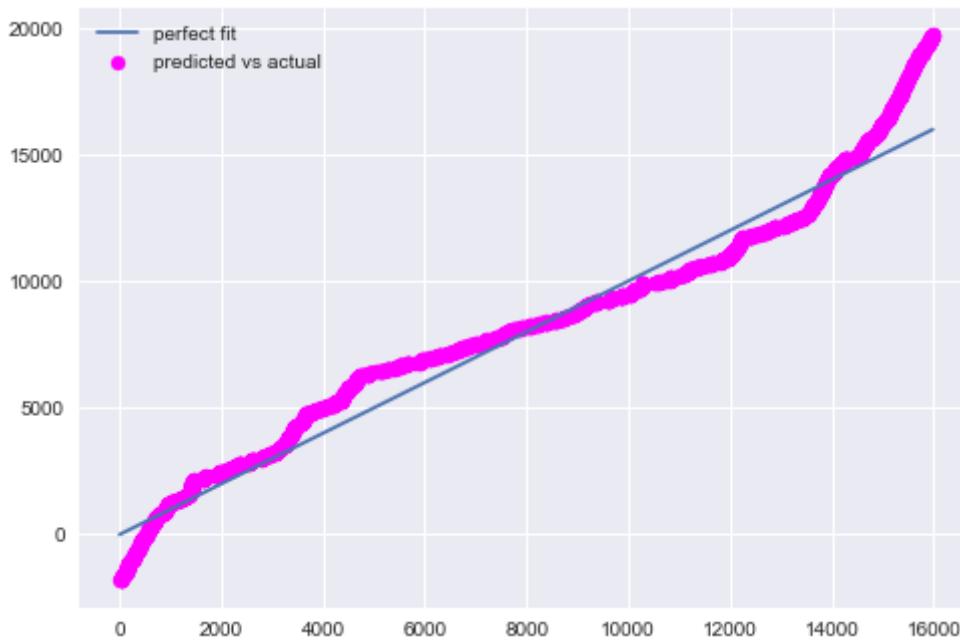
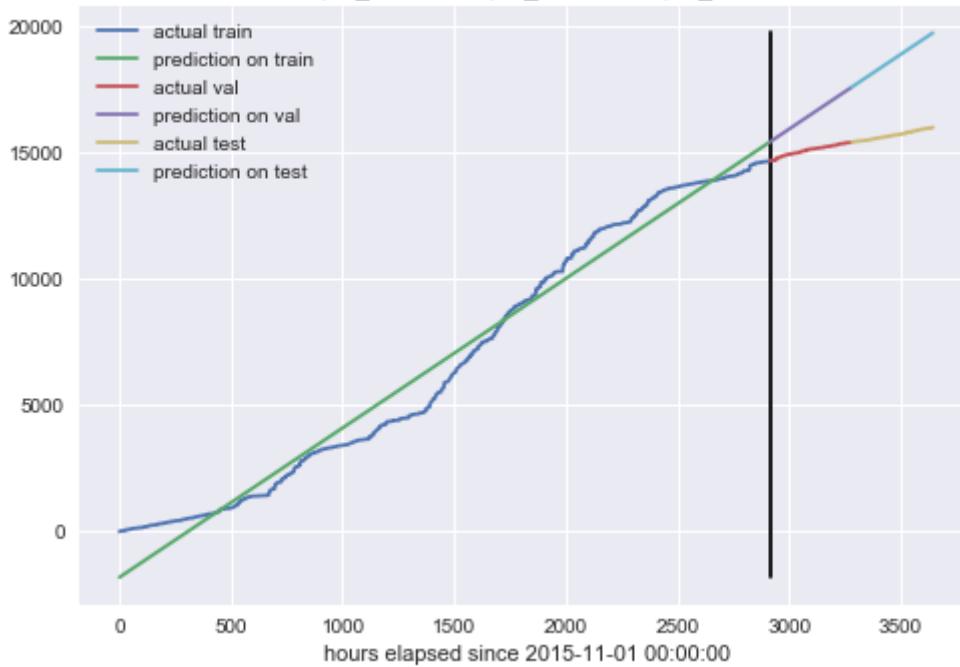
meterid 77; r2_train: 0.997; r2_val: 0.820; r2_test: 0.726



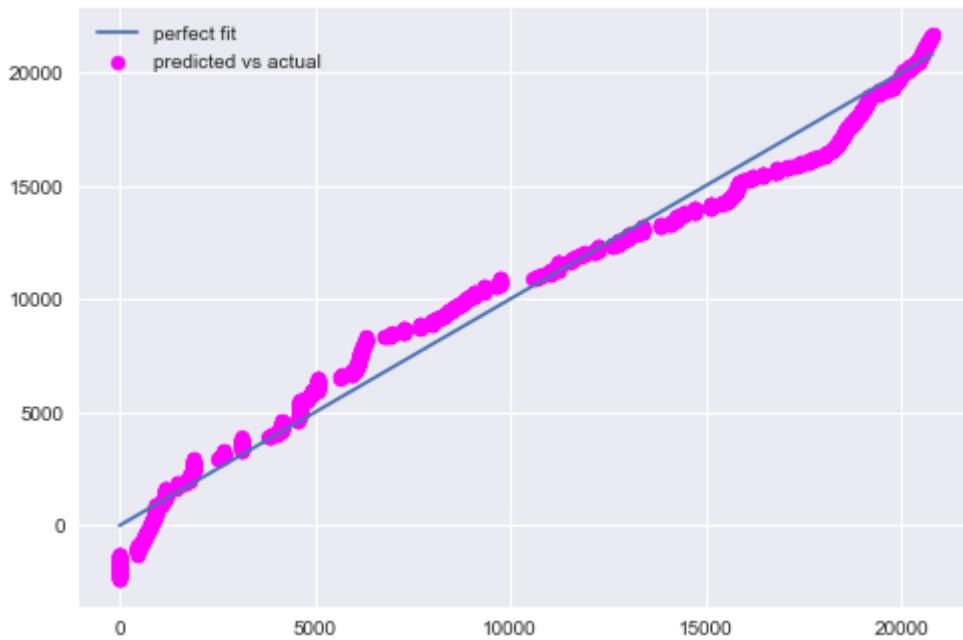
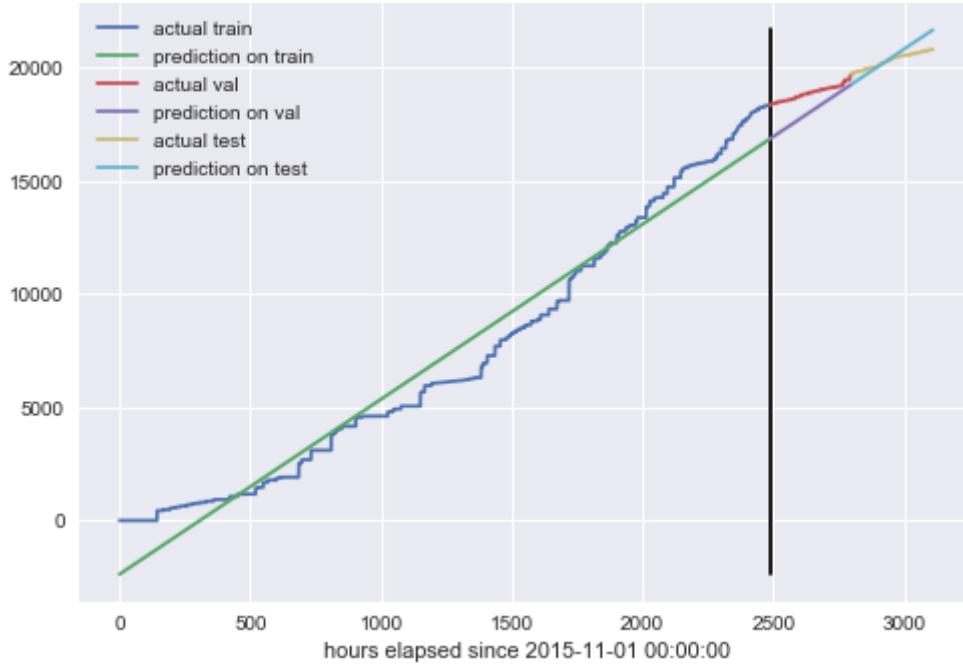
meterid 94; r2_train: 0.987; r2_val: -24.810; r2_test: -47.567



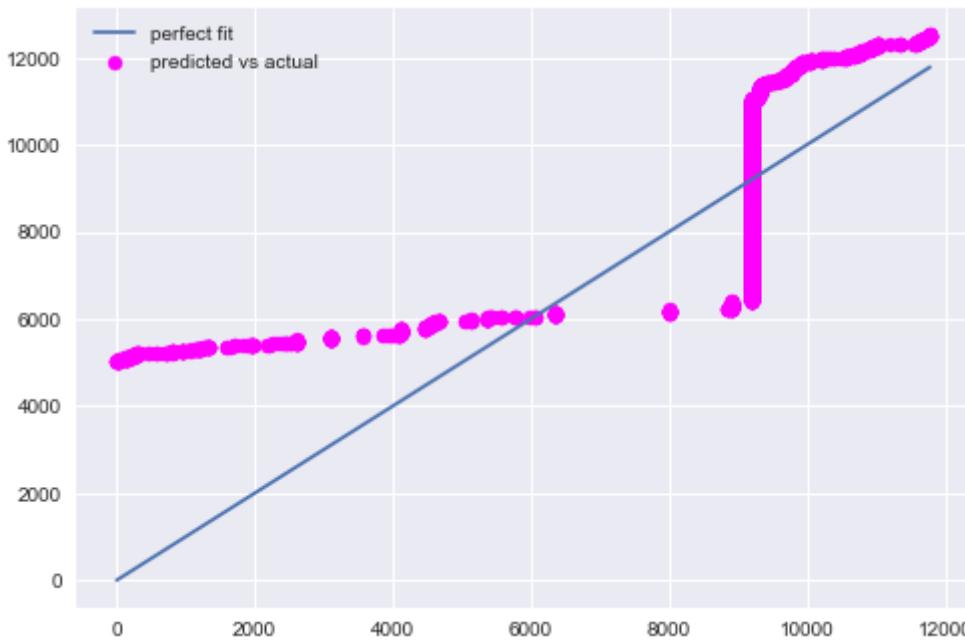
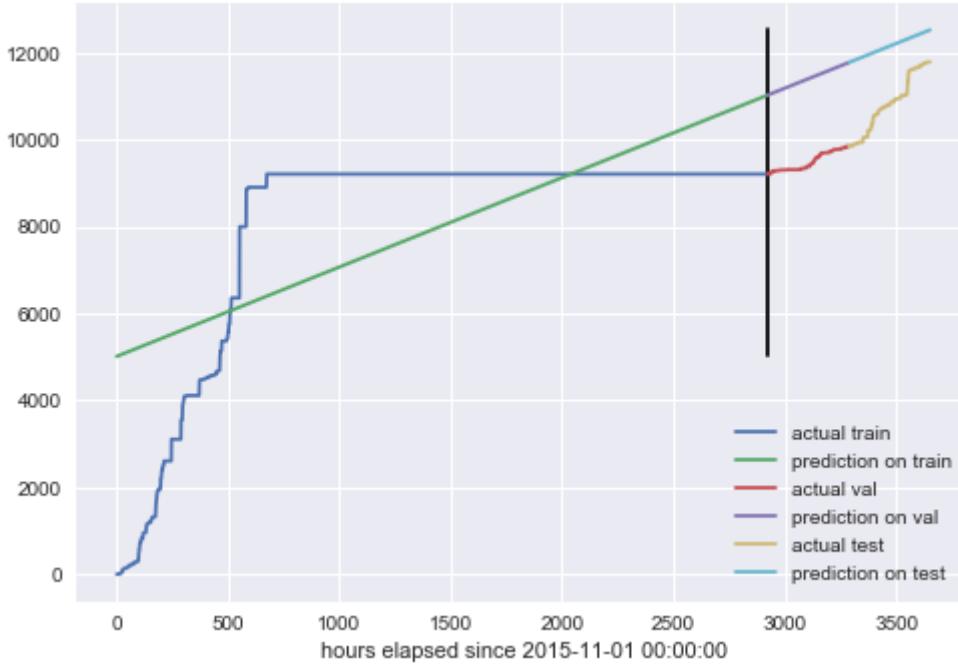
meterid 114; r2_train: 0.977; r2_val: -49.971; r2_test: -294.724



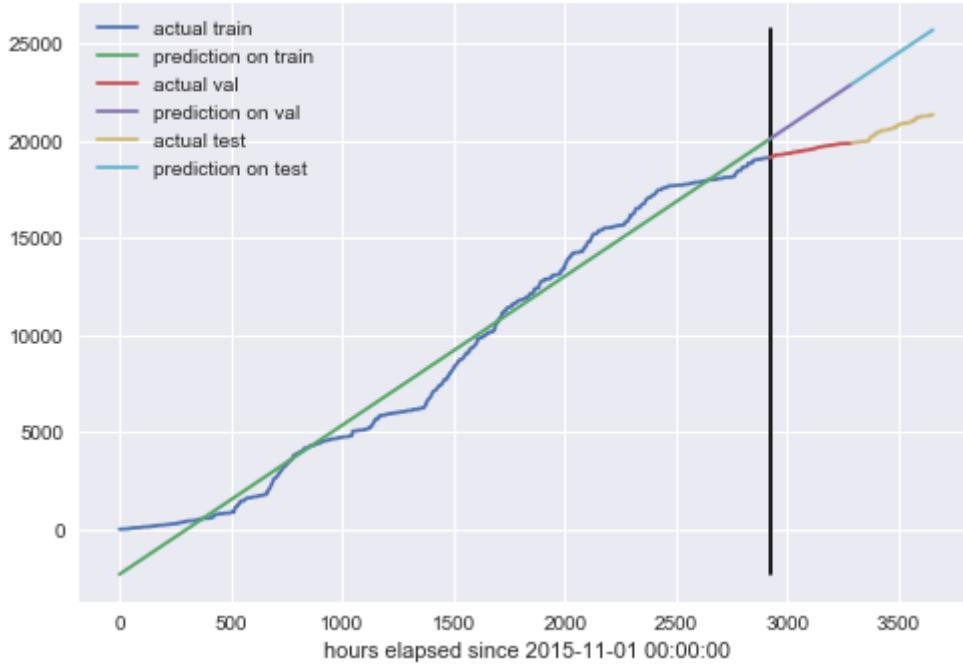
meterid 187; r2_train: 0.968; r2_val: -7.240; r2_test: -0.427

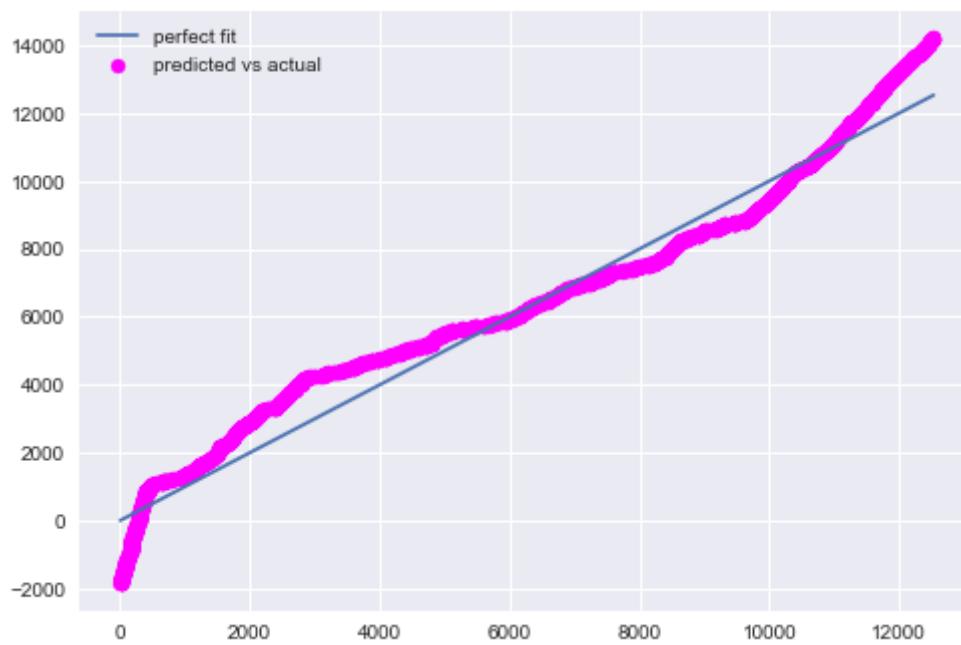
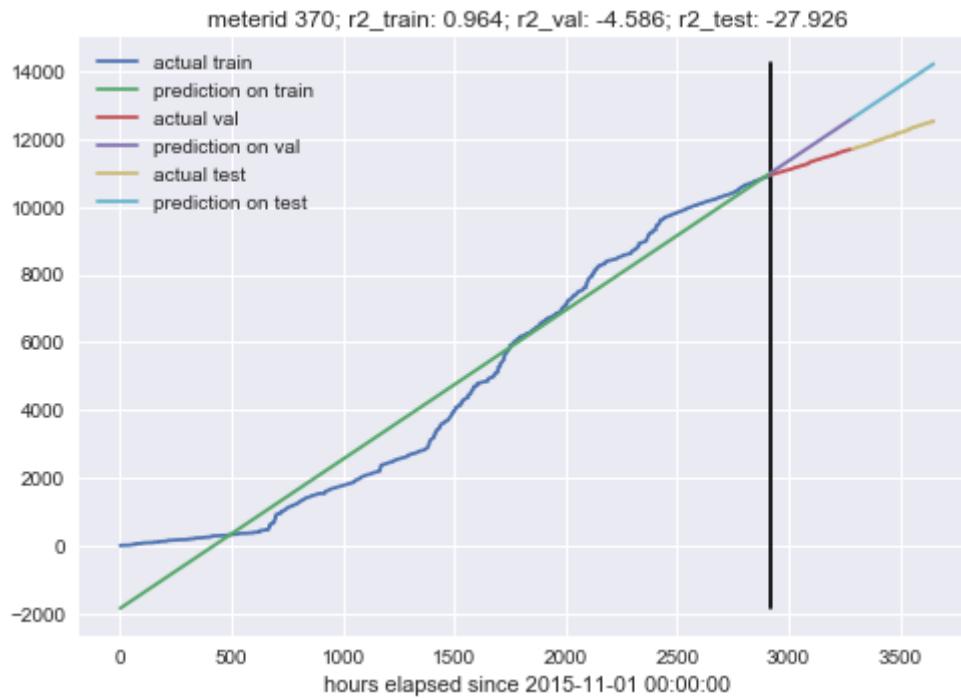


meterid 222; r2_train: 0.457; r2_val: -78.637; r2_test: -3.668

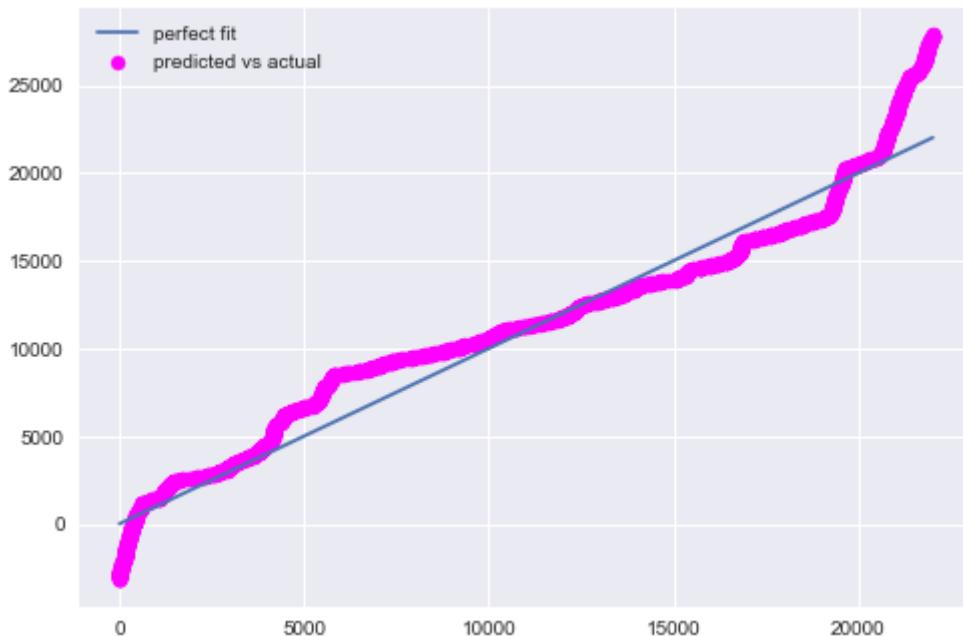
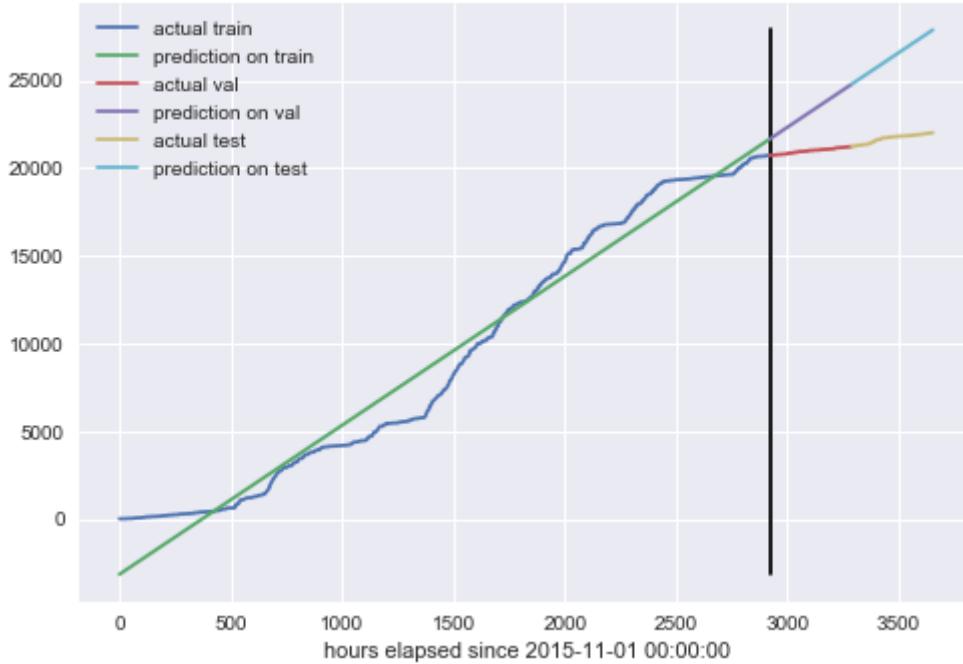


meterid 252; r2_train: 0.982; r2_val: -84.764; r2_test: -59.436

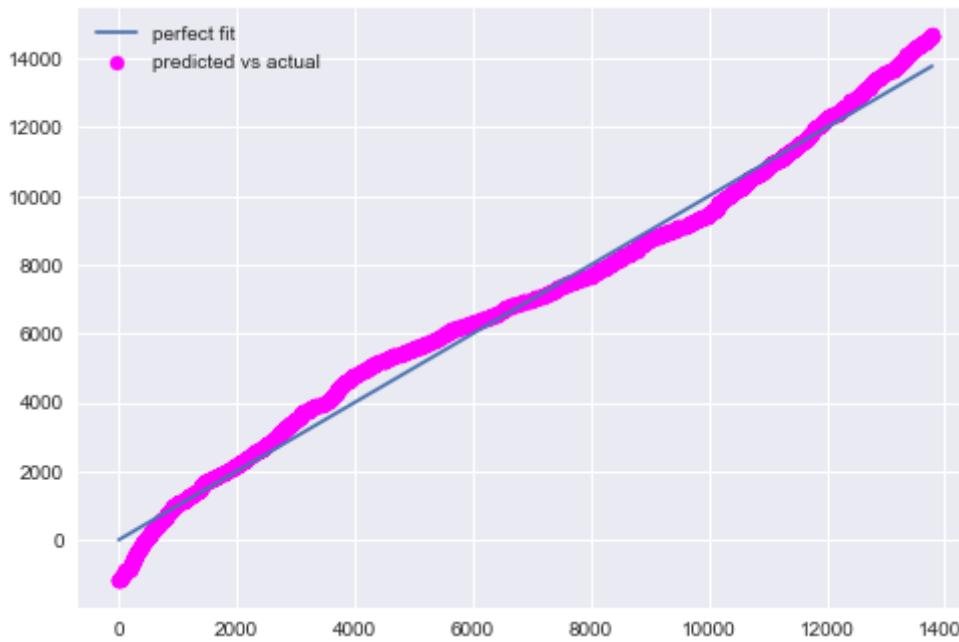
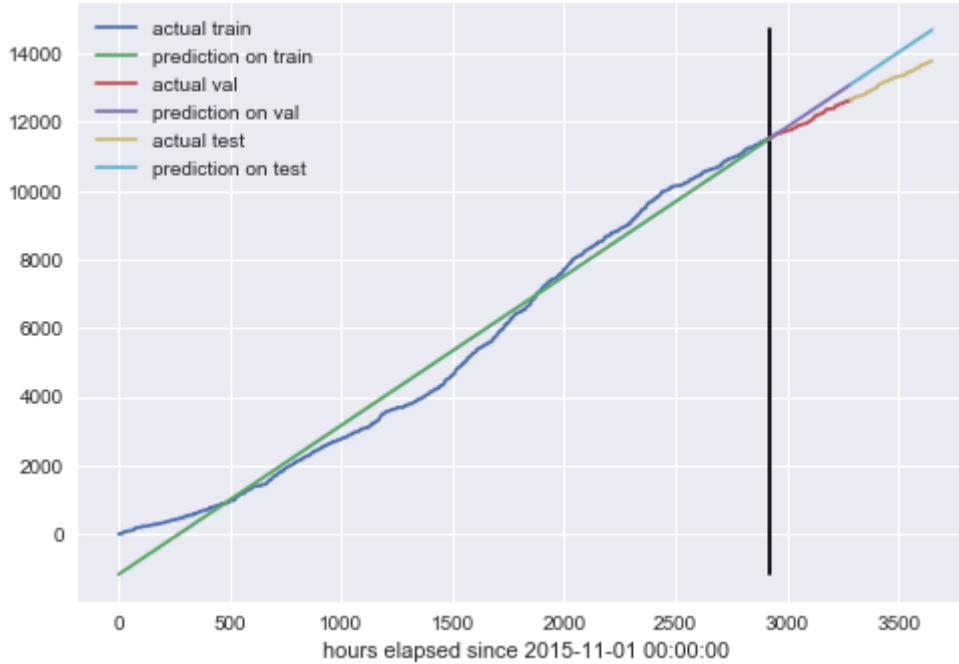




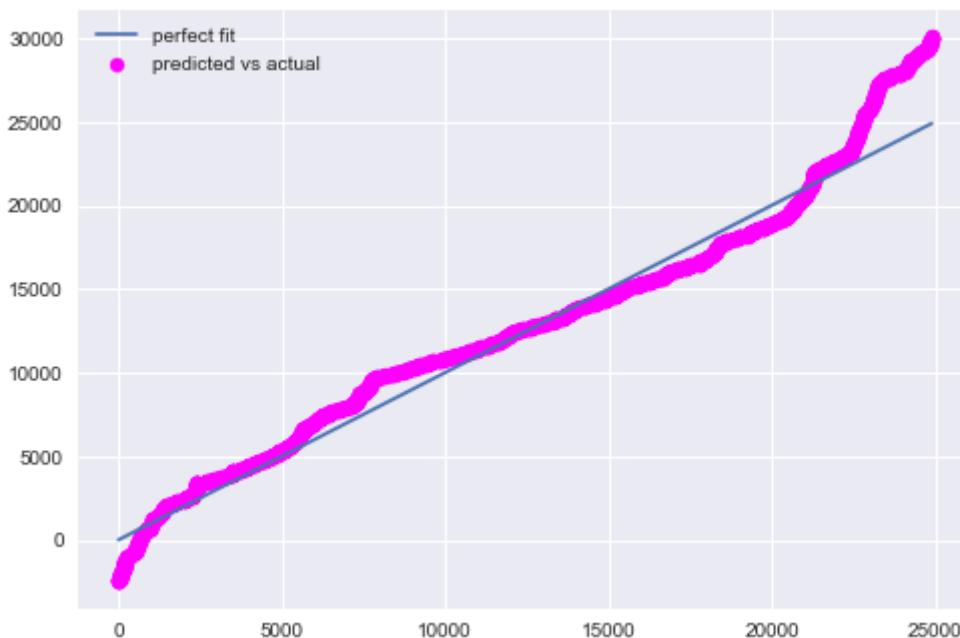
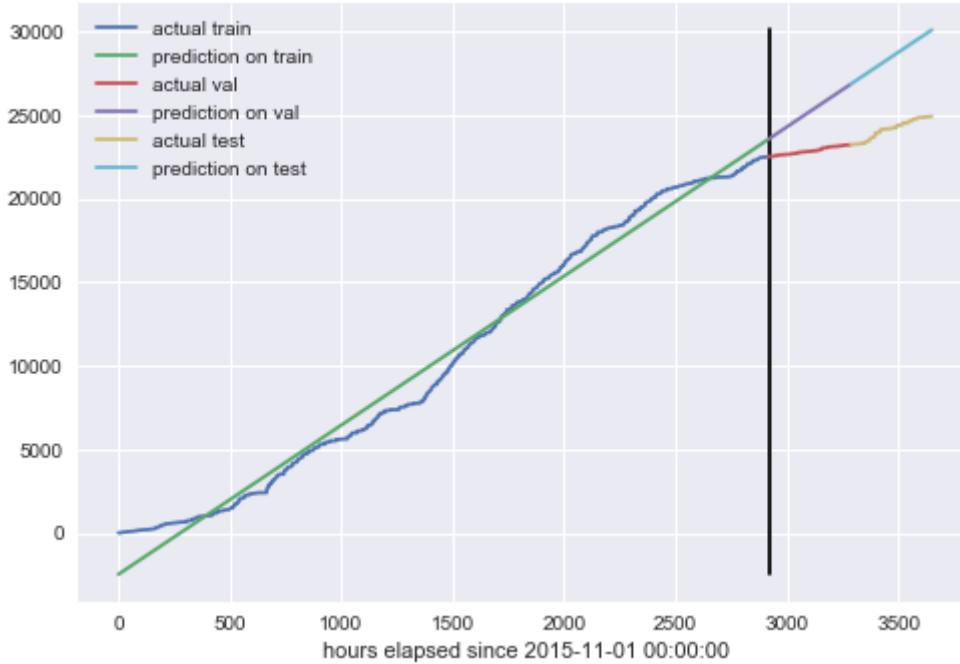
meterid 483; r2_train: 0.971; r2_val: -254.175; r2_test: -379.613

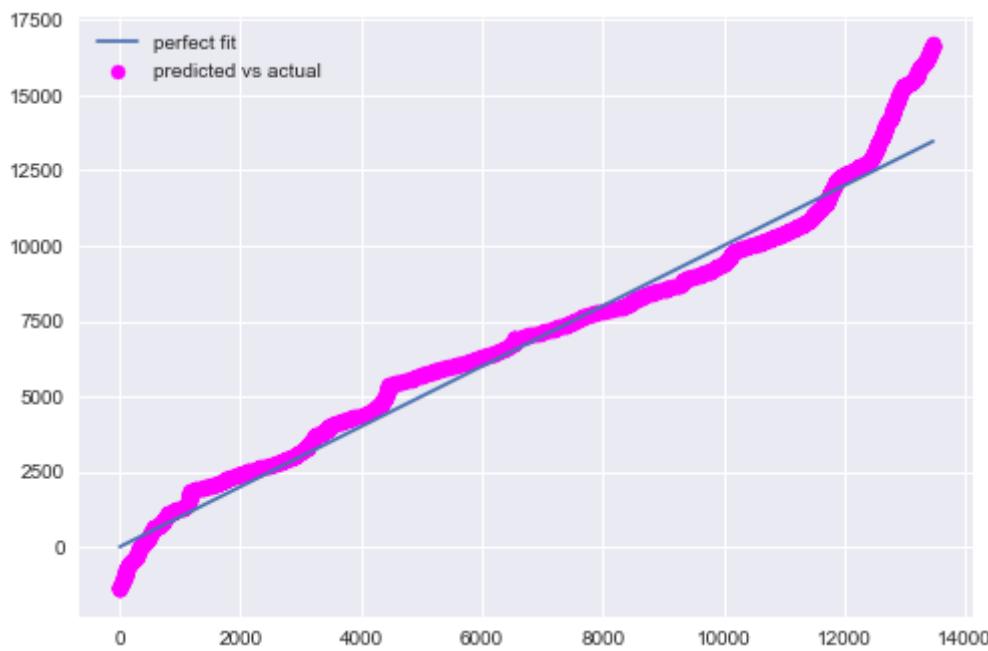
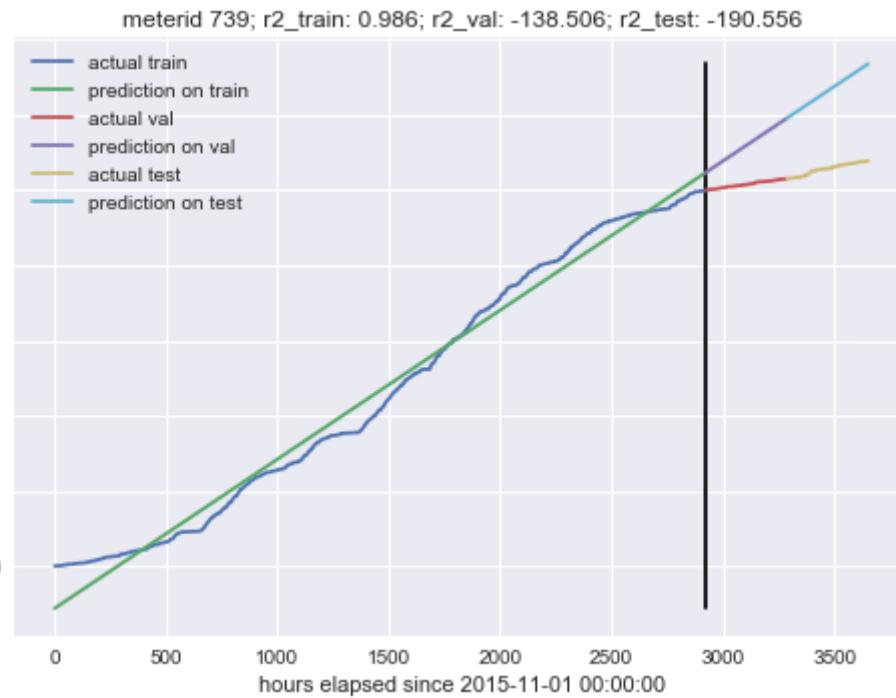


meterid 484; r2_train: 0.985; r2_val: 0.395; r2_test: -3.265

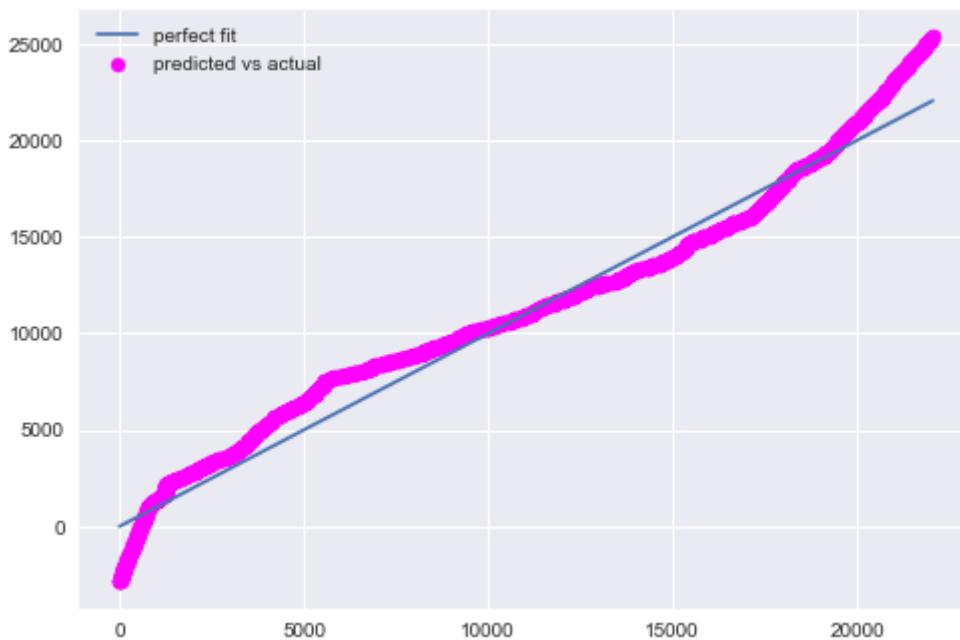
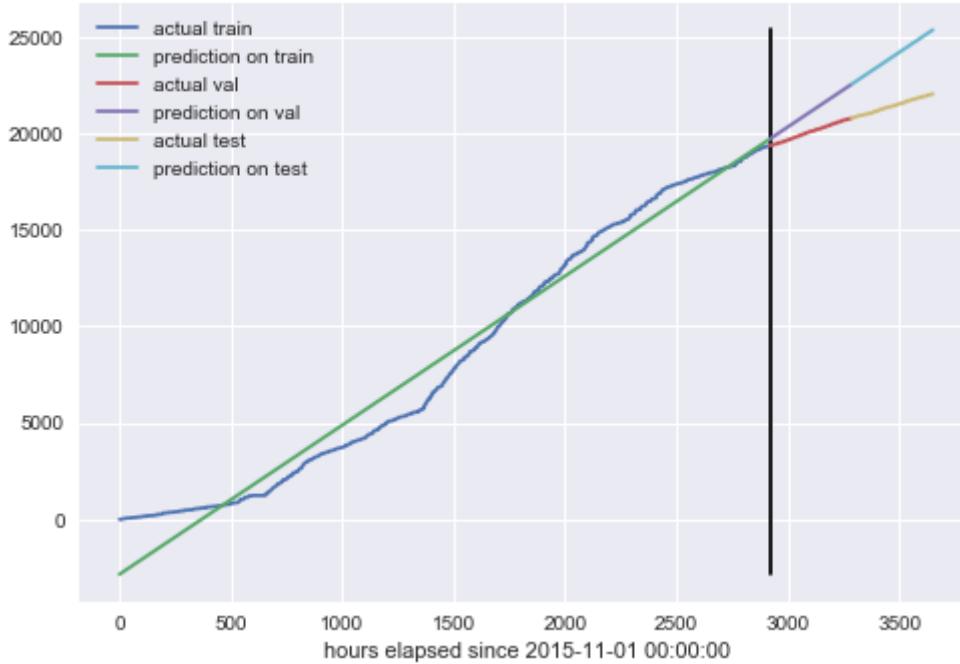


meterid 661; r2_train: 0.986; r2_val: -122.132; r2_test: -55.369

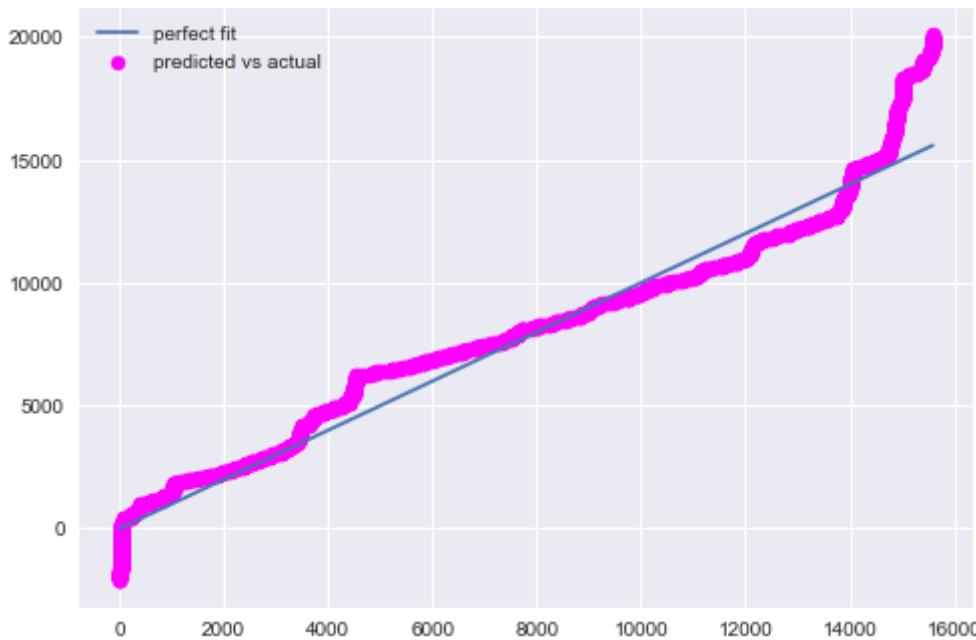
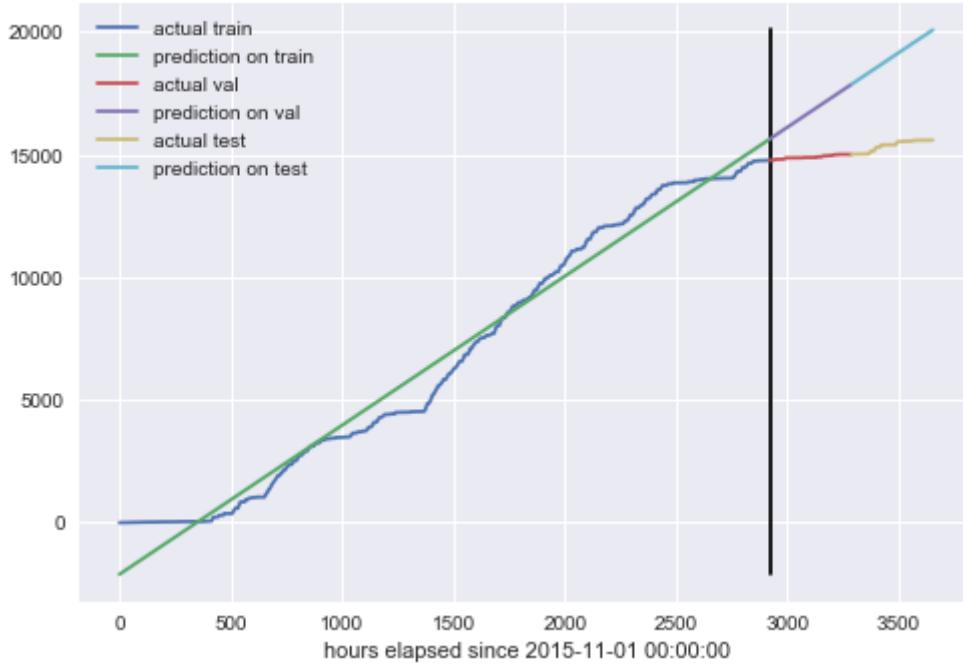




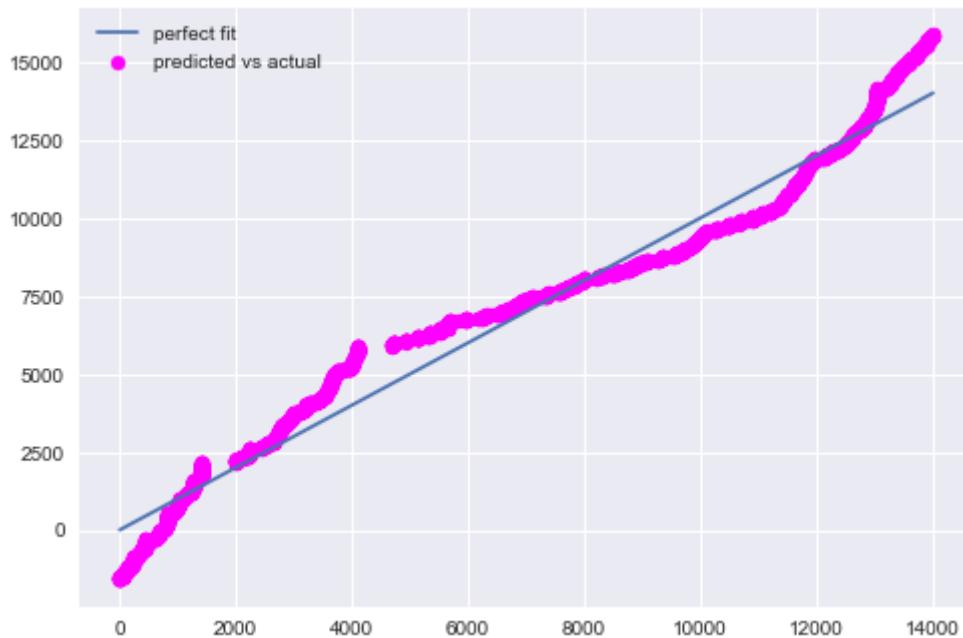
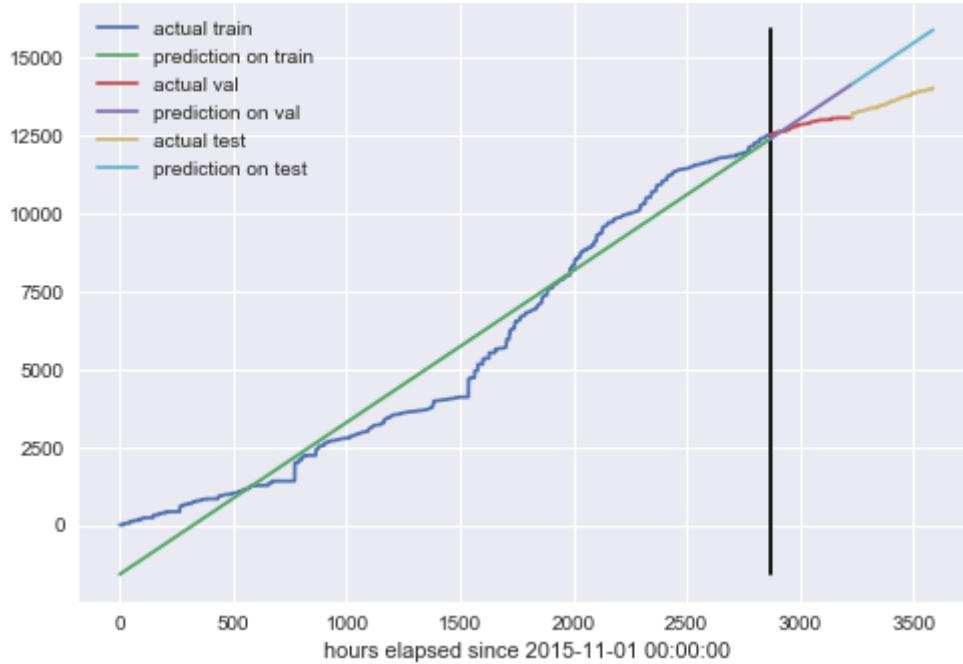
meterid 744; r2_train: 0.974; r2_val: -5.927; r2_test: -46.866



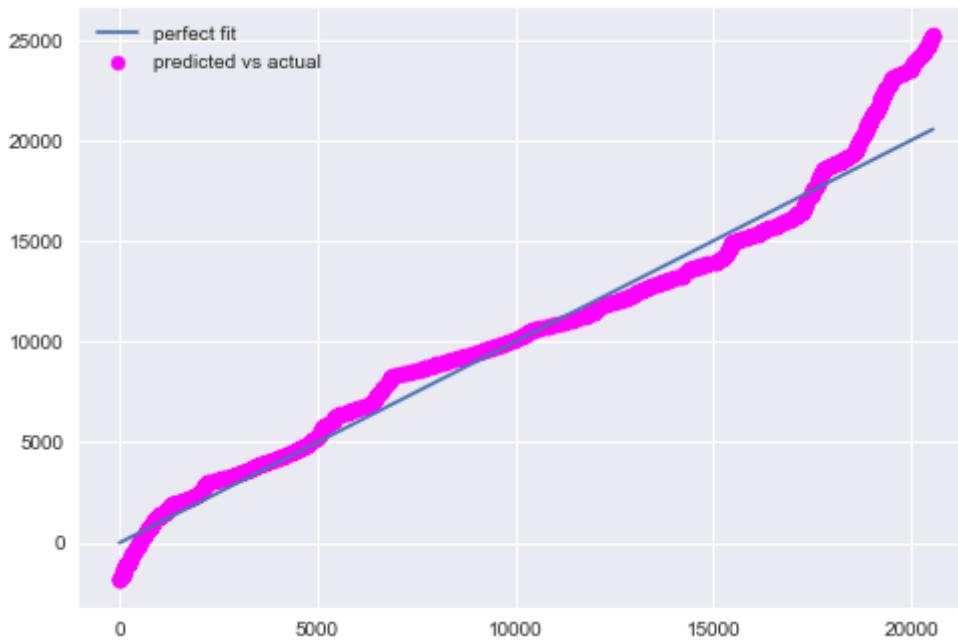
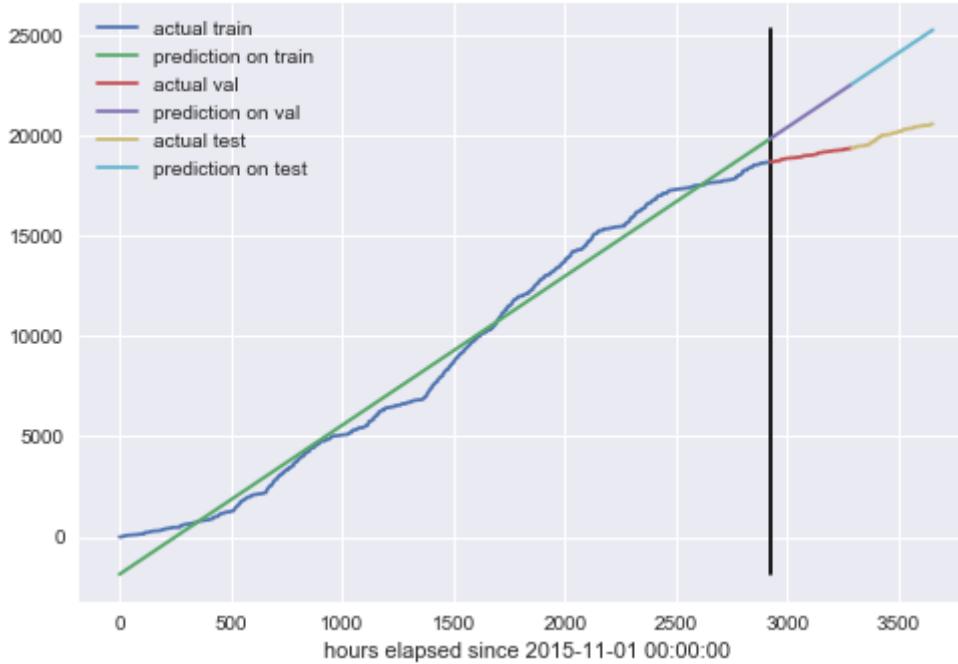
meterid 871; r2_train: 0.979; r2_val: -744.746; r2_test: -269.736



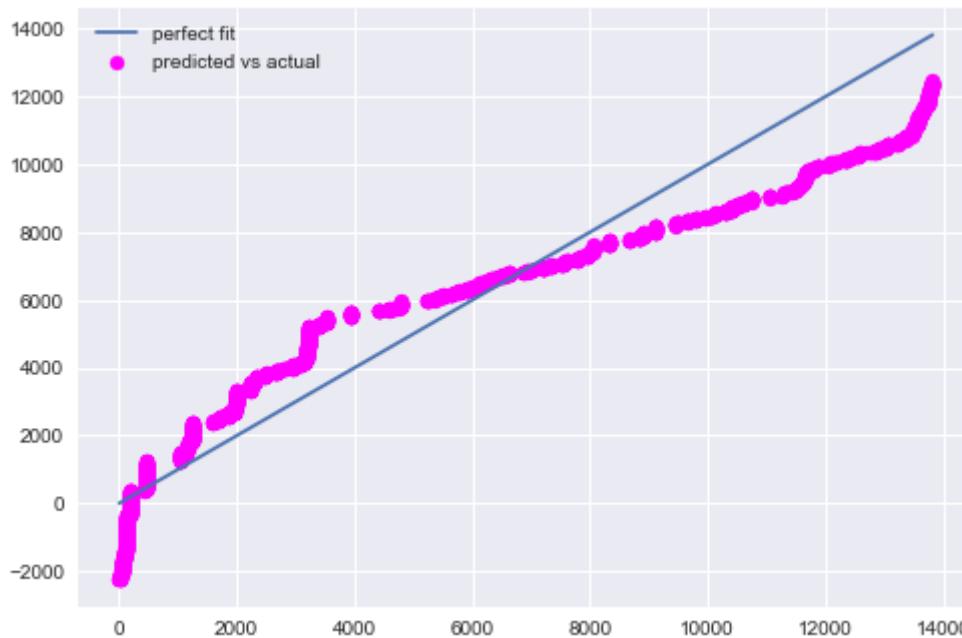
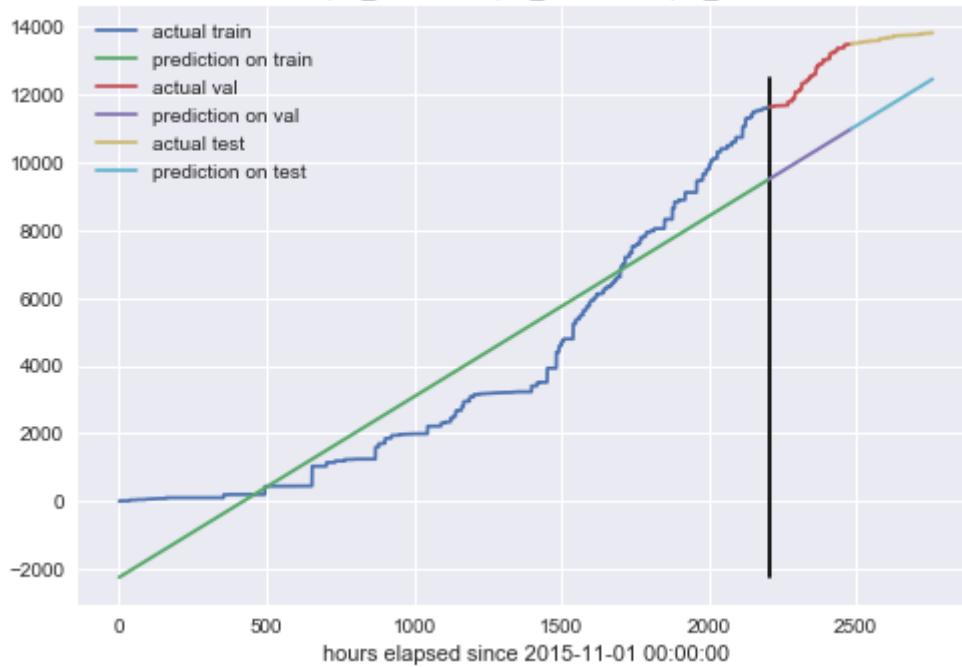
meterid 1042; r2_train: 0.964; r2_val: -7.976; r2_test: -31.689



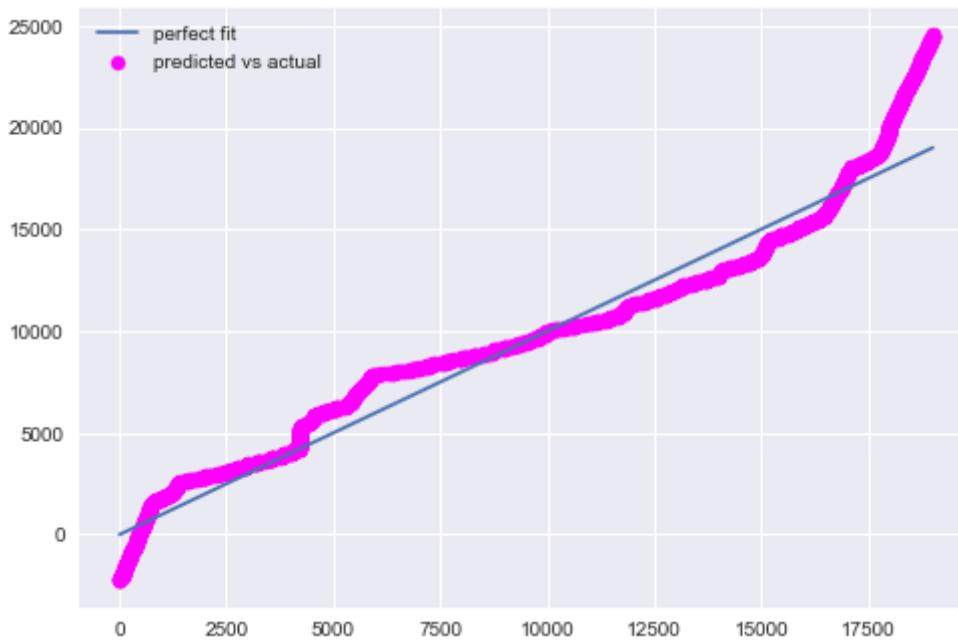
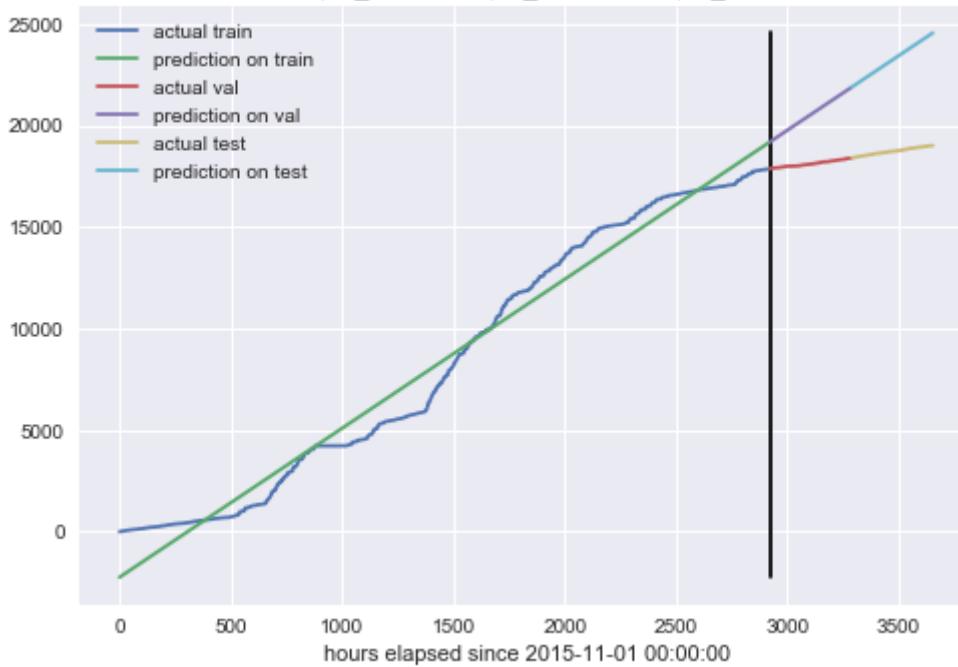
meterid 1086; r2_train: 0.987; r2_val: -121.656; r2_test: -101.074



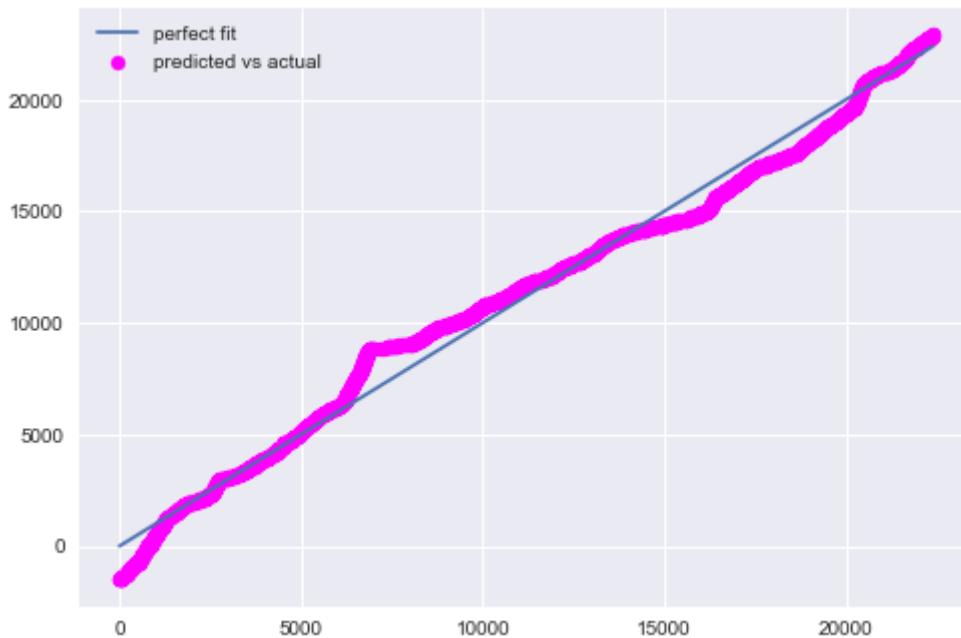
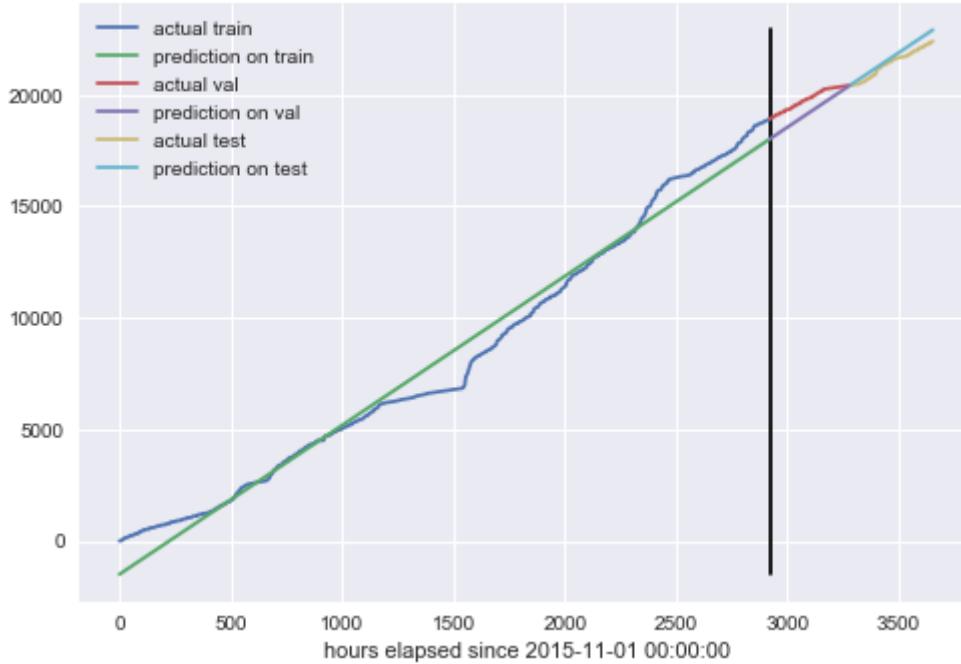
meterid 1103; r2_train: 0.891; r2_val: -10.832; r2_test: -388.336

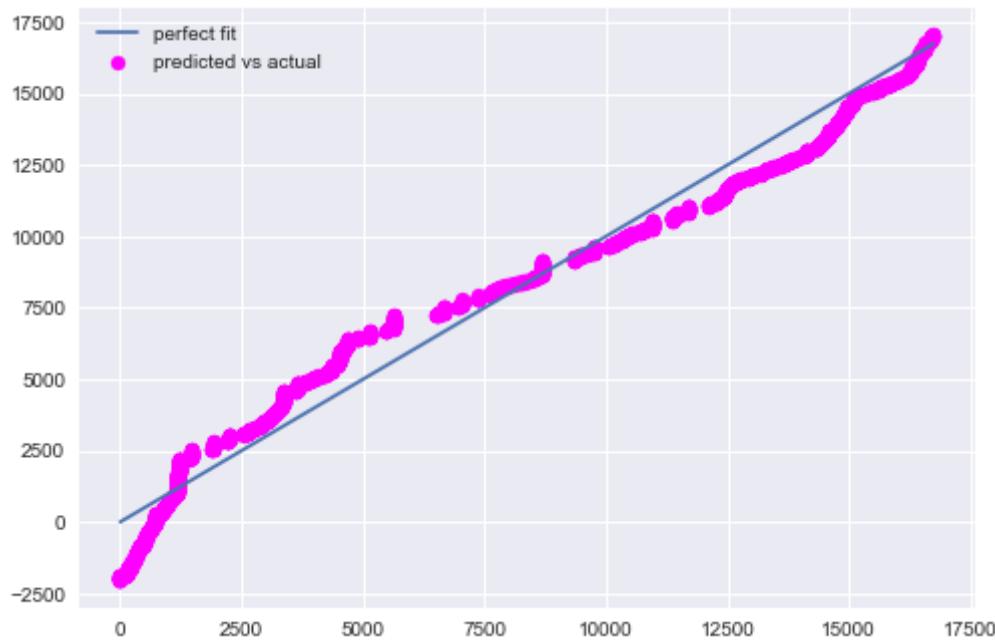
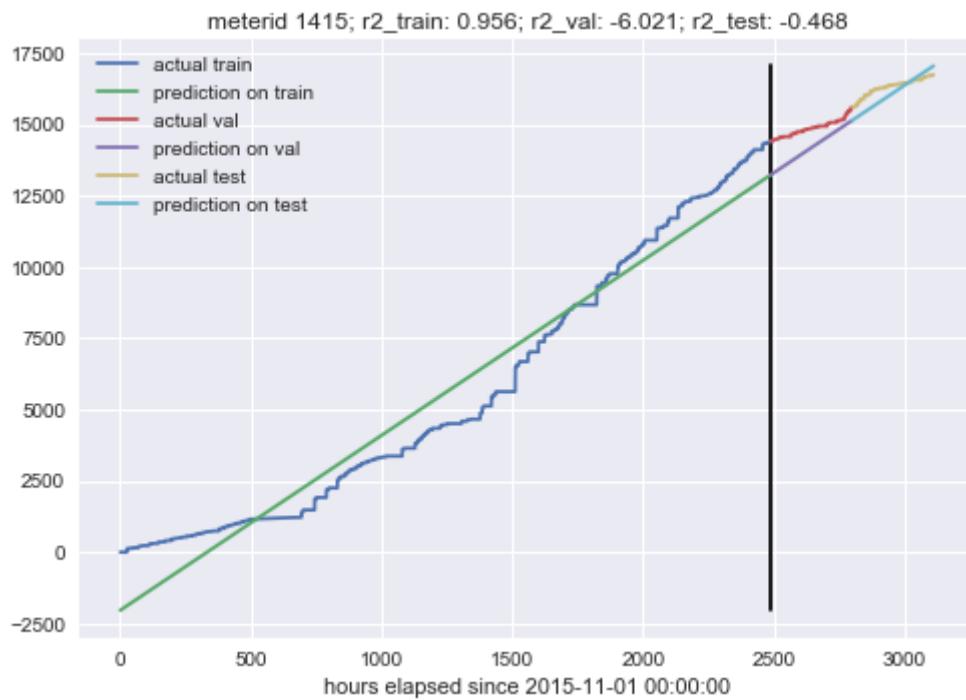


meterid 1185; r2_train: 0.977; r2_val: -278.277; r2_test: -657.119

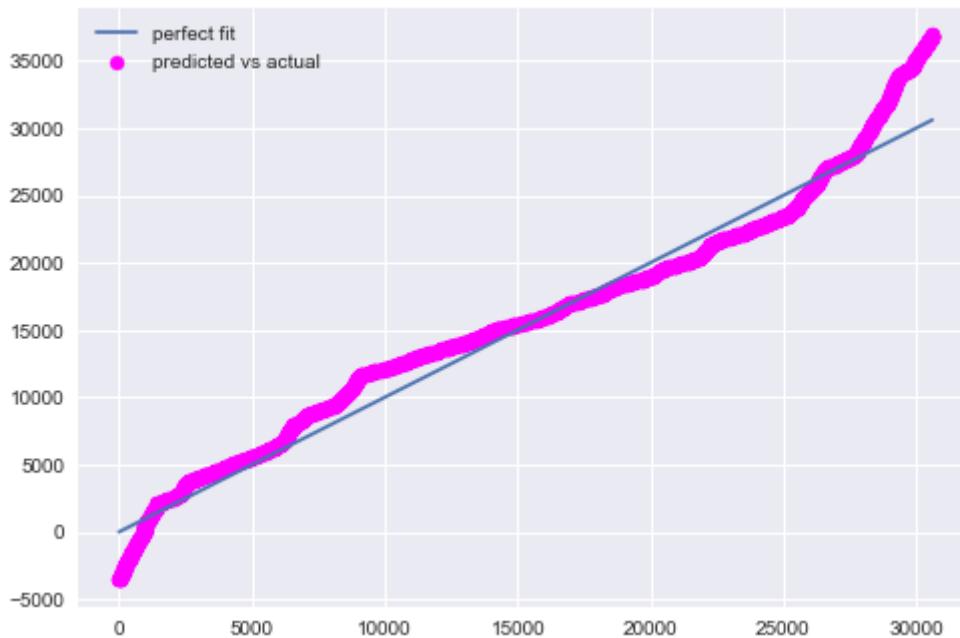
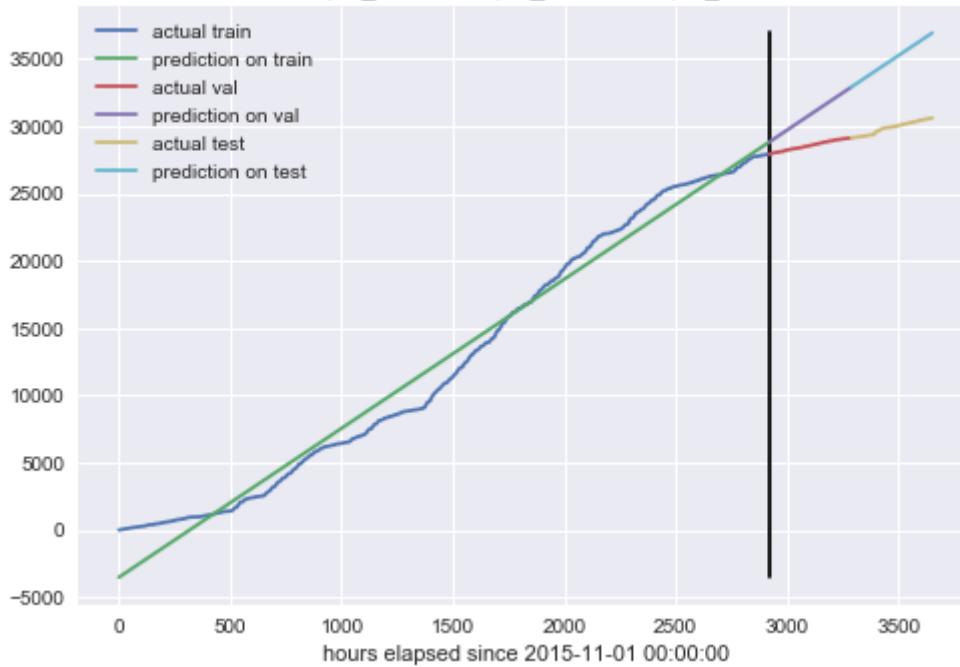


meterid 1283; r2_train: 0.983; r2_val: -0.776; r2_test: 0.745

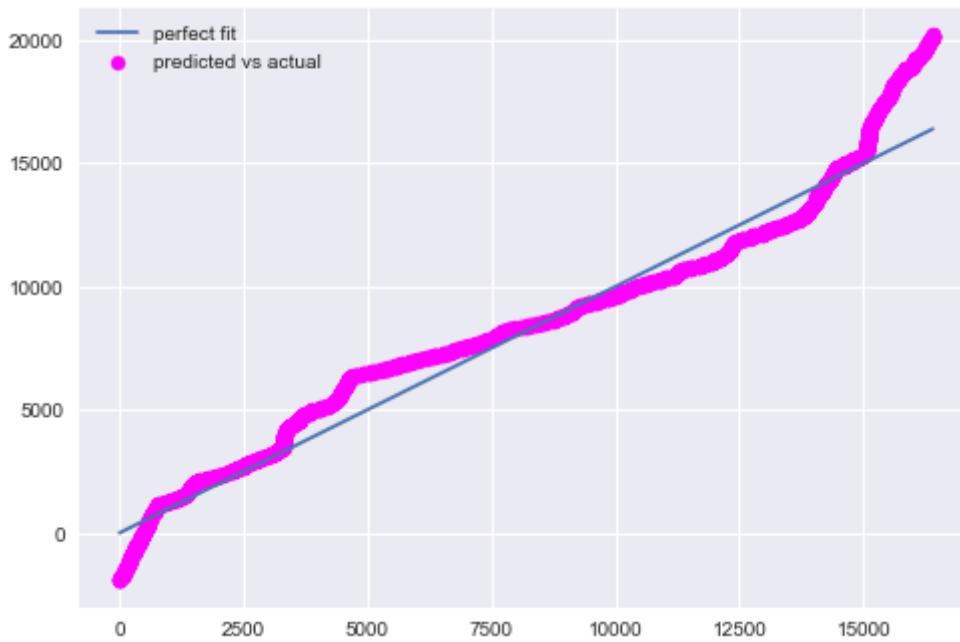
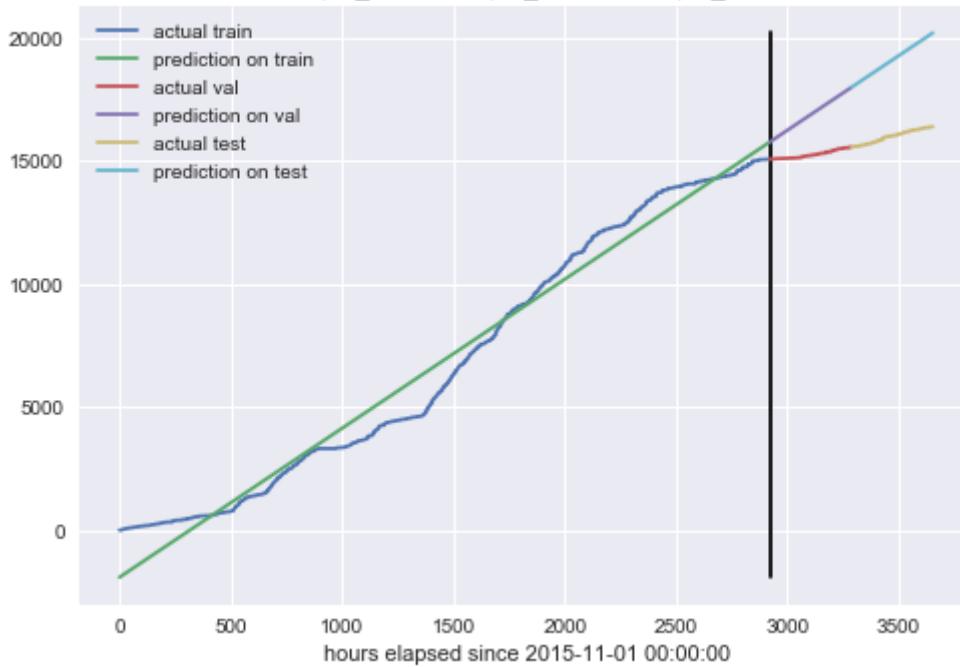




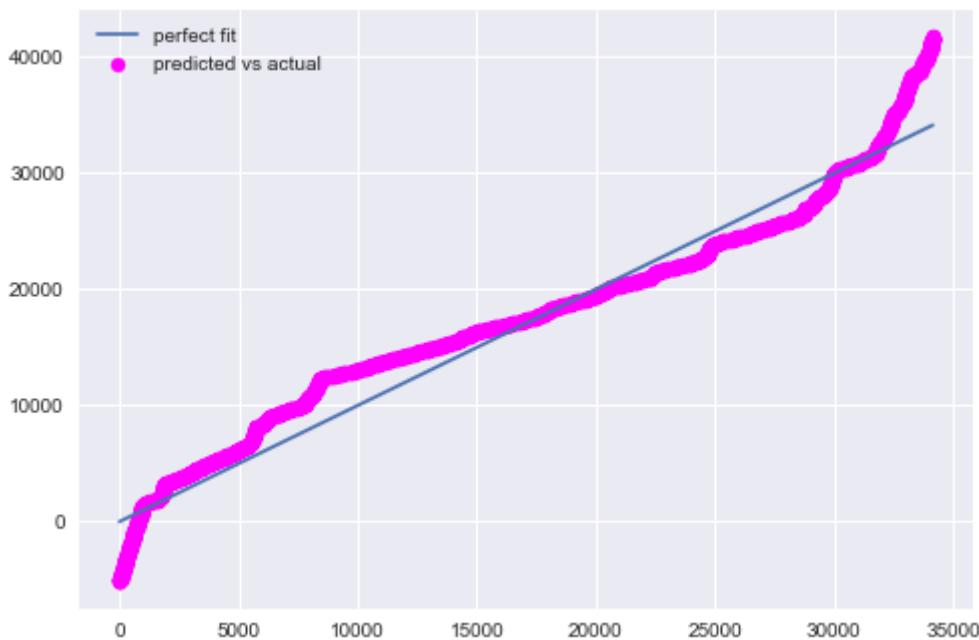
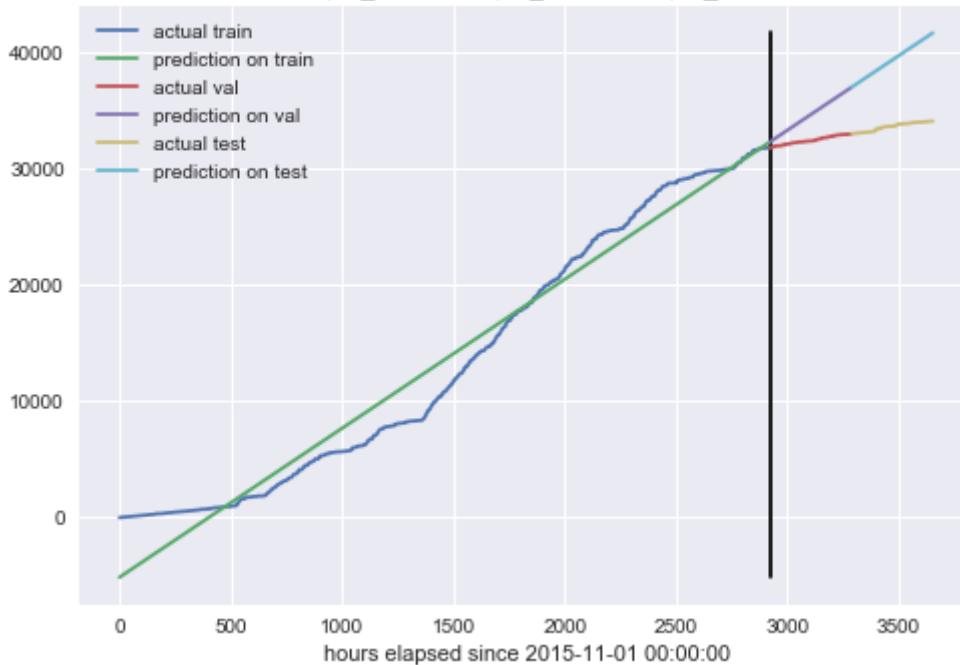
meterid 1507; r2_train: 0.981; r2_val: -44.690; r2_test: -115.381



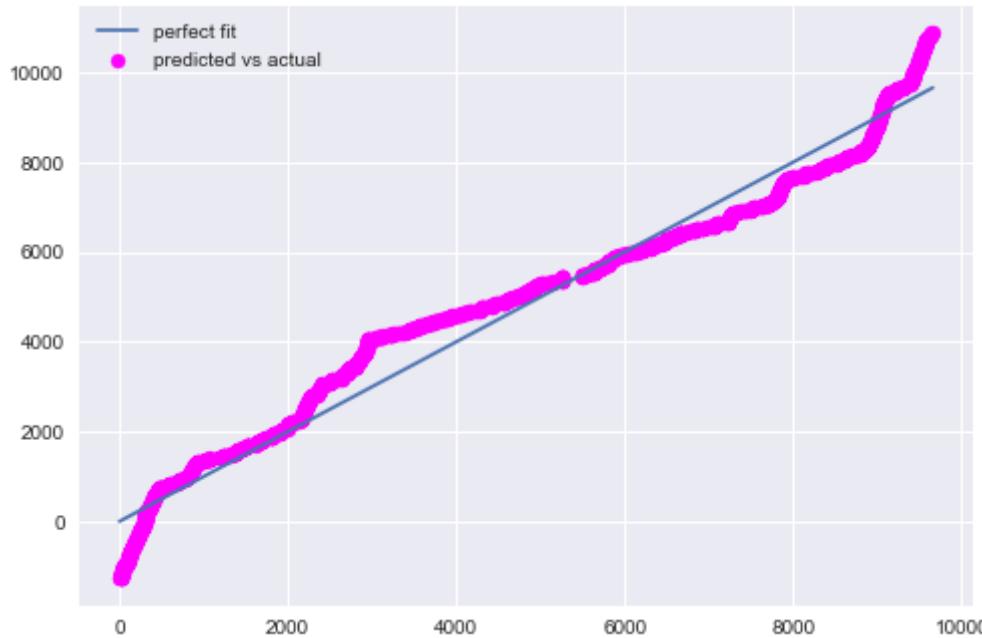
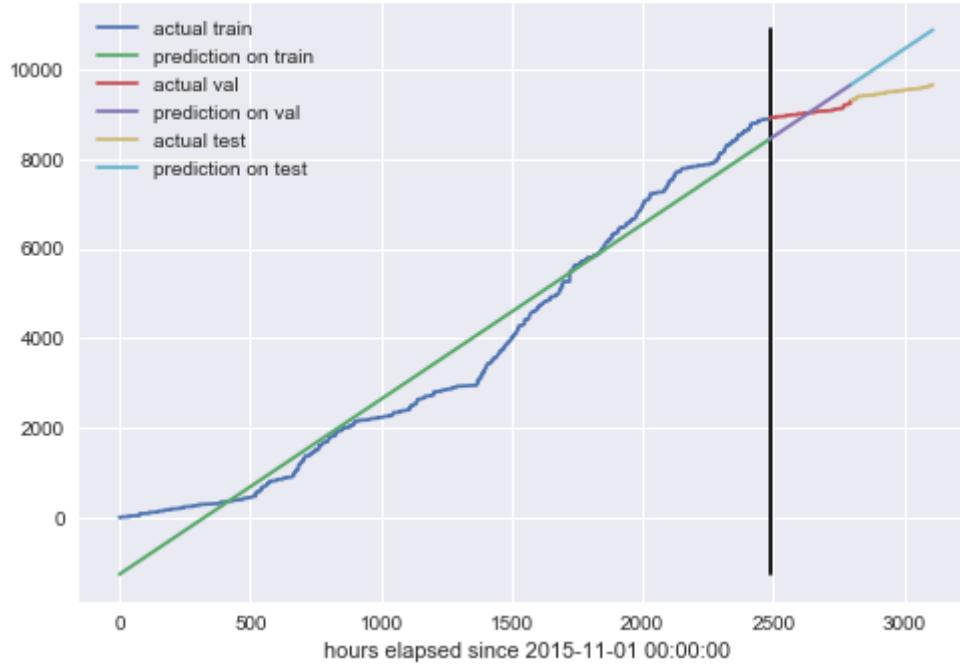
meterid 1556; r2_train: 0.977; r2_val: -111.823; r2_test: -137.802



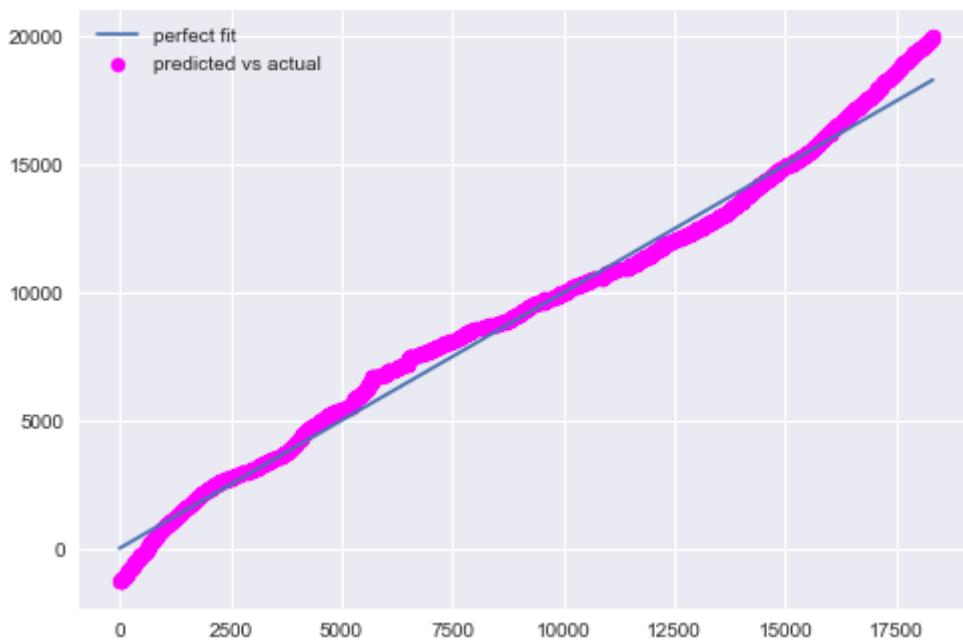
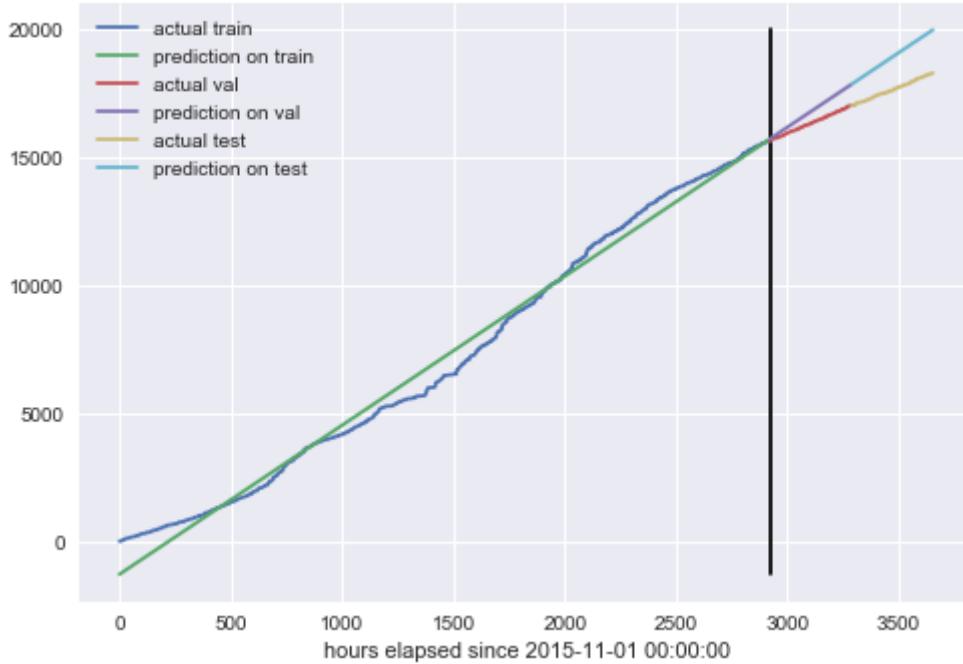
meterid 1589; r2_train: 0.968; r2_val: -47.073; r2_test: -249.586



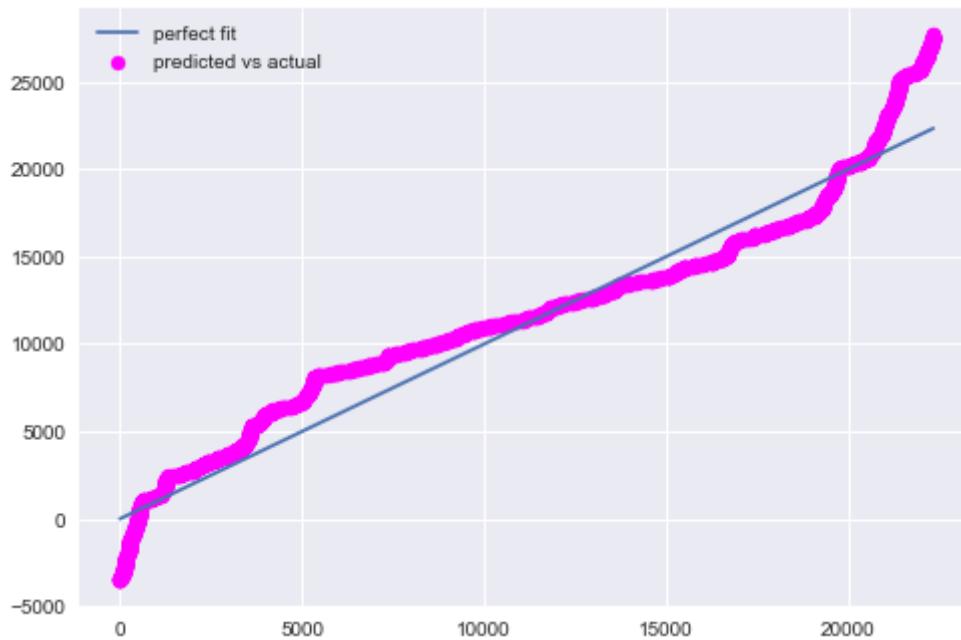
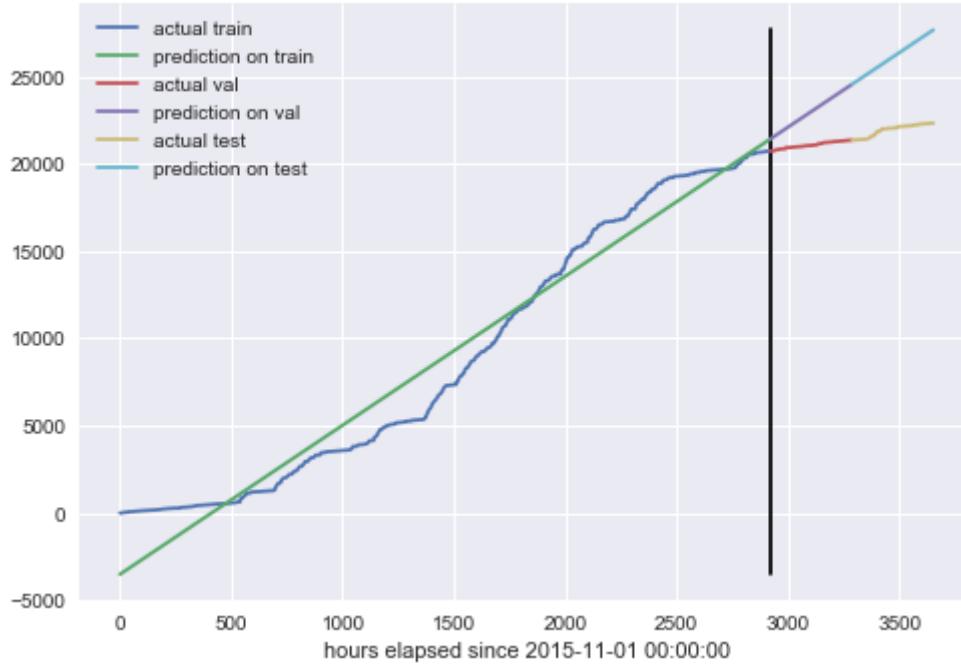
meterid 1619; r2_train: 0.966; r2_val: -9.462; r2_test: -99.969



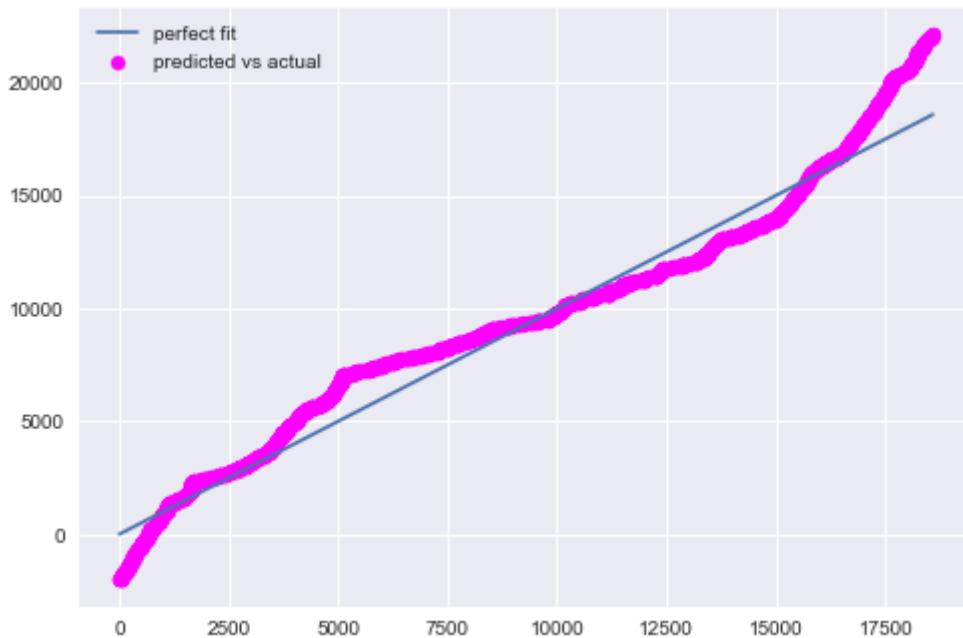
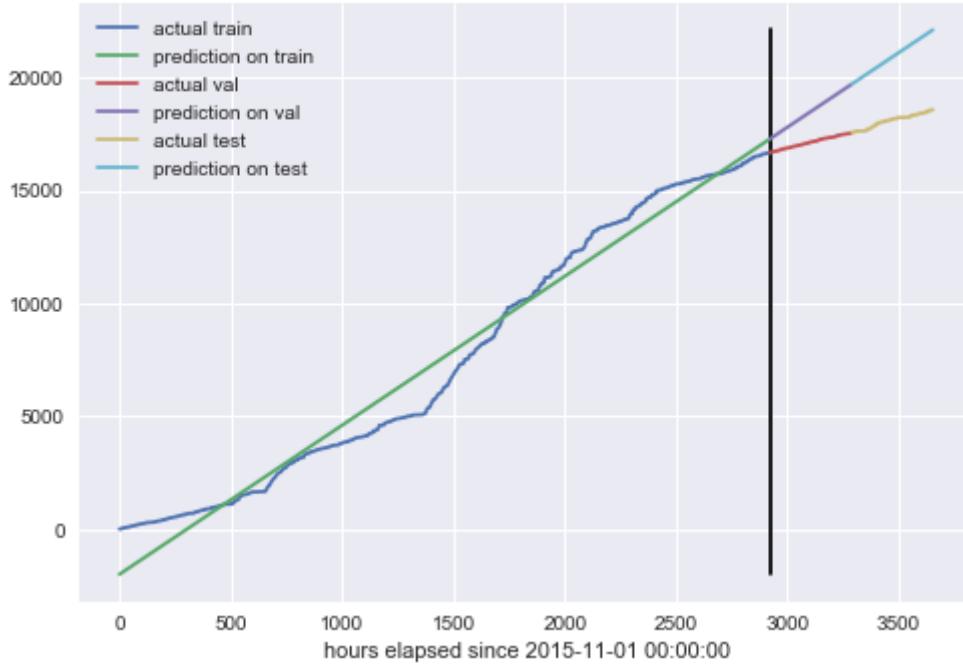
meterid 1697; r2_train: 0.990; r2_val: -0.614; r2_test: -10.957



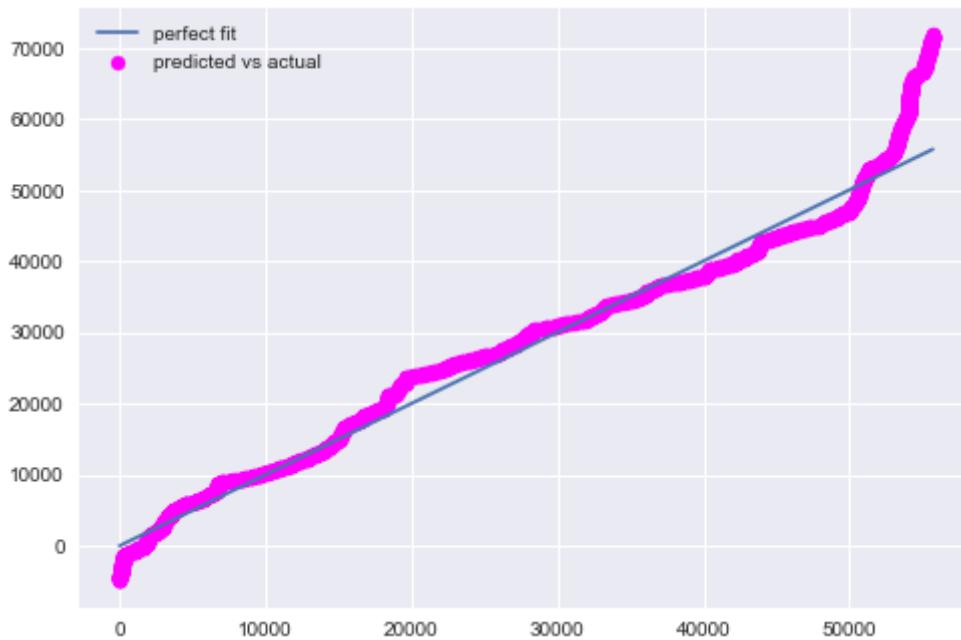
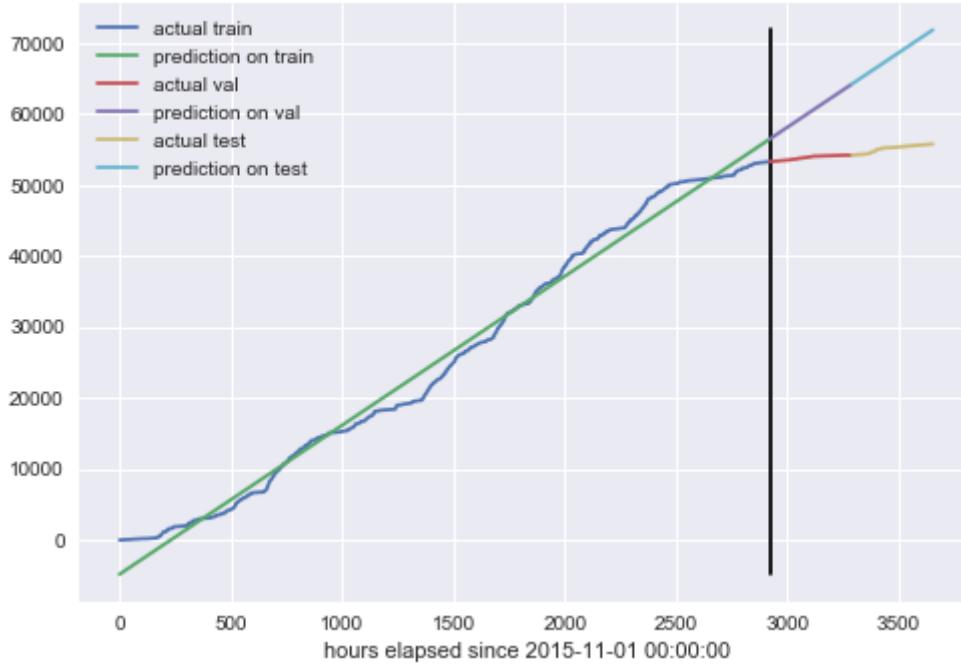
meterid 1714; r2_train: 0.962; r2_val: -123.822; r2_test: -154.902



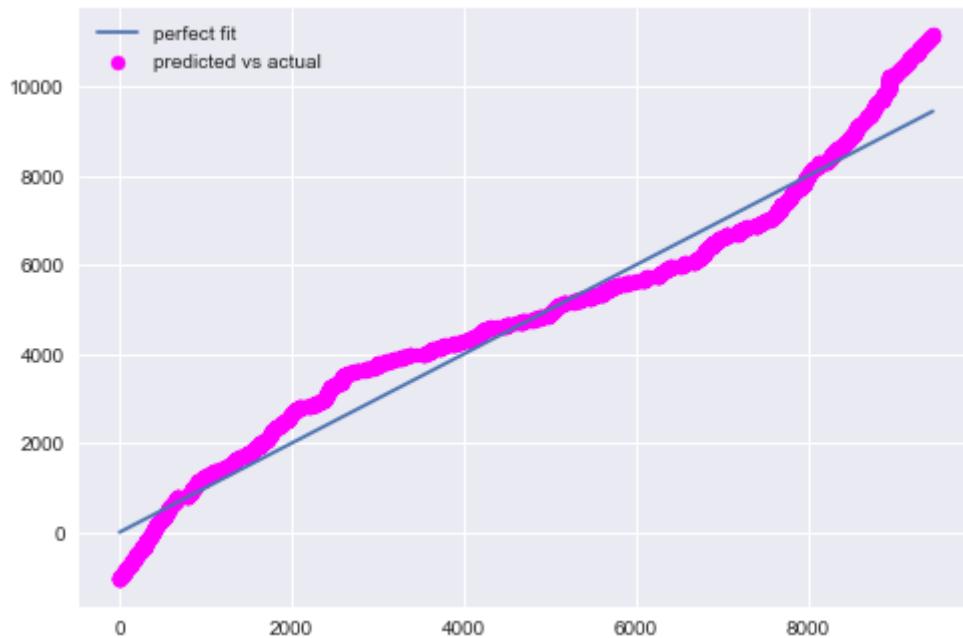
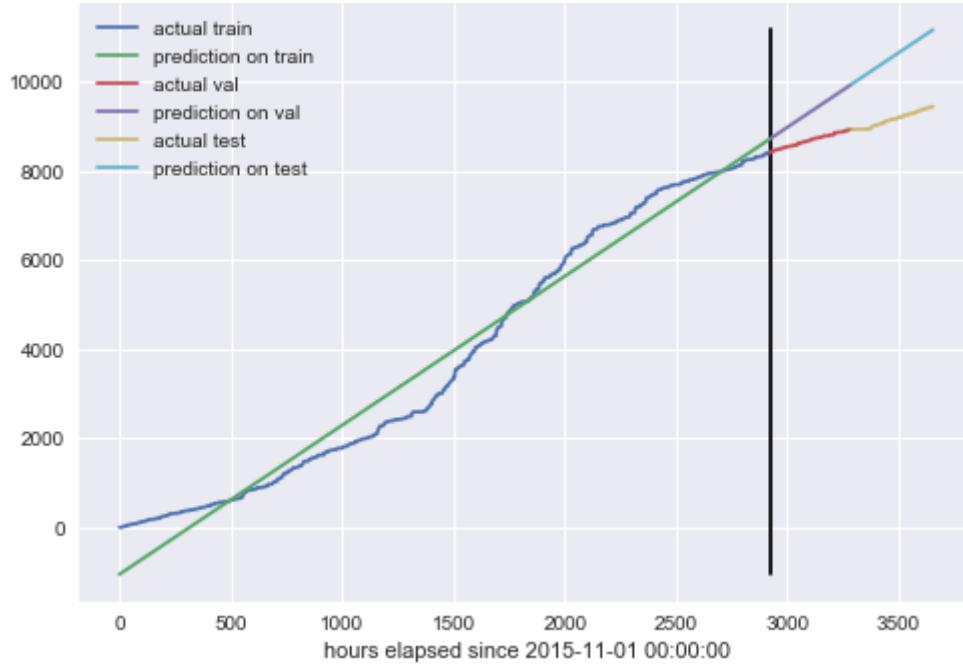
meterid 1718; r2_train: 0.976; r2_val: -30.295; r2_test: -88.032



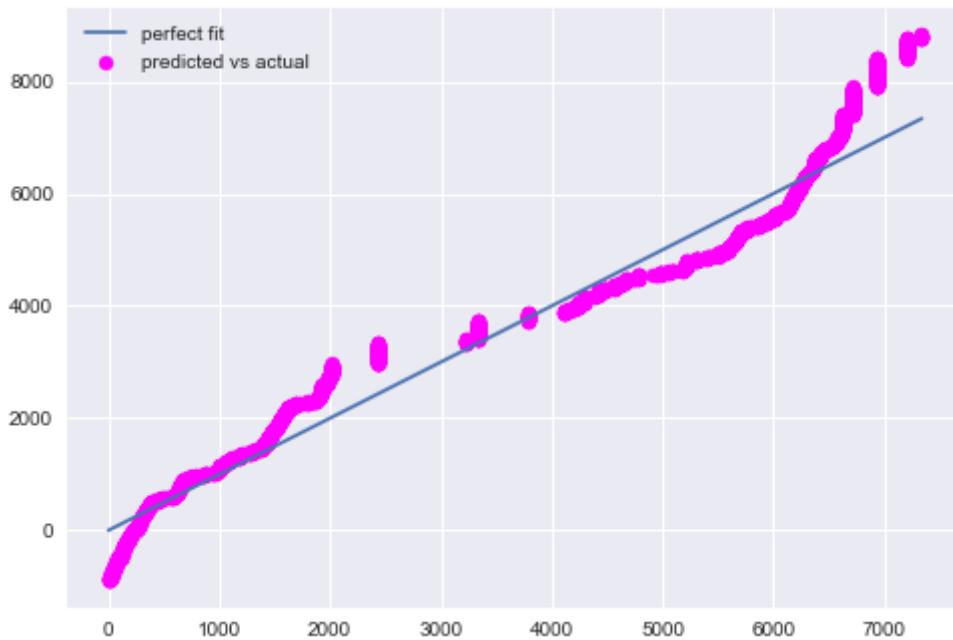
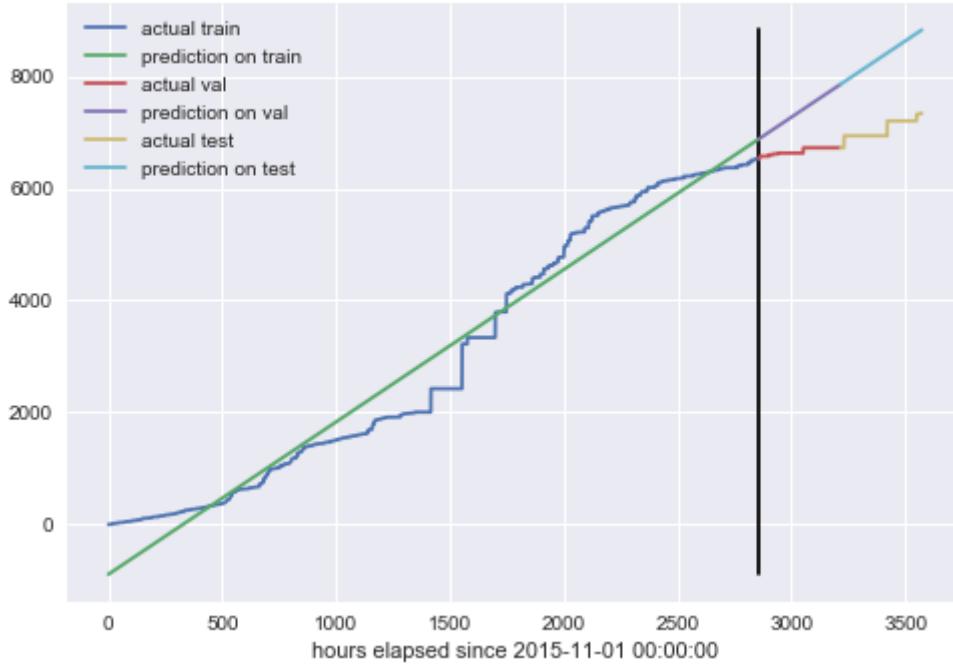
meterid 1790; r2_train: 0.989; r2_val: -481.912; r2_test: -641.092



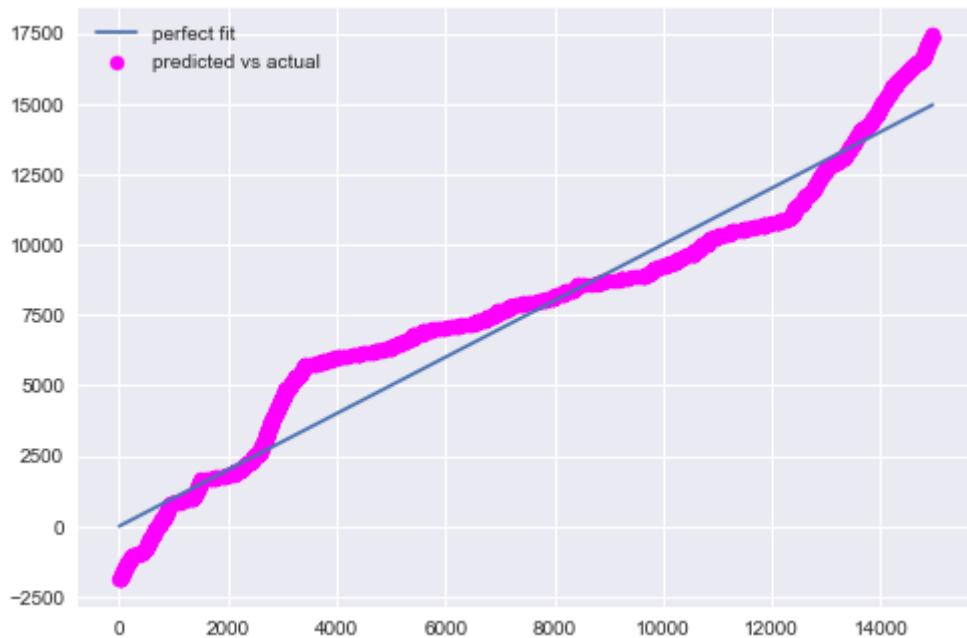
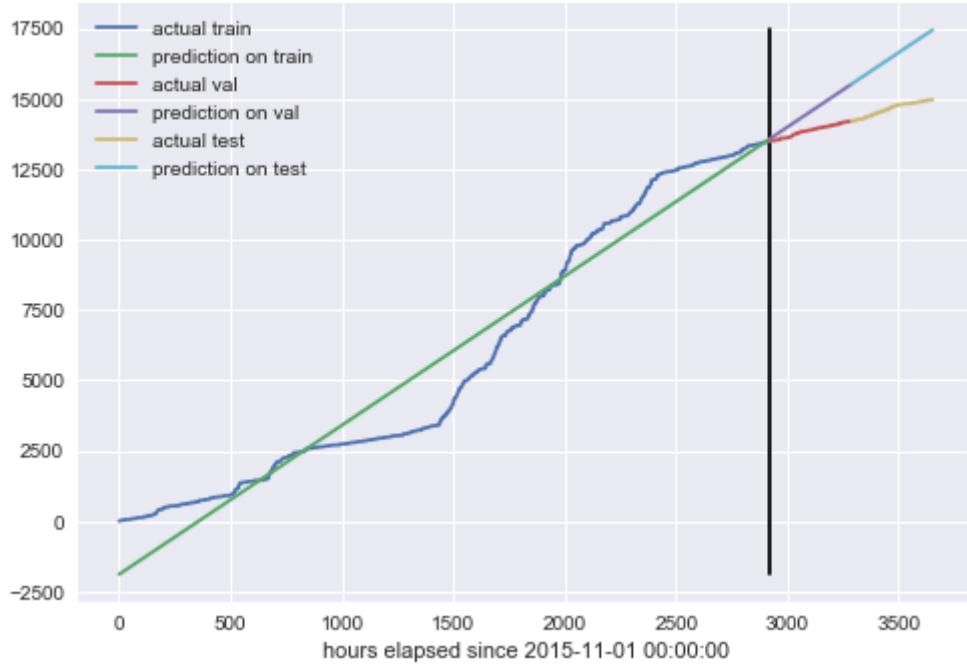
meterid 1791; r2_train: 0.973; r2_val: -19.917; r2_test: -69.982



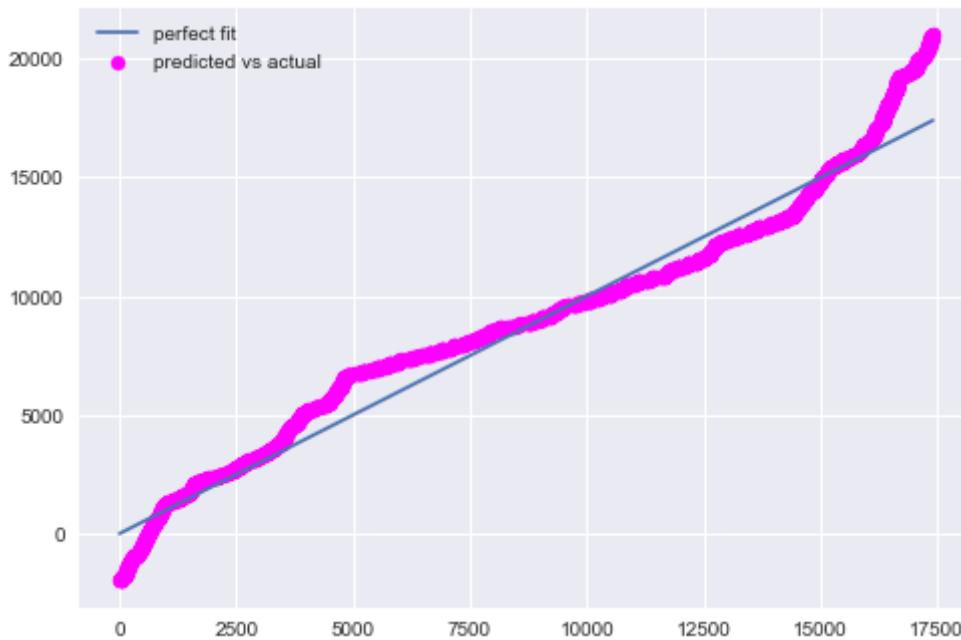
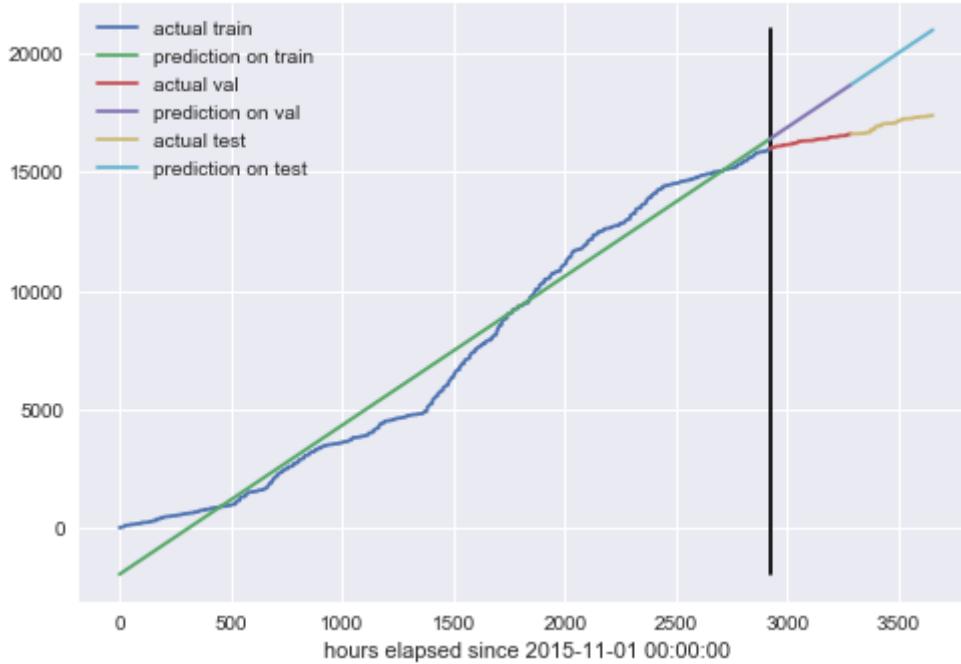
meterid 1792; r2_train: 0.967; r2_val: -159.672; r2_test: -66.899



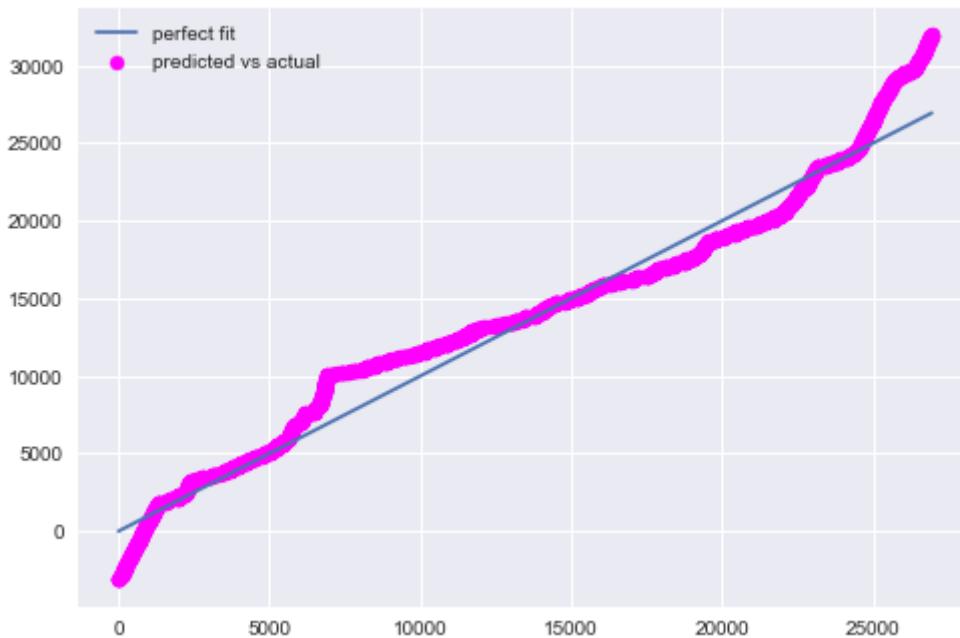
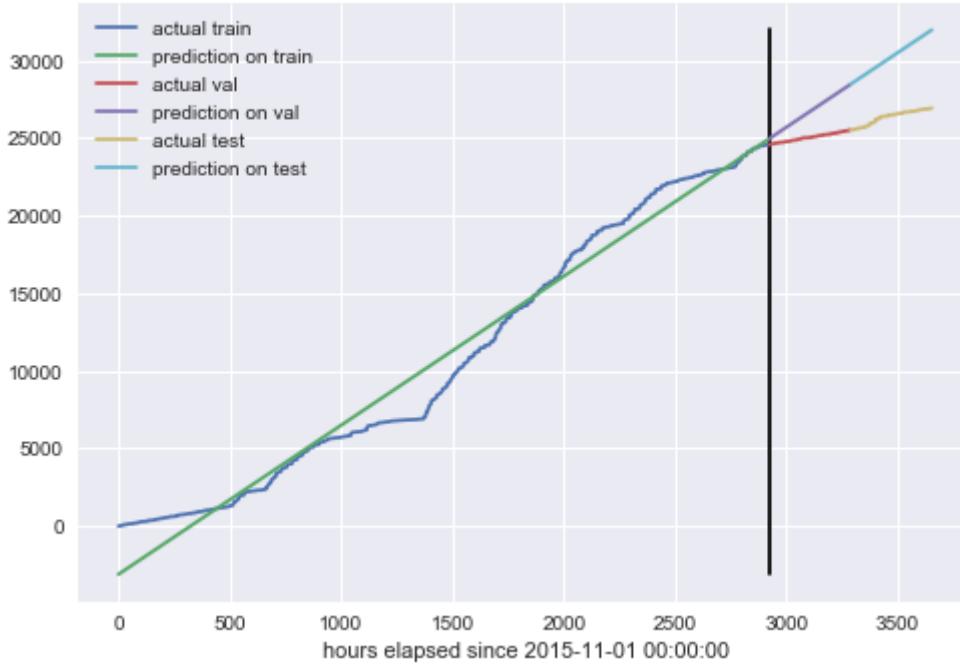
meterid 1800; r2_train: 0.949; r2_val: -11.398; r2_test: -58.719



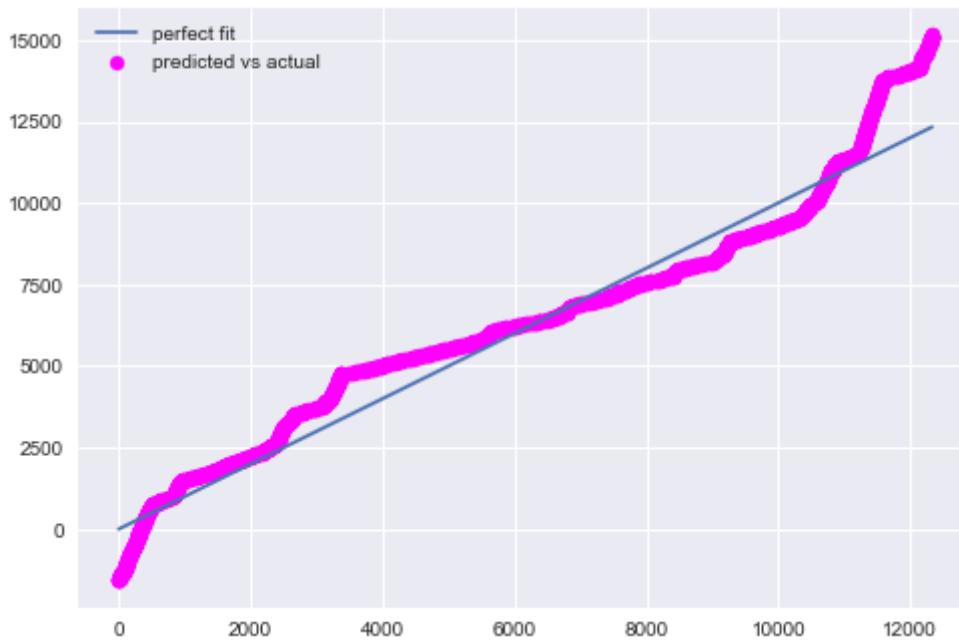
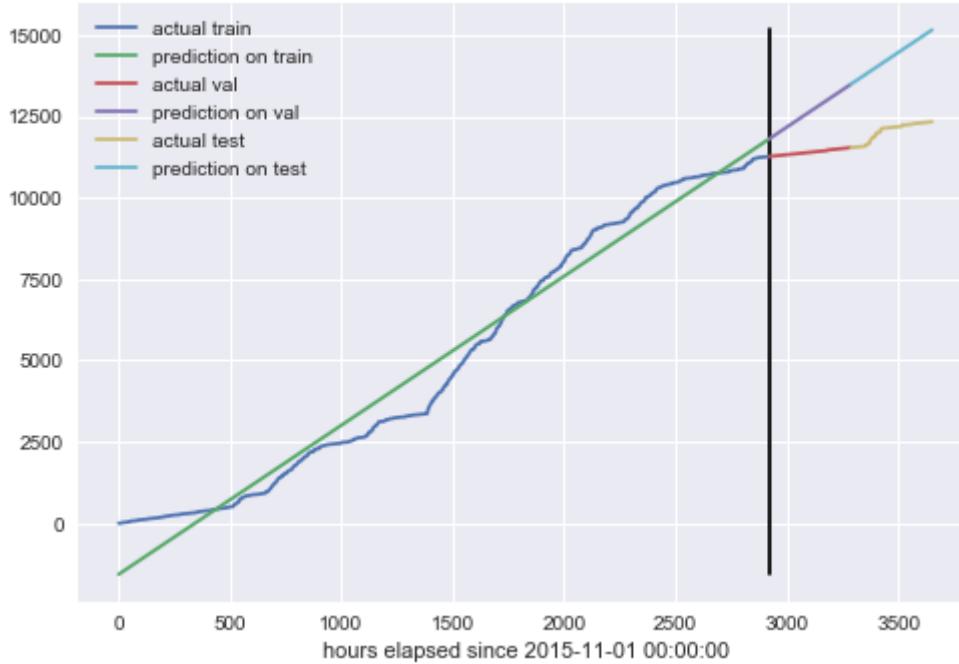
meterid 1801; r2_train: 0.977; r2_val: -63.359; r2_test: -111.285



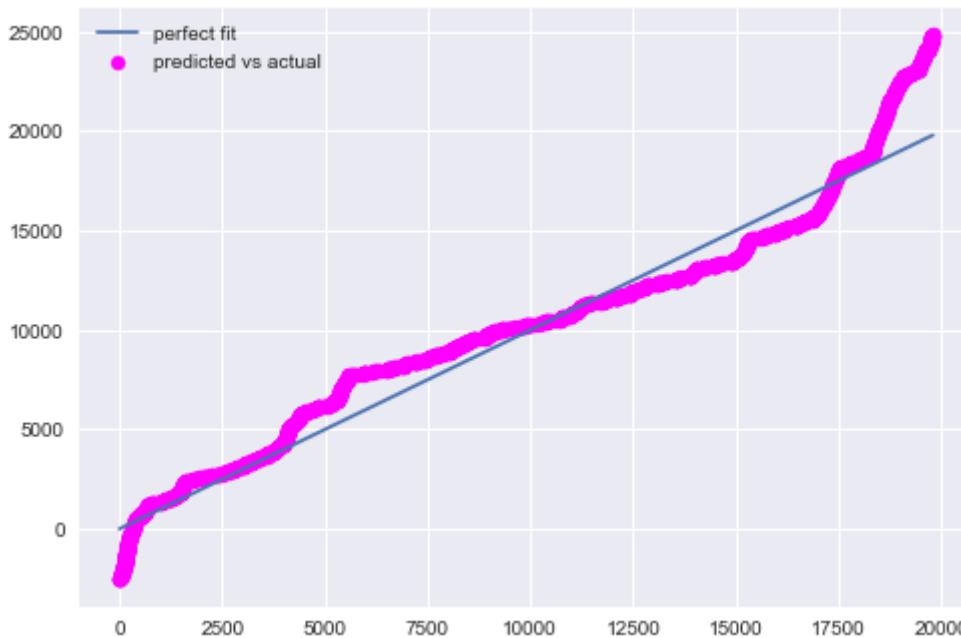
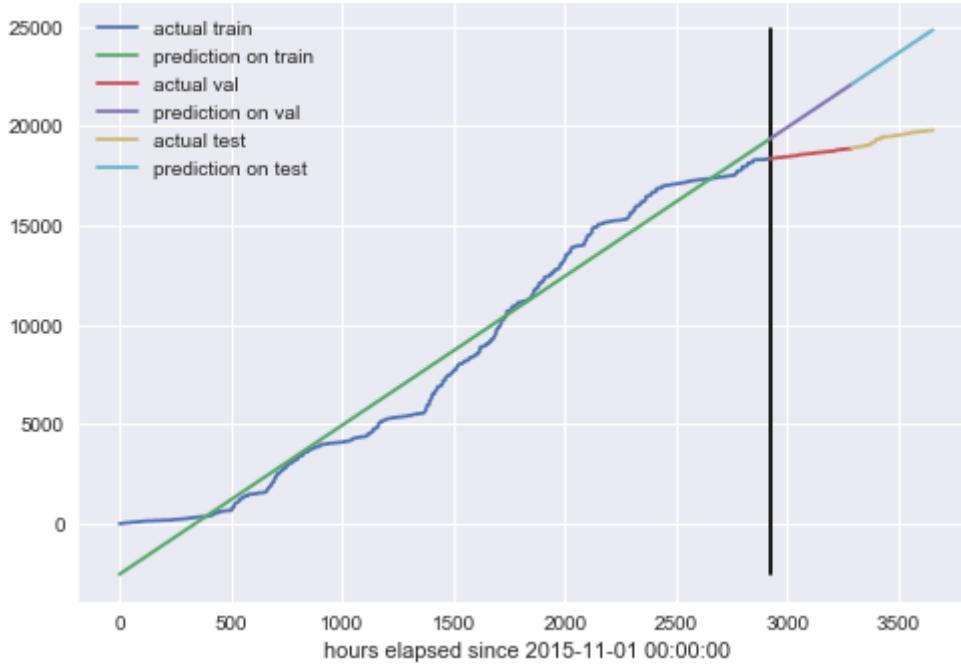
meterid 2018; r2_train: 0.976; r2_val: -45.883; r2_test: -74.573



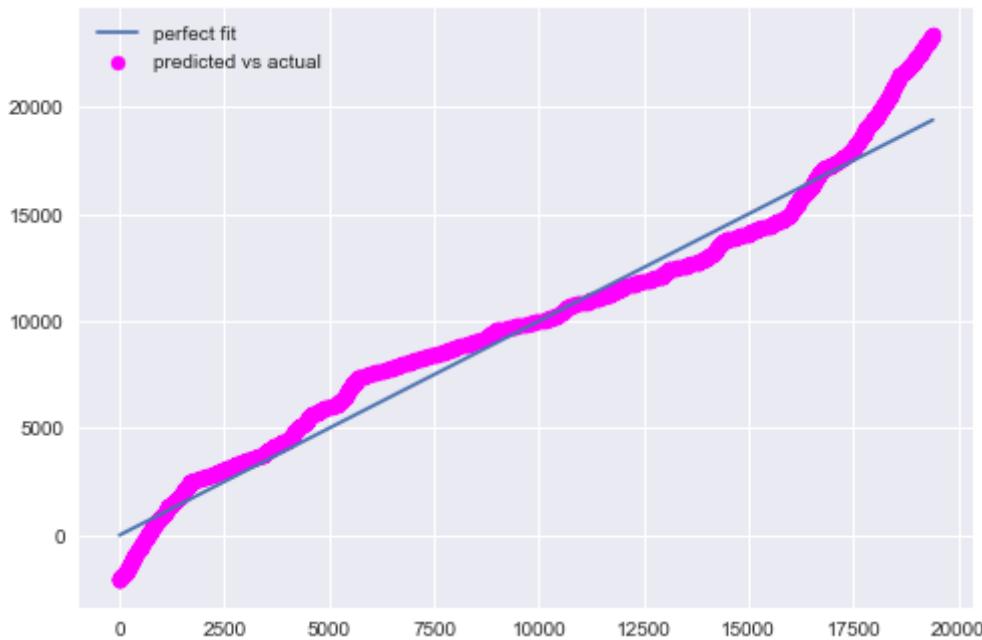
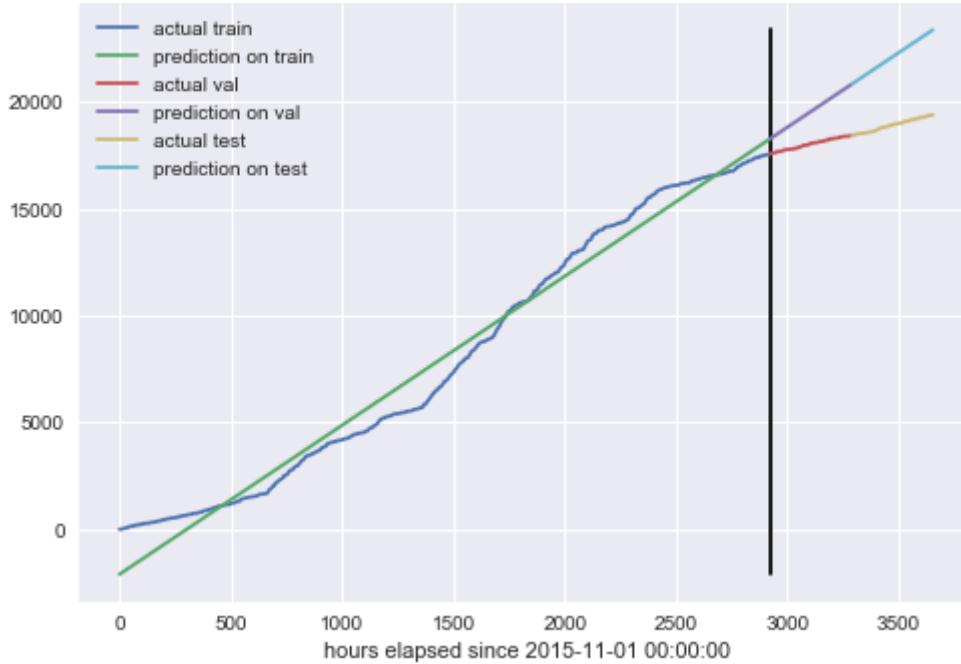
meterid 2034; r2_train: 0.974; r2_val: -273.068; r2_test: -65.606



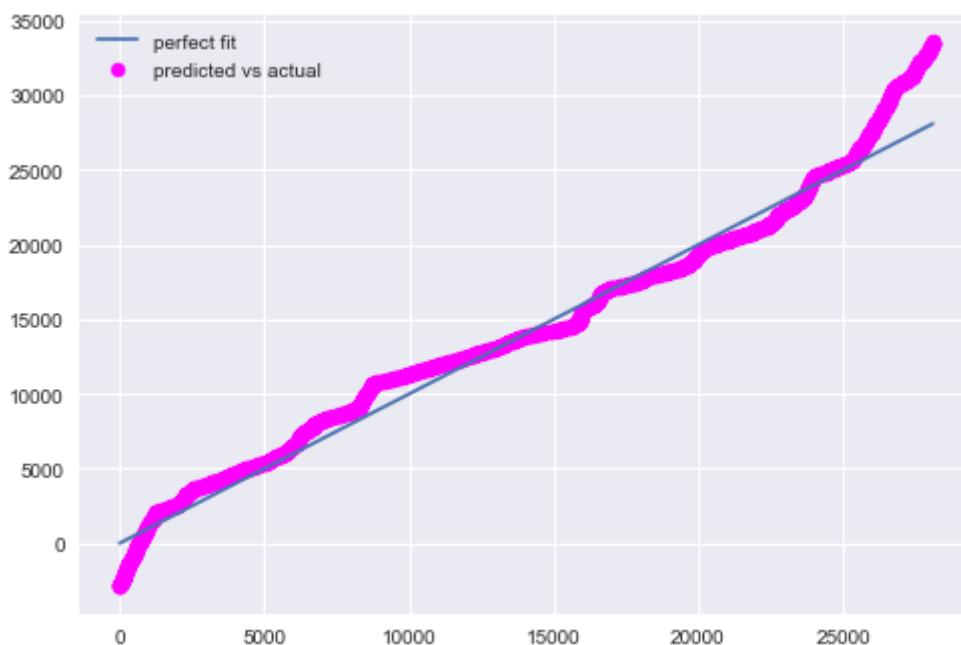
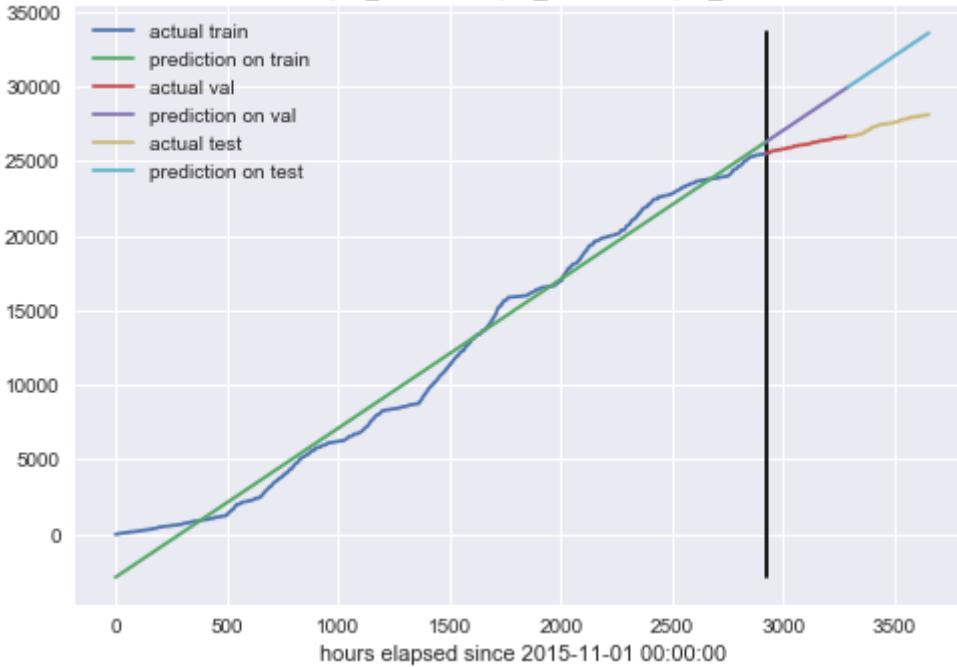
meterid 2072; r2_train: 0.975; r2_val: -215.773; r2_test: -198.096



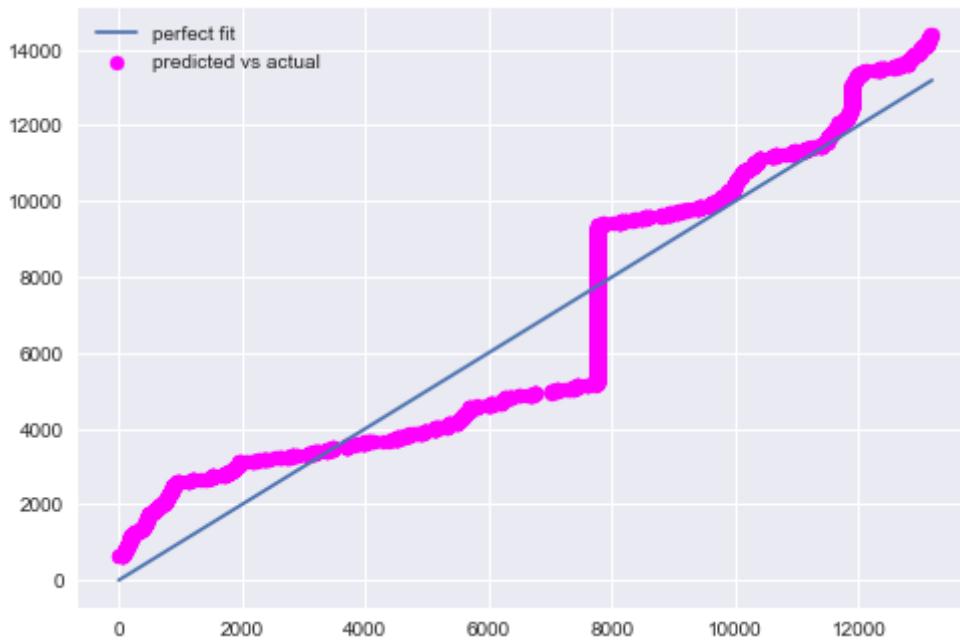
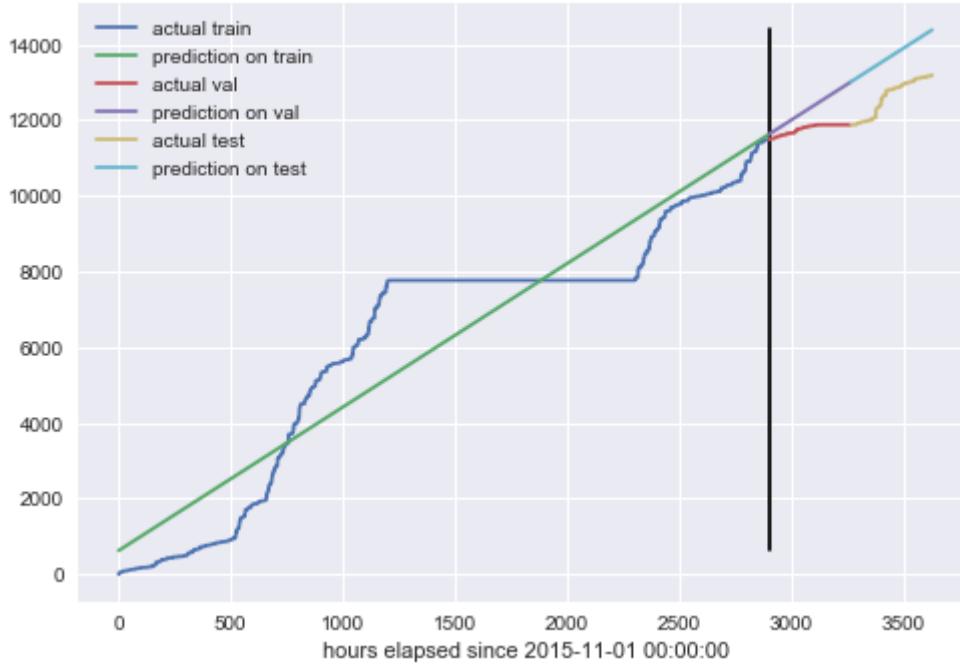
meterid 2094; r2_train: 0.980; r2_val: -37.260; r2_test: -117.753



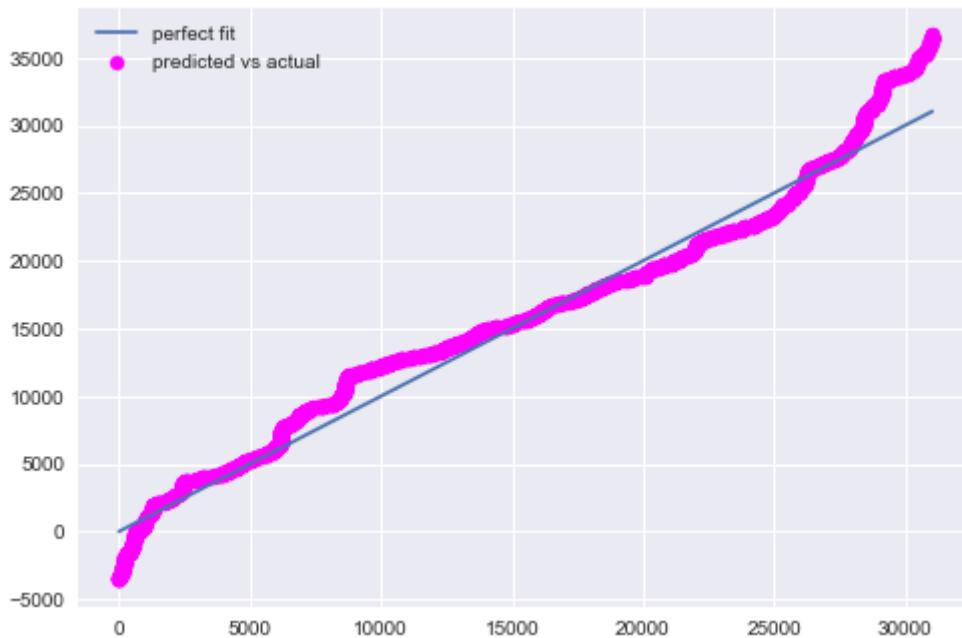
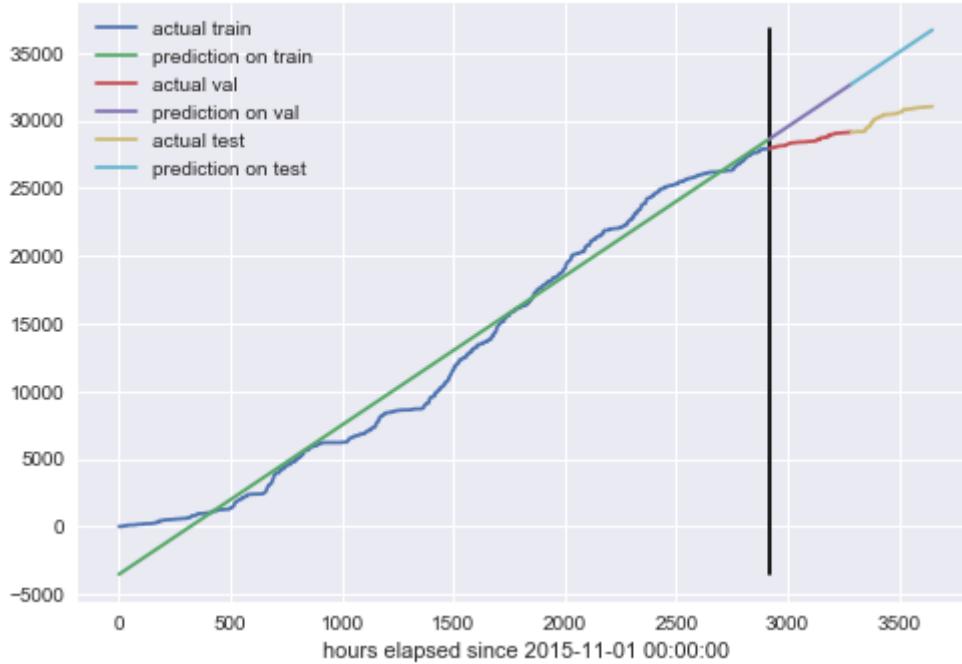
meterid 2129; r2_train: 0.987; r2_val: -41.434; r2_test: -87.716



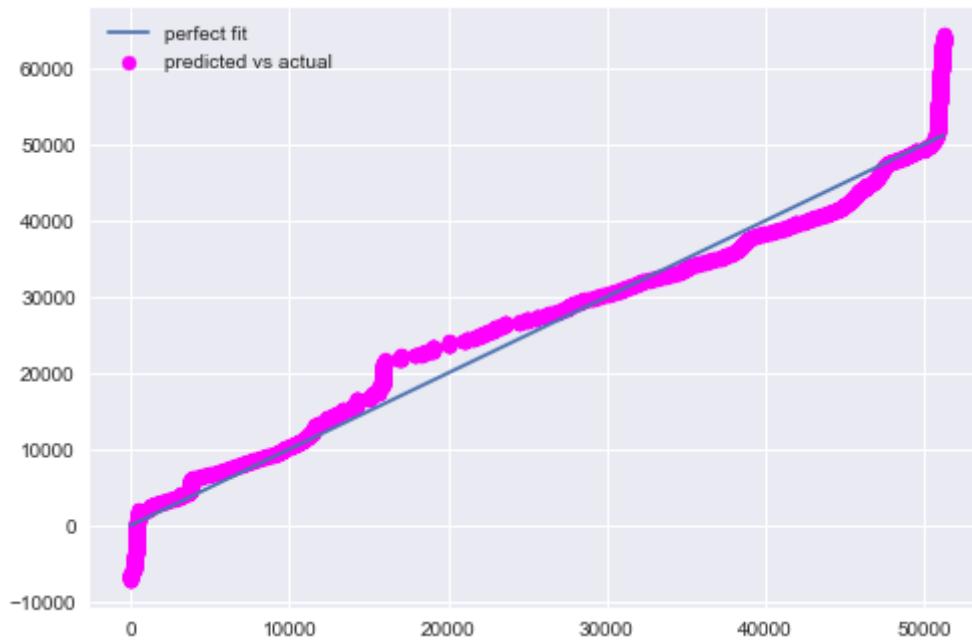
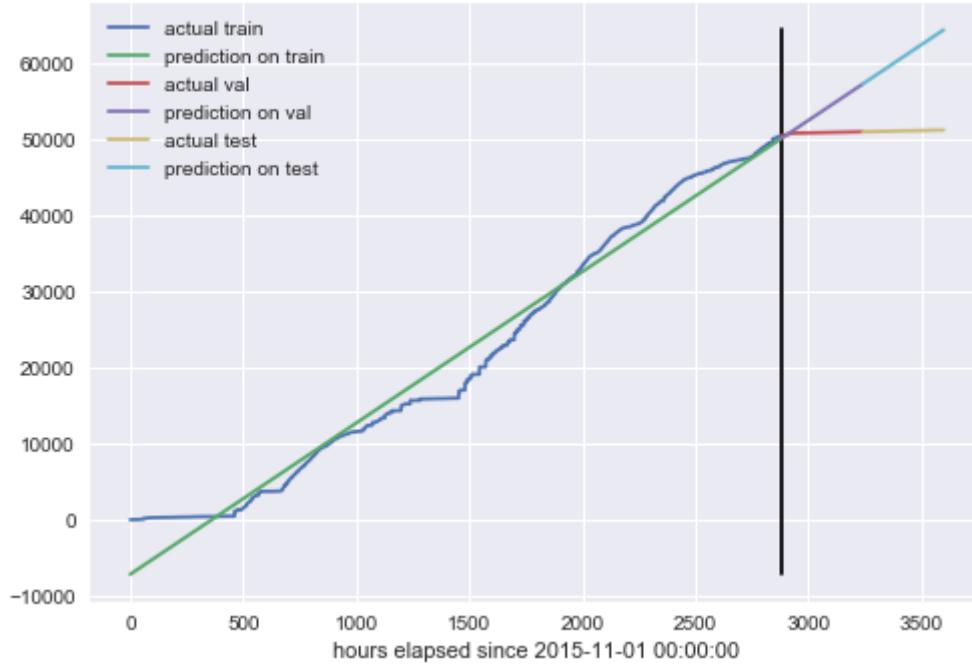
meterid 2233; r2_train: 0.883; r2_val: -21.569; r2_test: -4.272



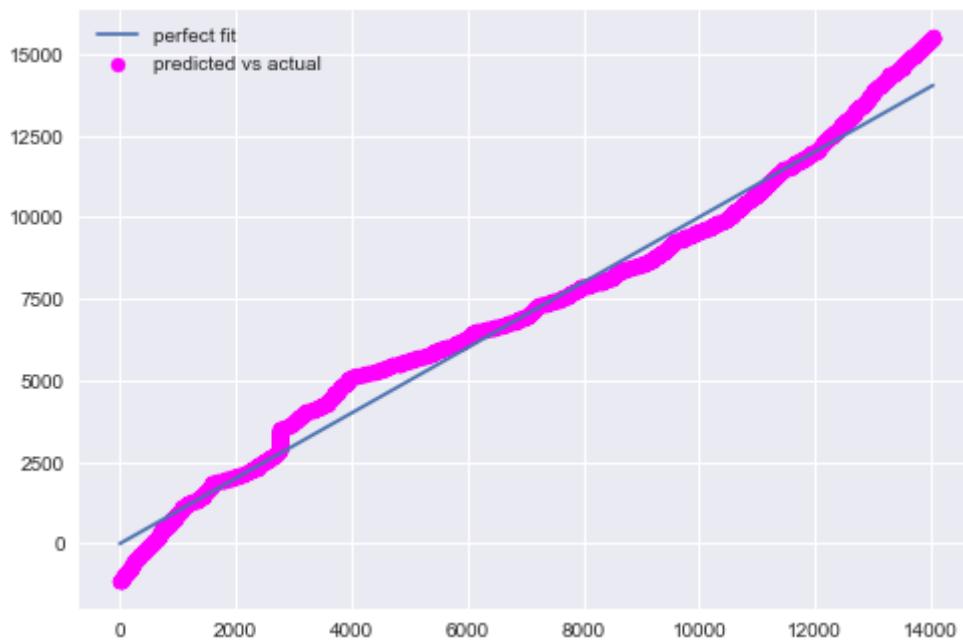
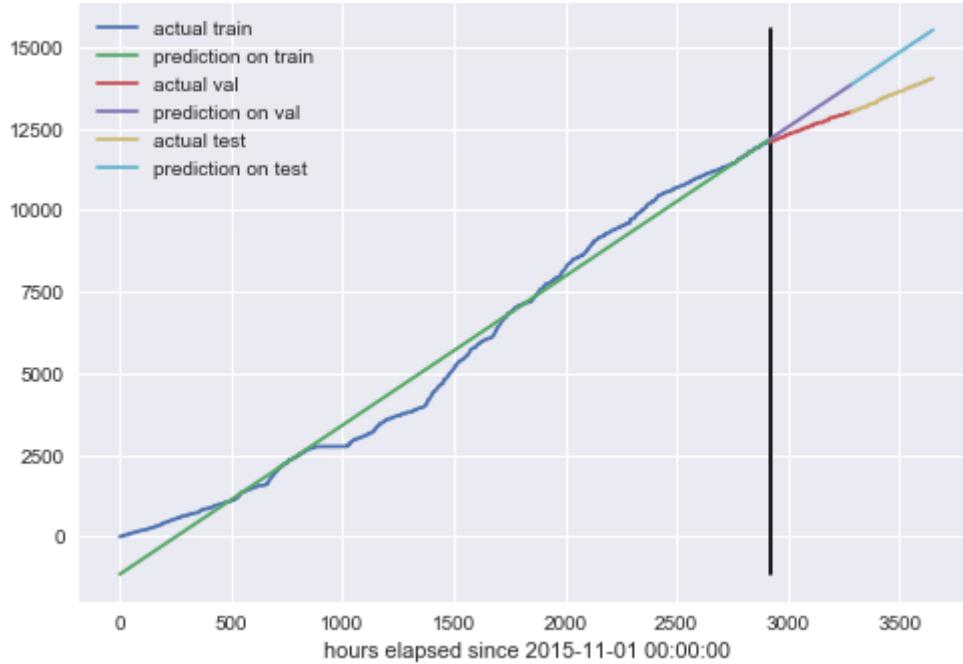
meterid 2335; r2_train: 0.981; r2_val: -36.752; r2_test: -44.637



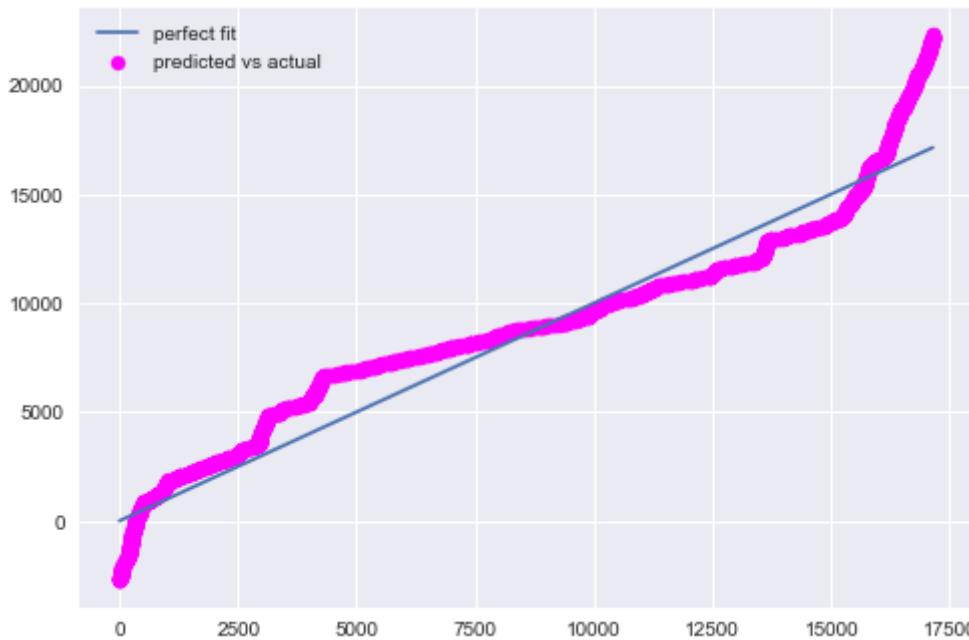
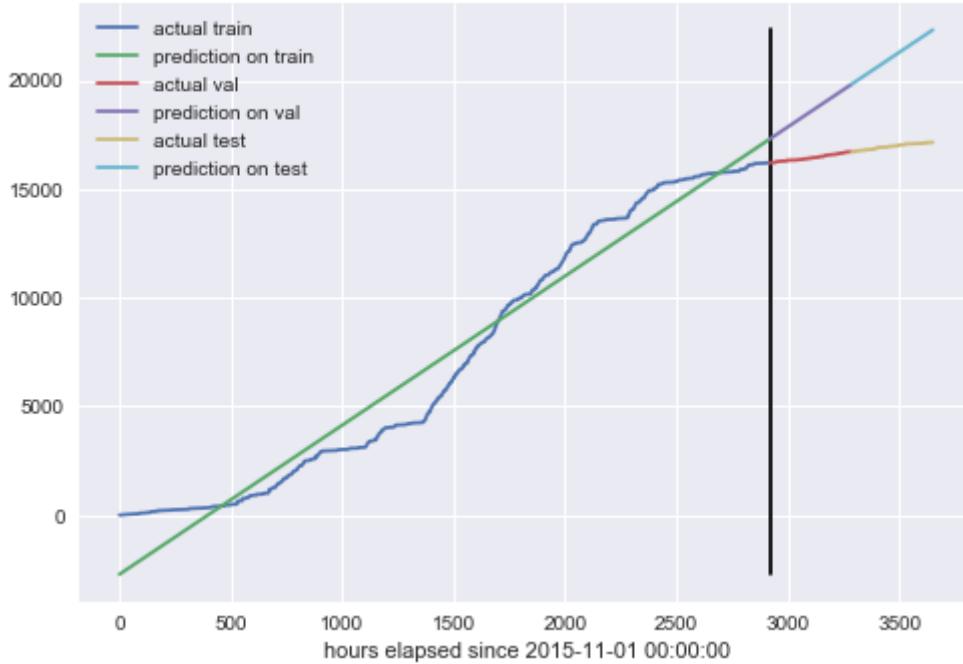
meterid 2378; r2_train: 0.978; r2_val: -837.877; r2_test: -19777.933



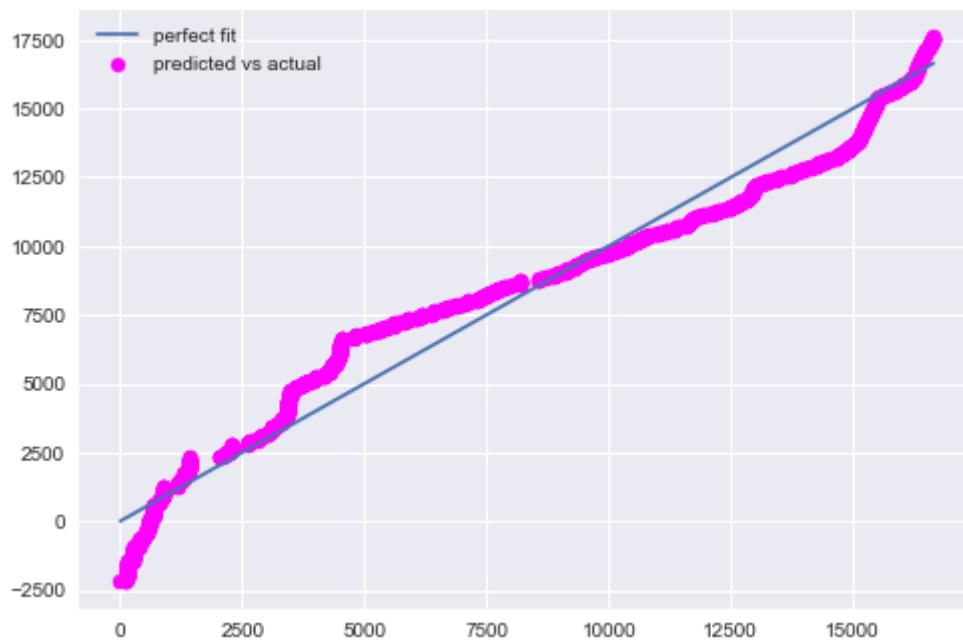
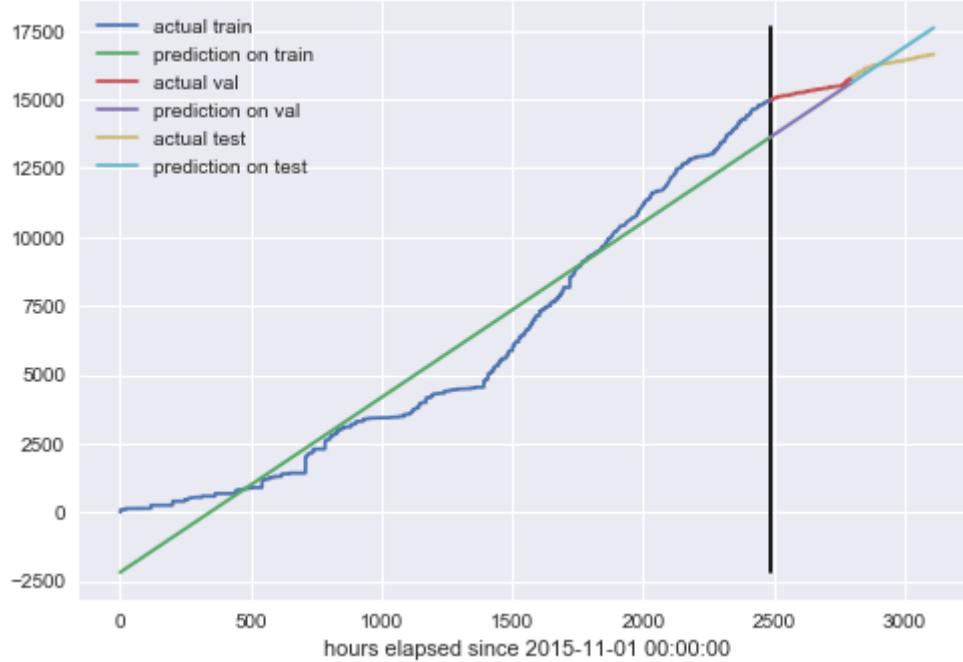
meterid 2449; r2_train: 0.984; r2_val: -2.360; r2_test: -13.860



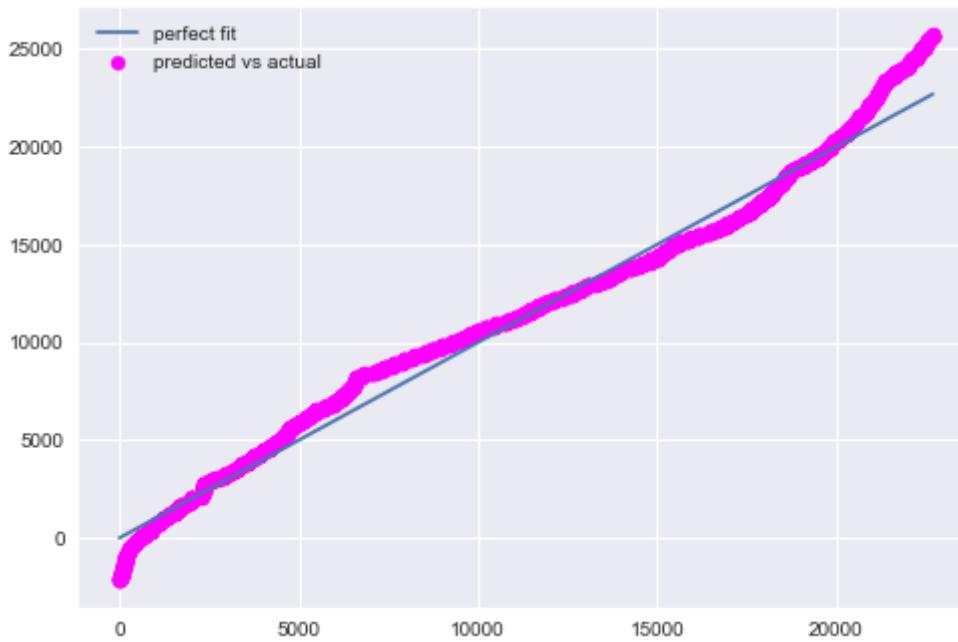
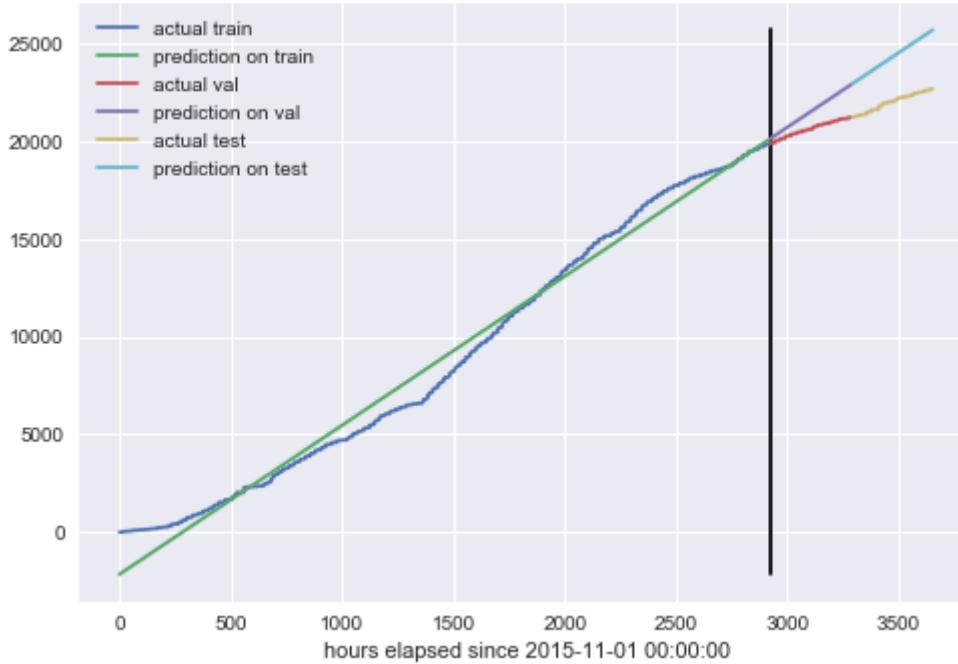
meterid 2461; r2_train: 0.962; r2_val: -215.856; r2_test: -975.750



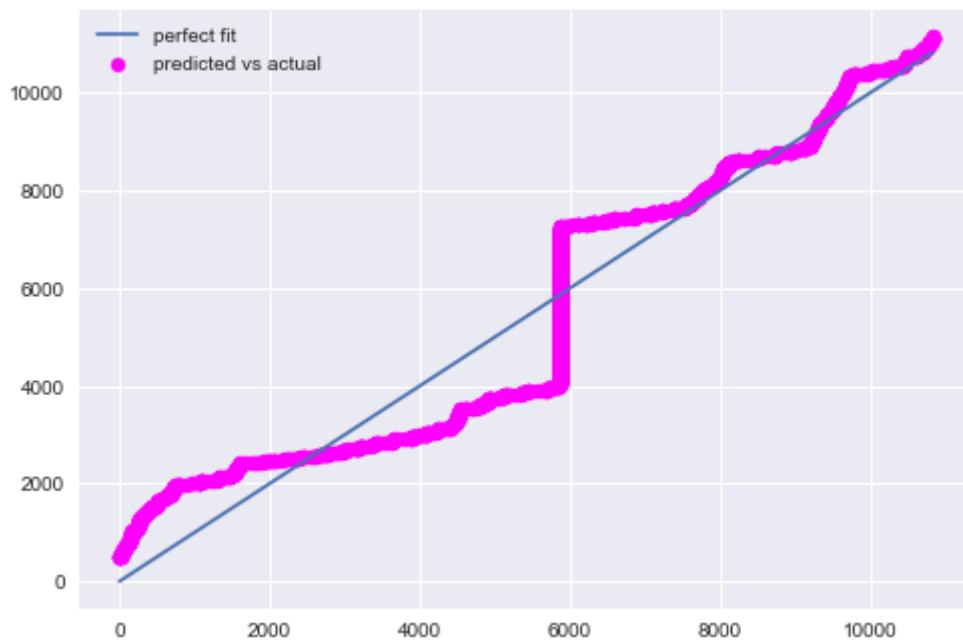
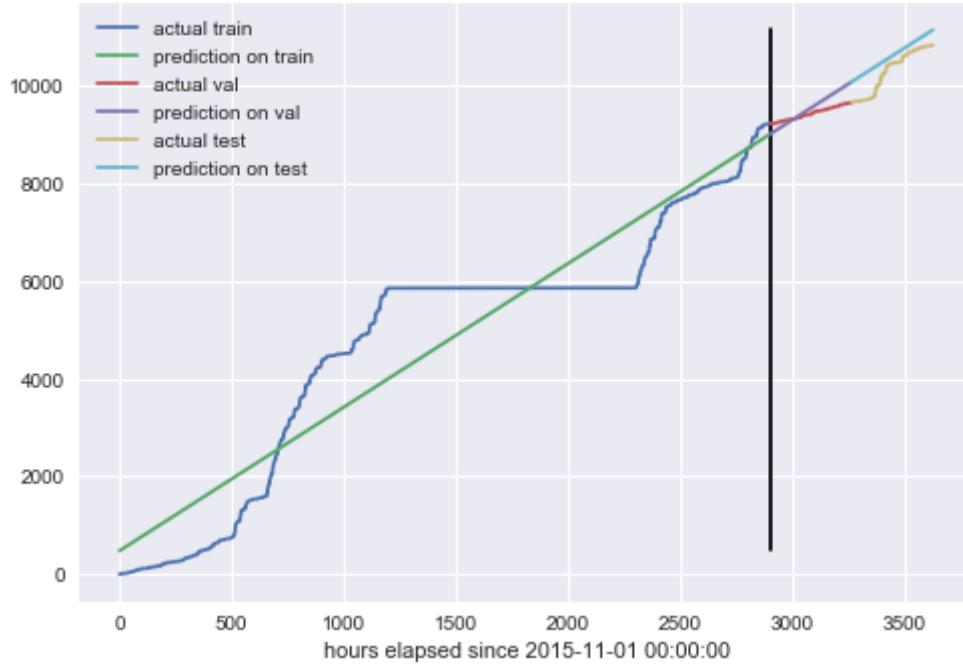
meterid 2470; r2_train: 0.951; r2_val: -18.129; r2_test: -4.091



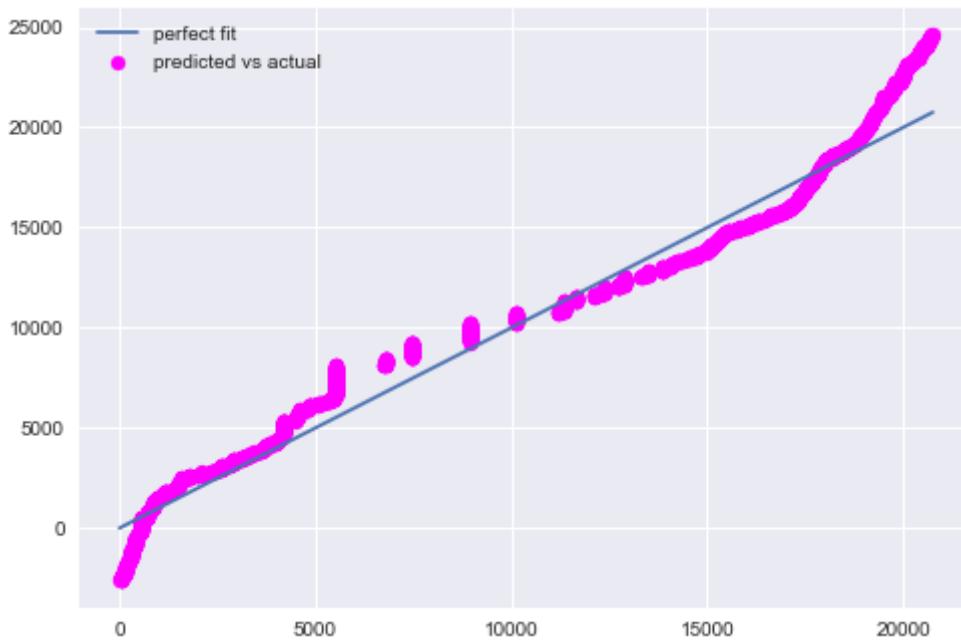
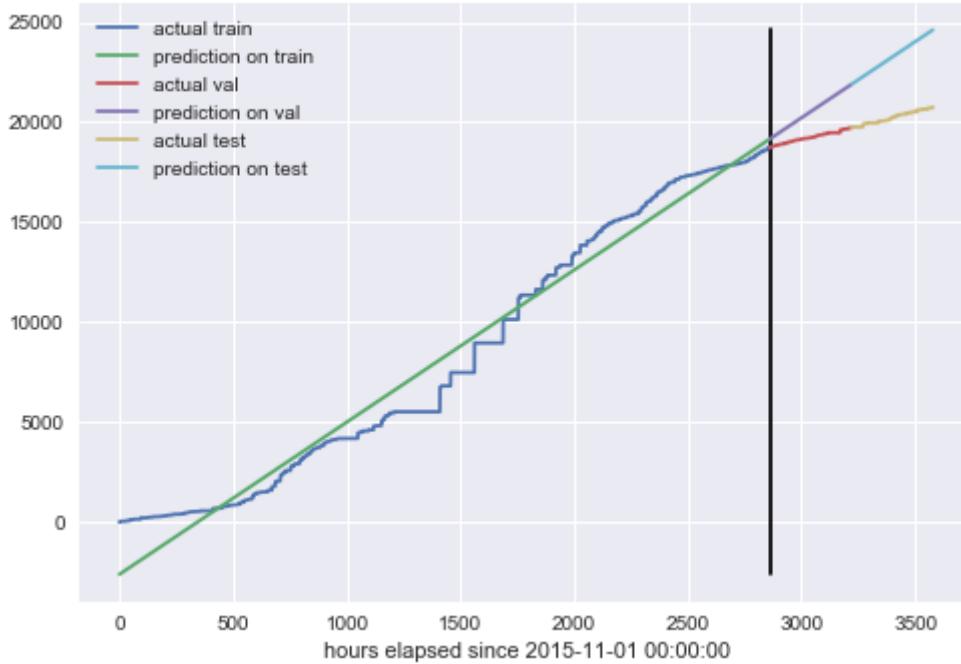
meterid 2575; r2_train: 0.986; r2_val: -5.092; r2_test: -25.413



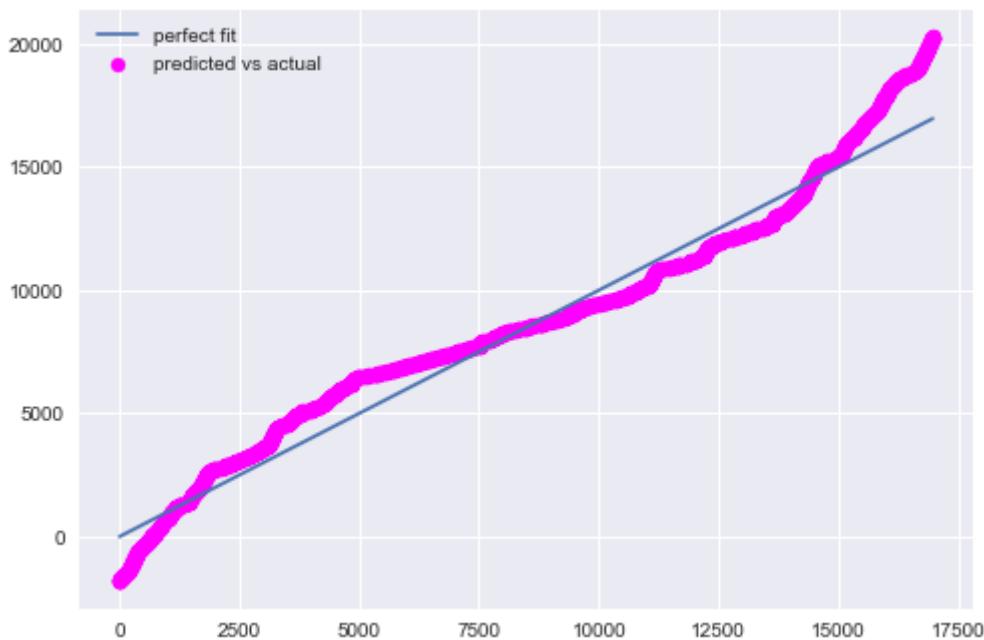
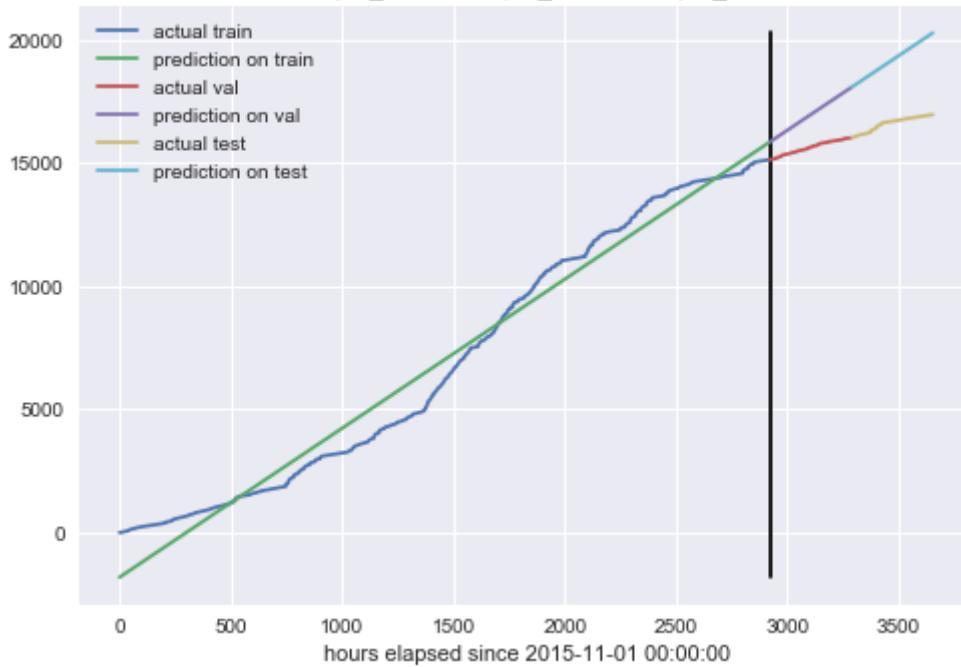
meterid 2638; r2_train: 0.884; r2_val: -1.431; r2_test: 0.356

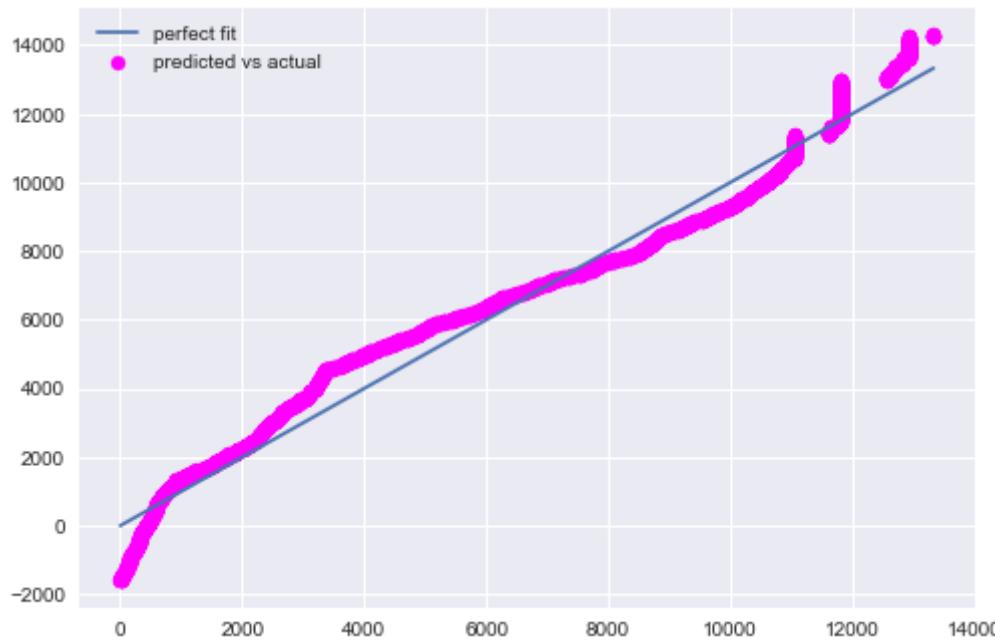
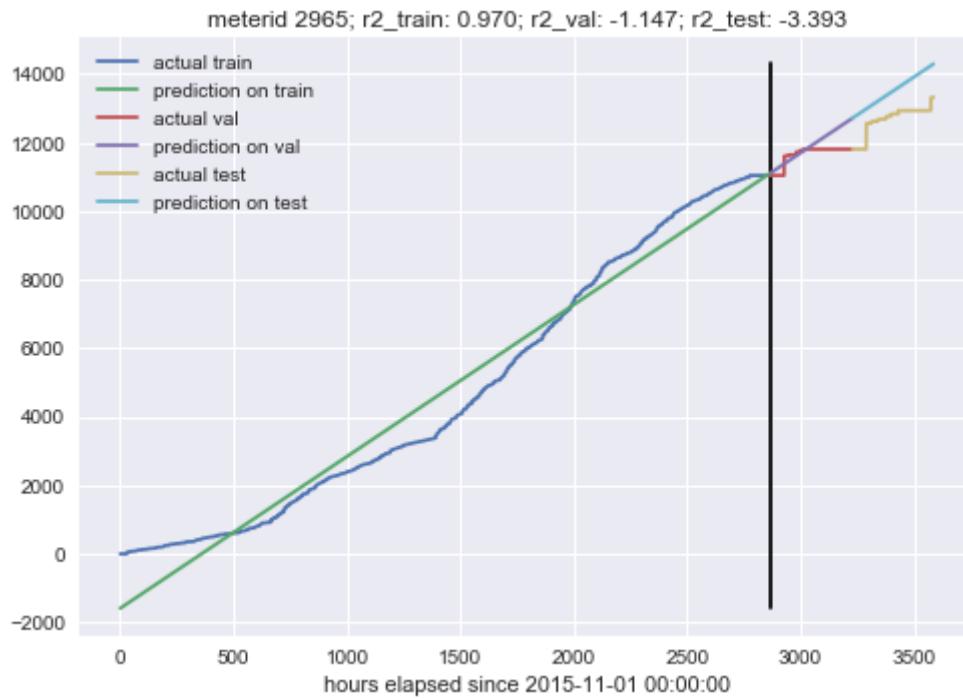


meterid 2818; r2_train: 0.973; r2_val: -25.743; r2_test: -94.954

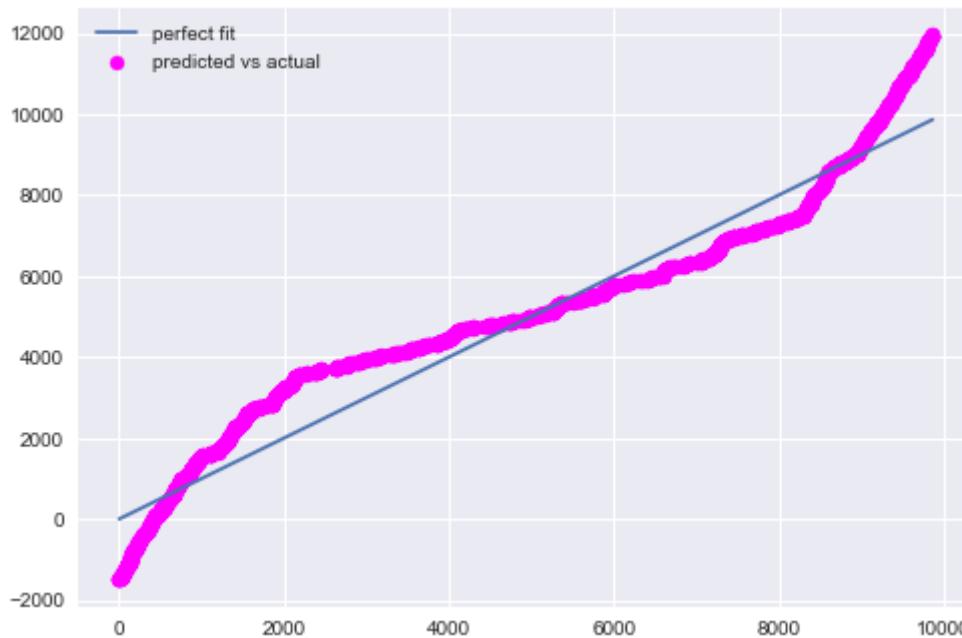
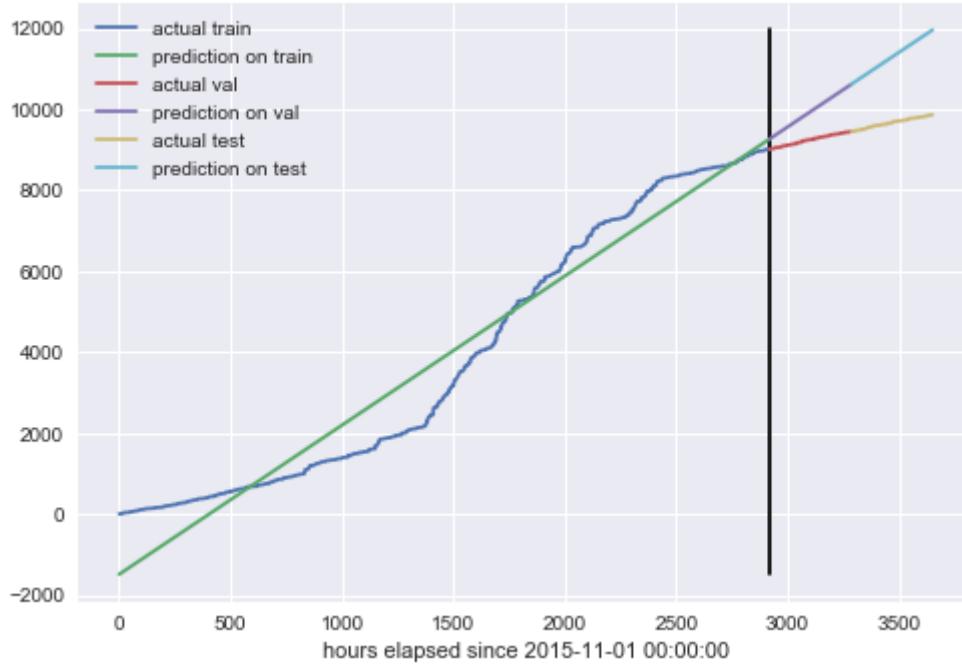


meterid 2945; r2_train: 0.977; r2_val: -25.531; r2_test: -78.714

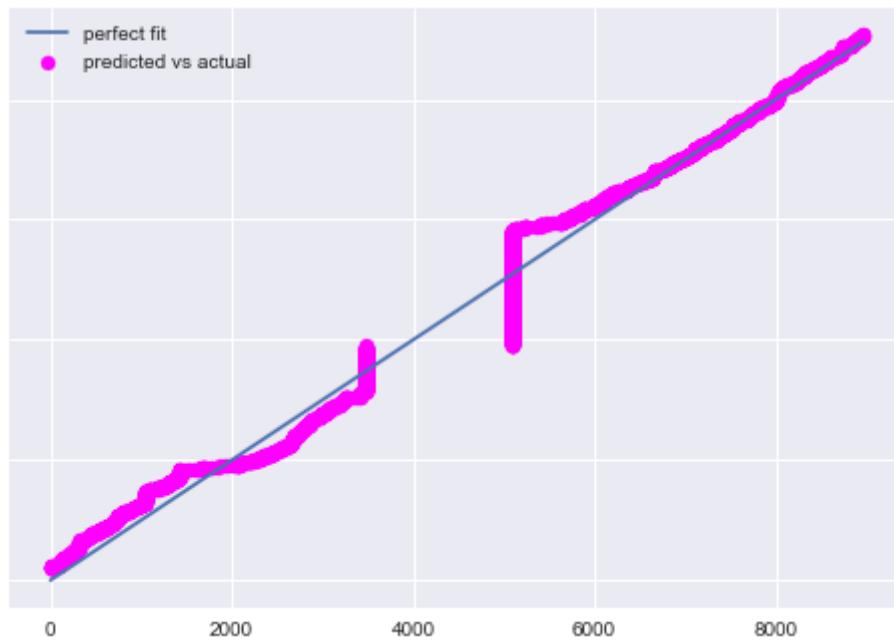
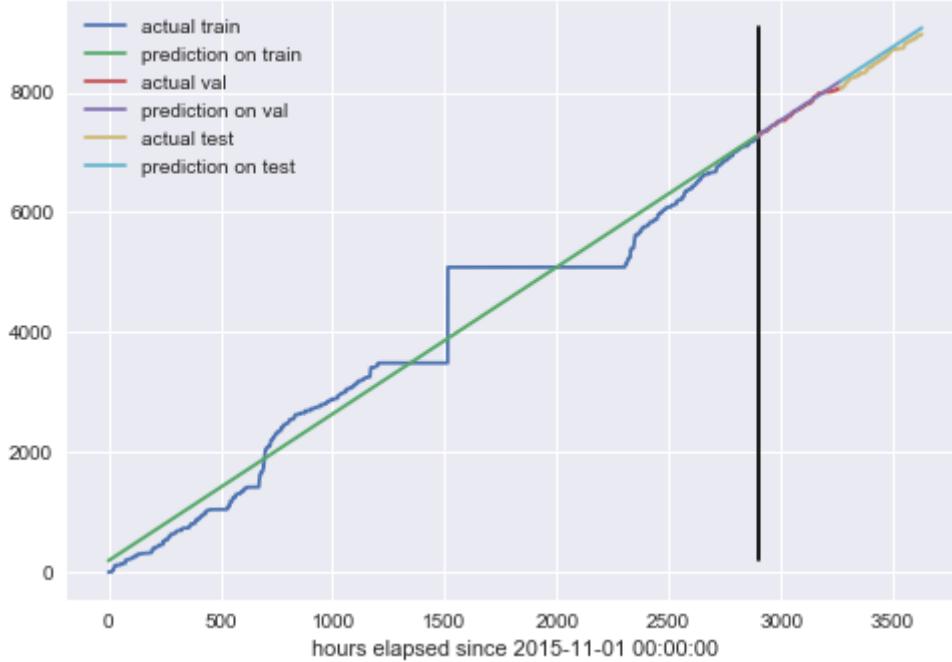




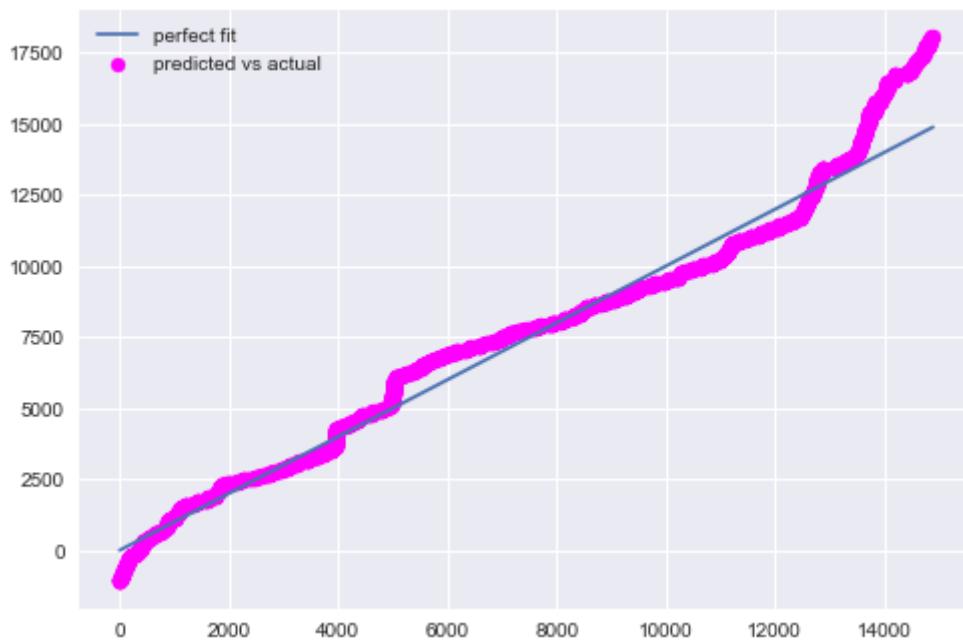
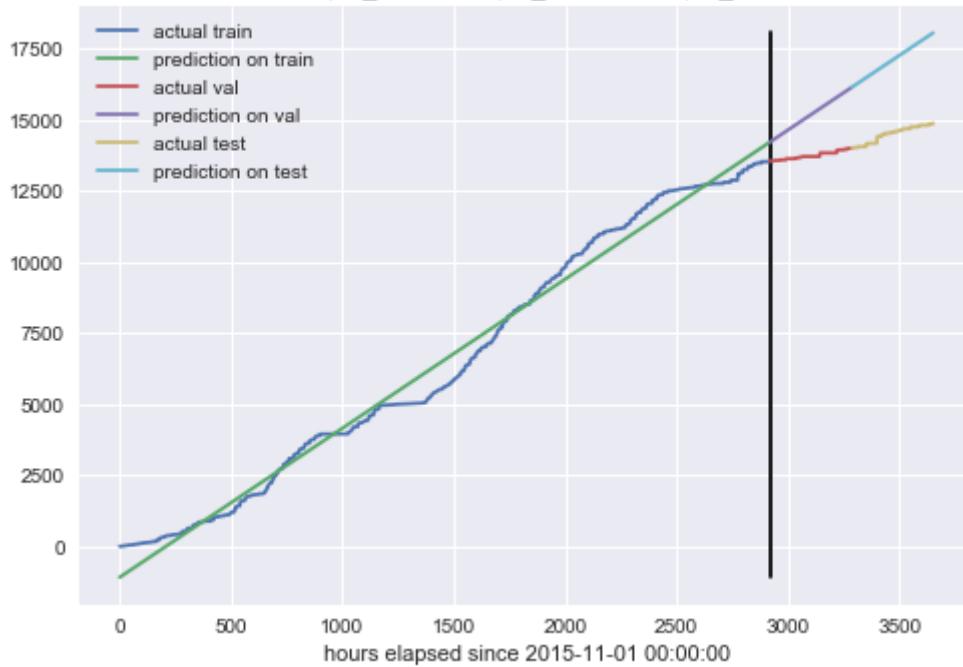
meterid 2980; r2_train: 0.952; r2_val: -31.560; r2_test: -181.519



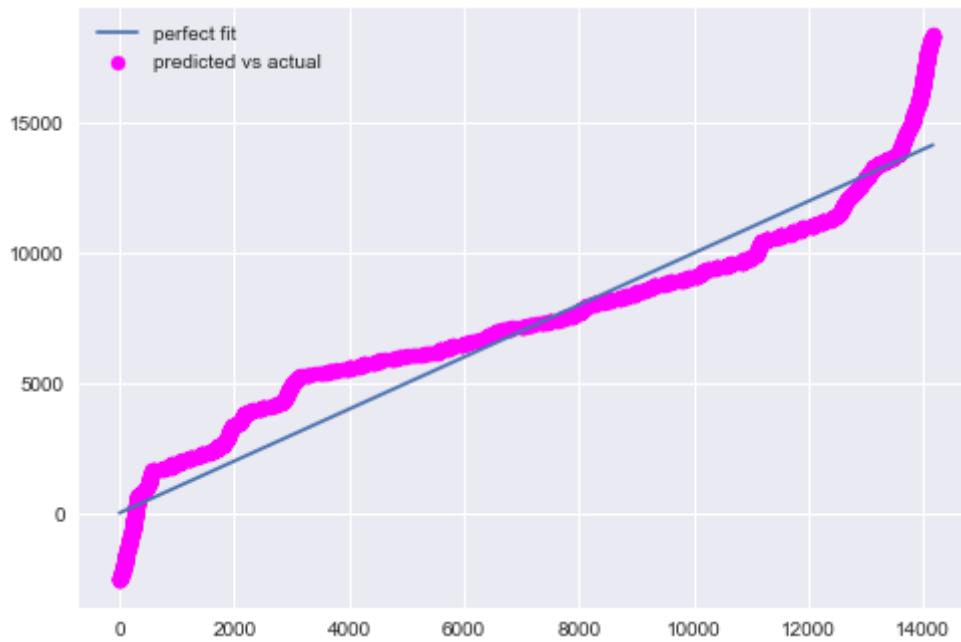
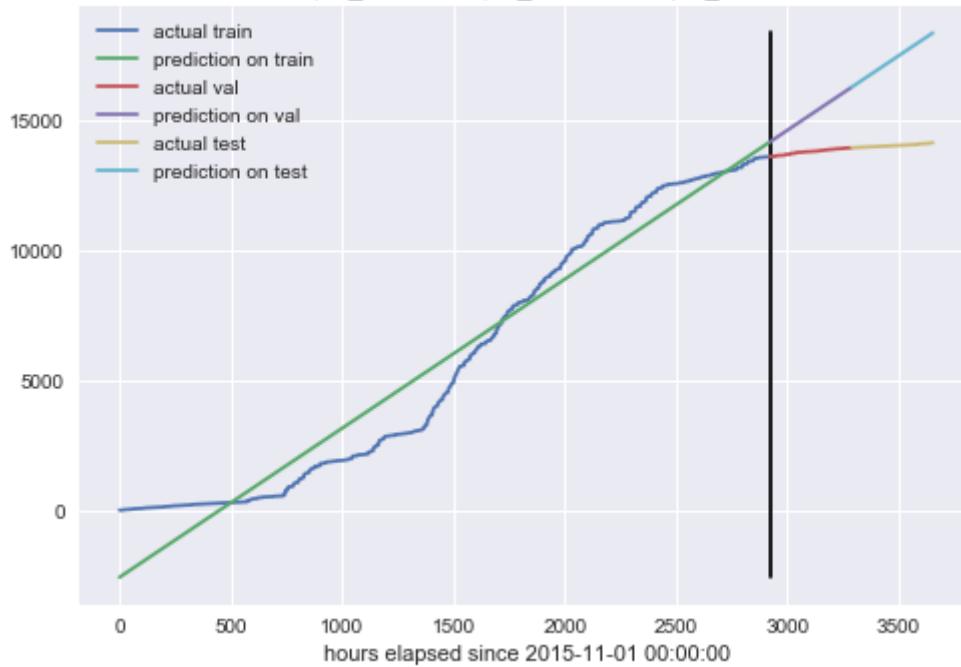
meterid 3039; r2_train: 0.966; r2_val: 0.975; r2_test: 0.874



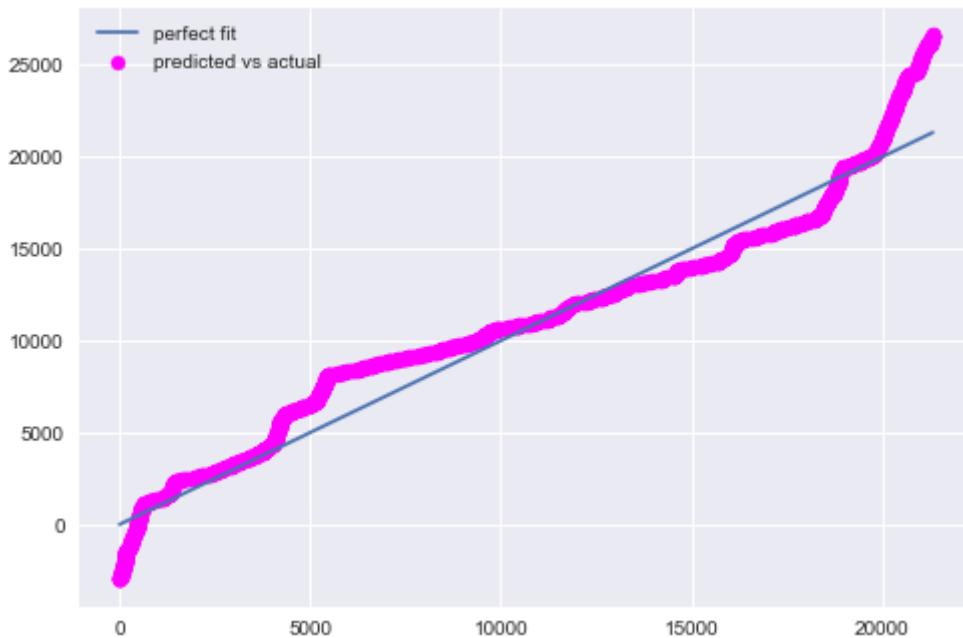
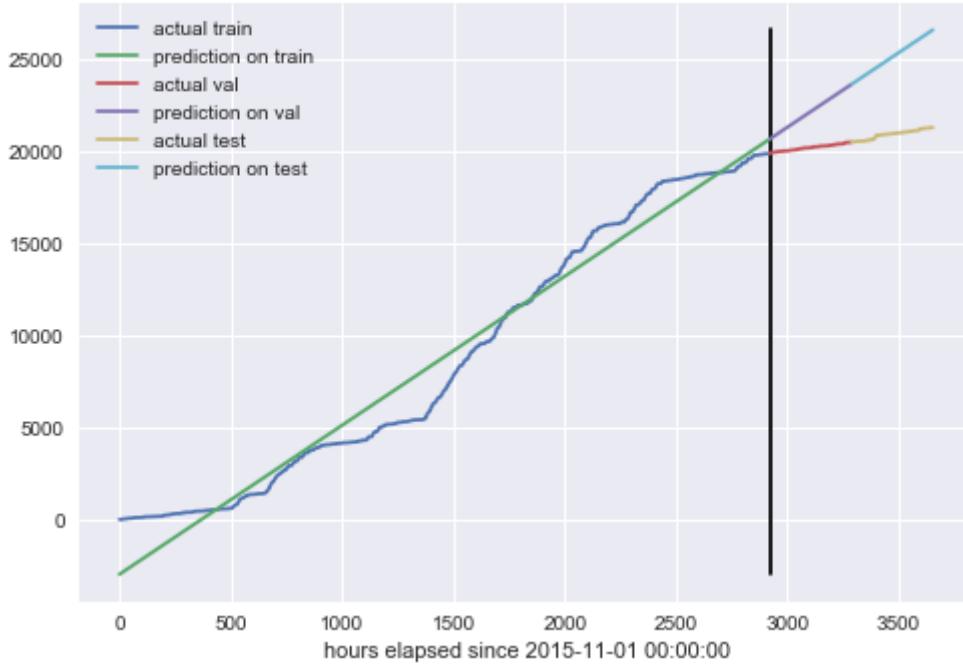
meterid 3134; r2_train: 0.988; r2_val: -121.916; r2_test: -83.002



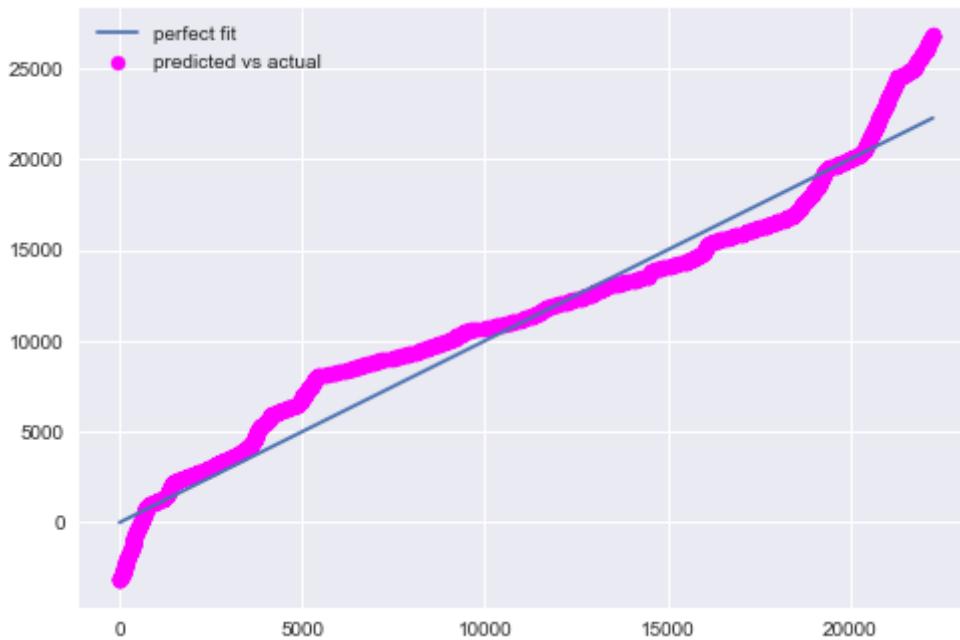
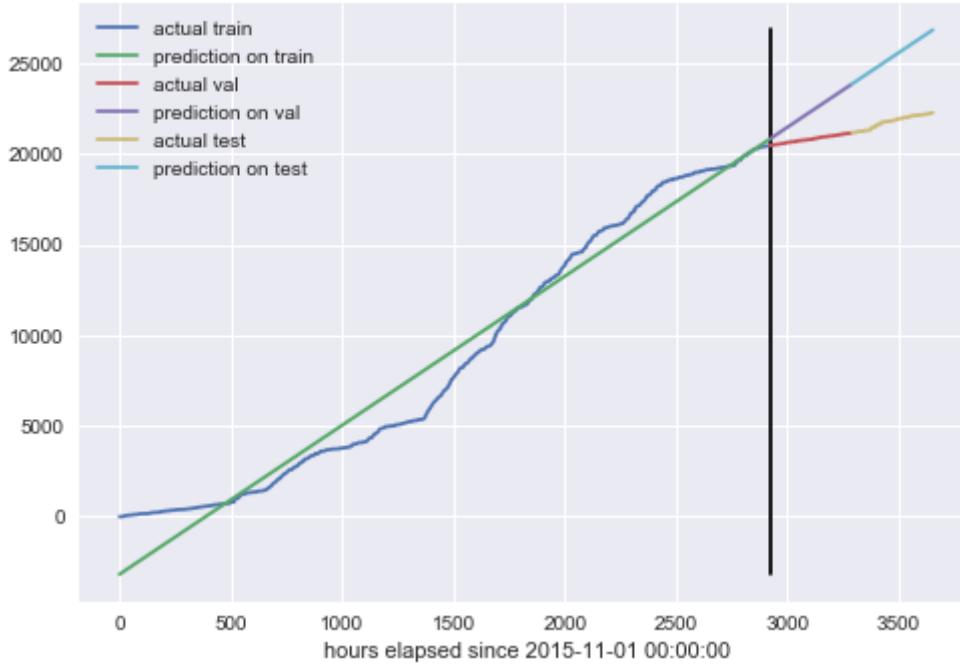
meterid 3310; r2_train: 0.953; r2_val: -223.705; r2_test: -4063.628



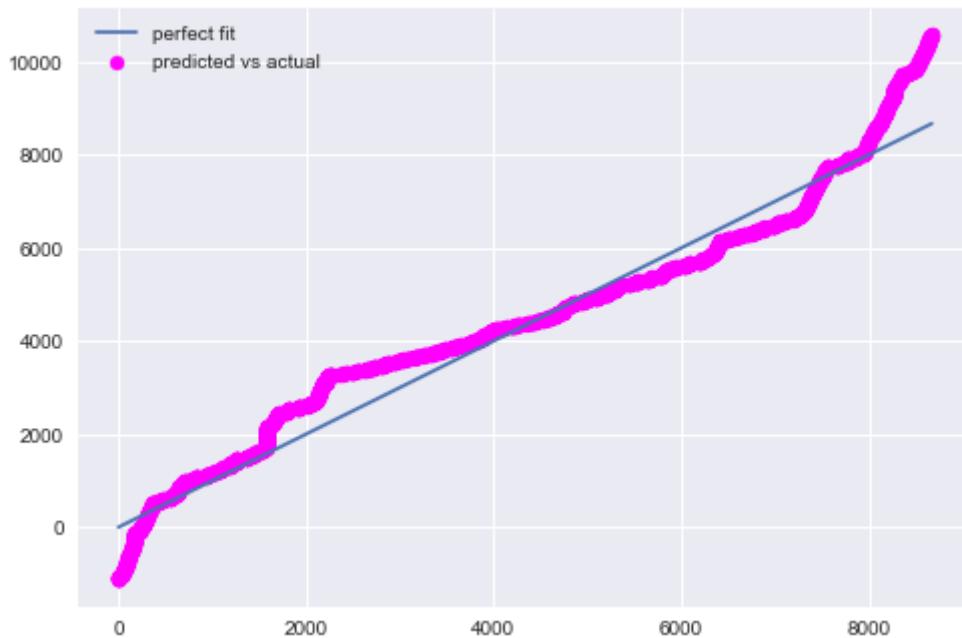
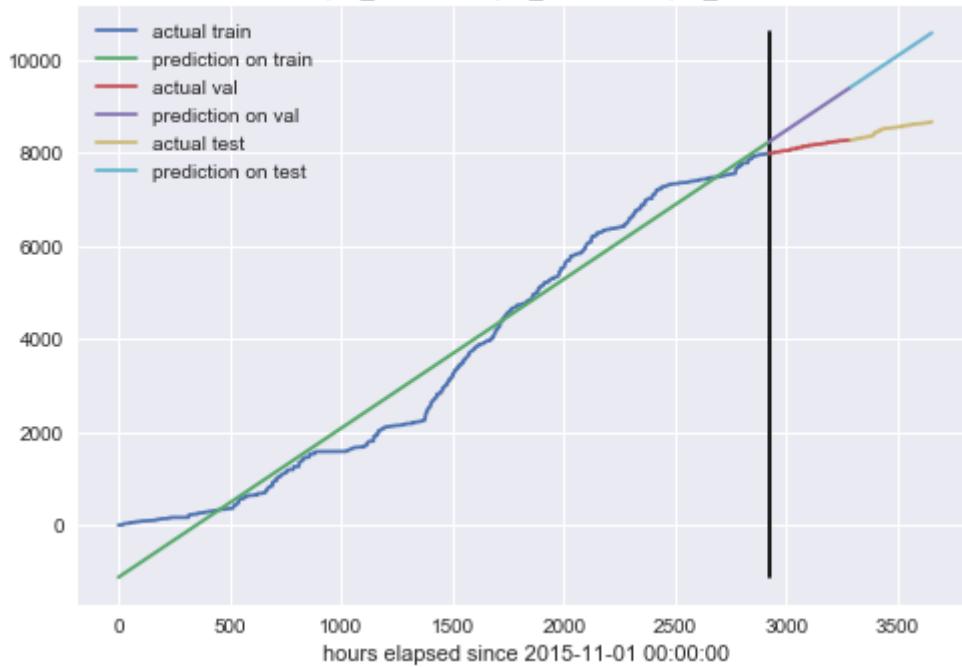
meterid 3367; r2_train: 0.970; r2_val: -158.717; r2_test: -295.600

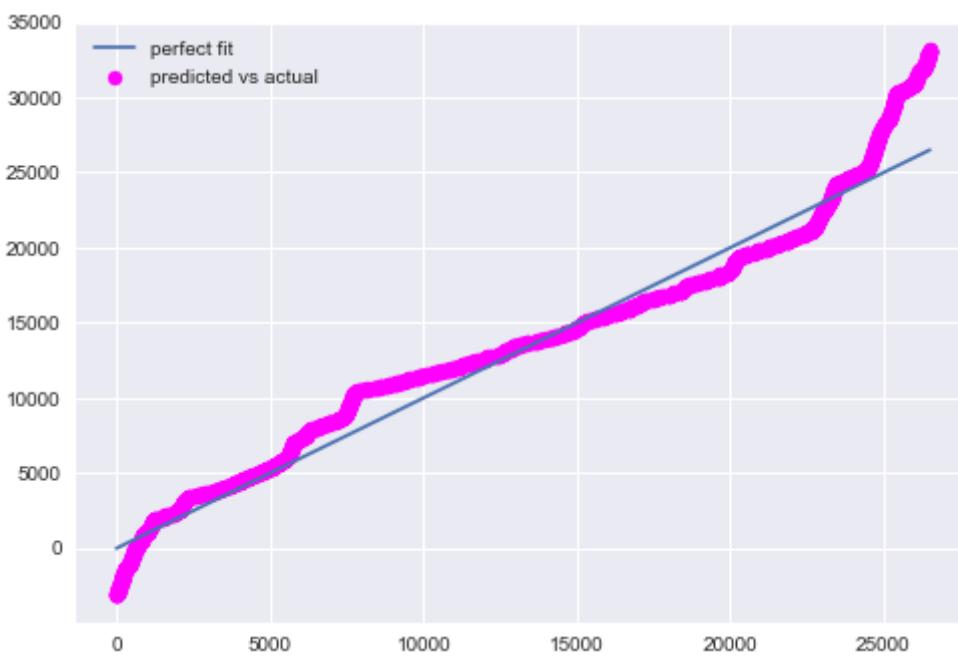
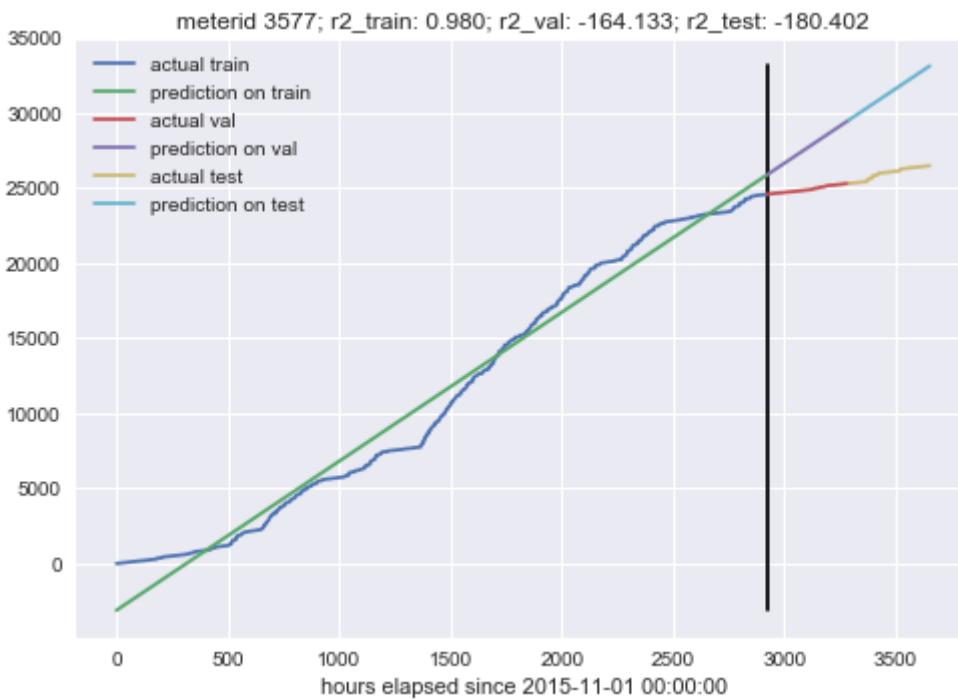


meterid 3527; r2_train: 0.967; r2_val: -68.210; r2_test: -102.039

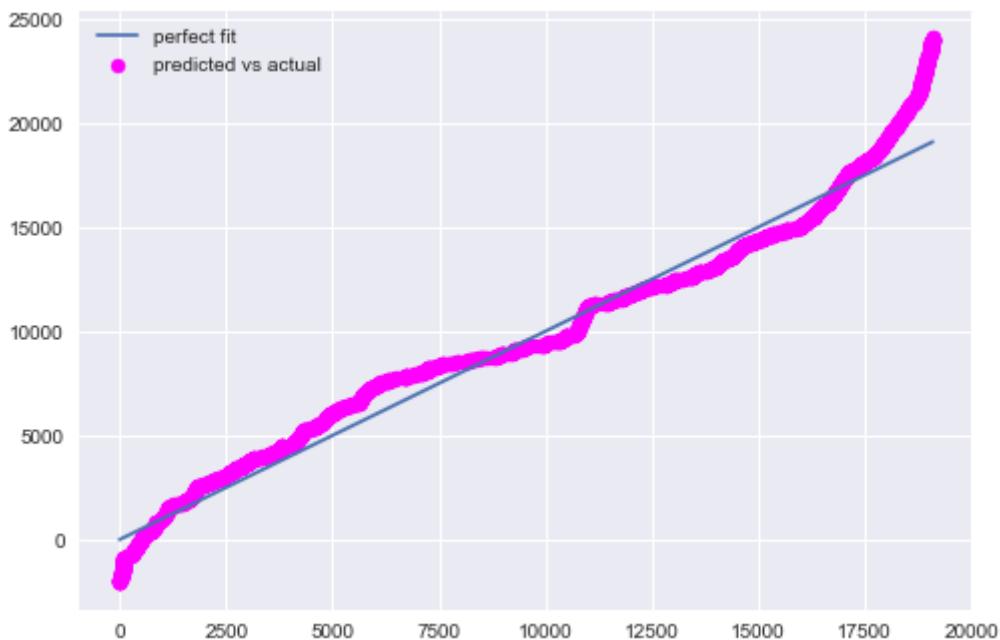
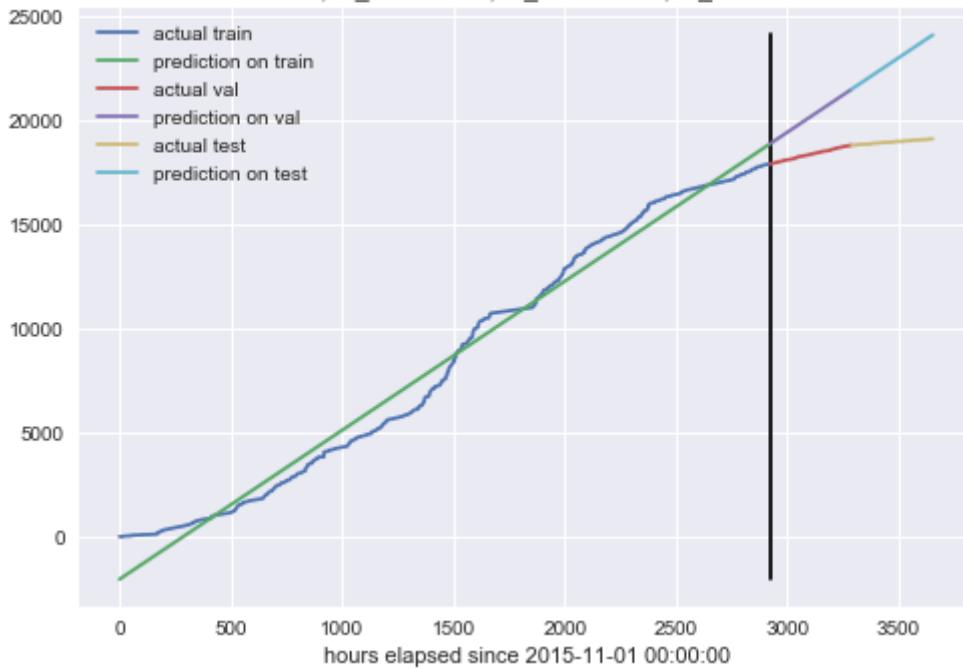


meterid 3544; r2_train: 0.972; r2_val: -65.087; r2_test: -147.166

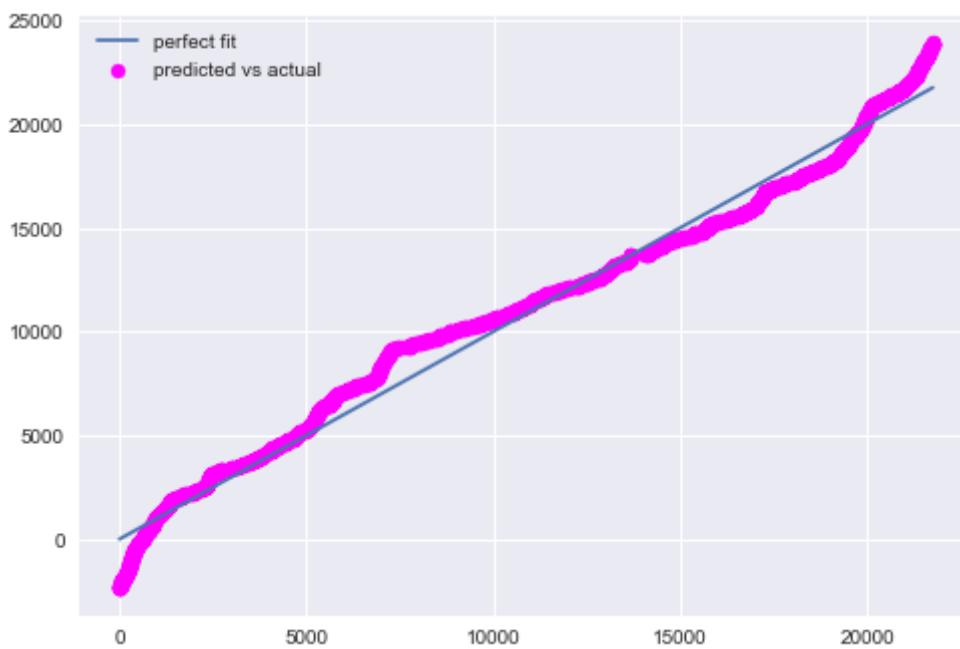
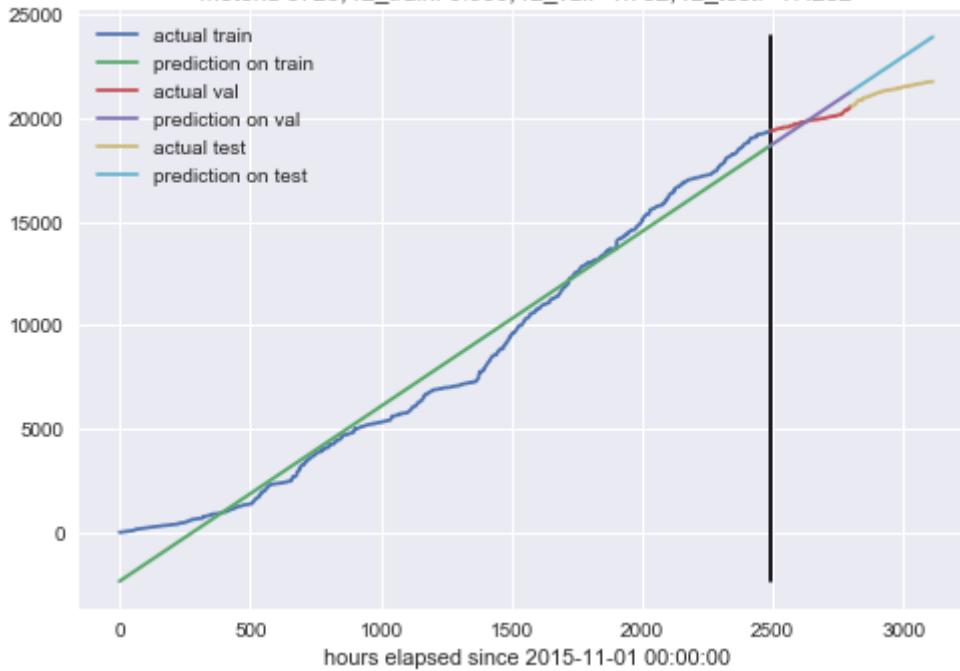




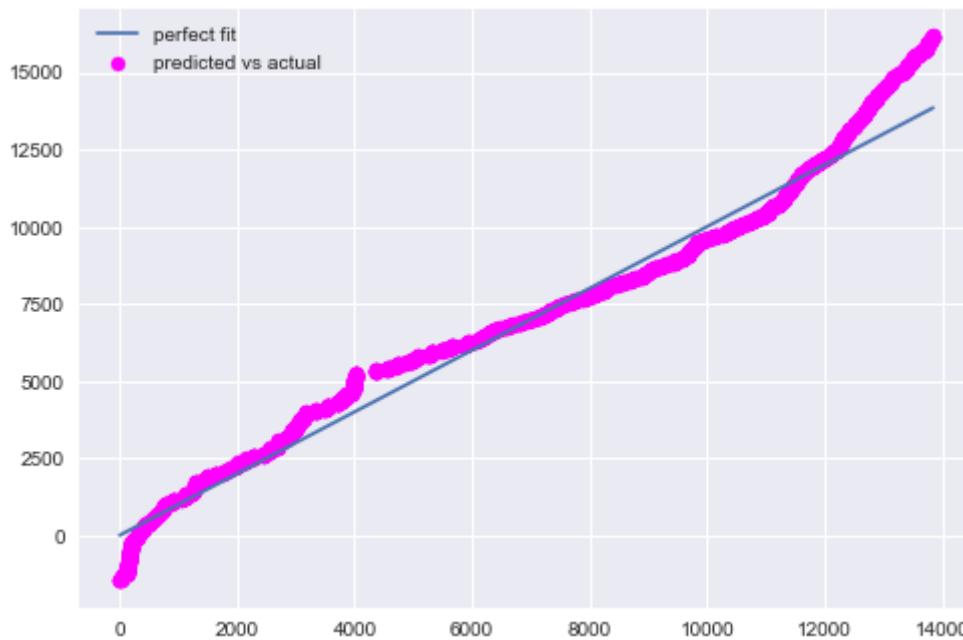
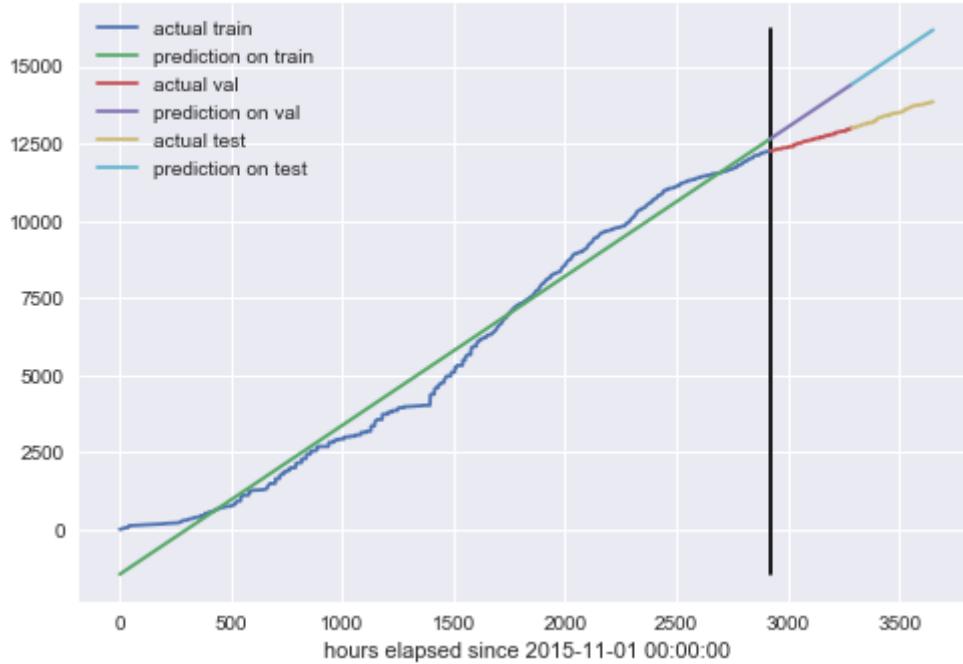
meterid 3635; r2_train: 0.984; r2_val: -50.158; r2_test: -1924.855



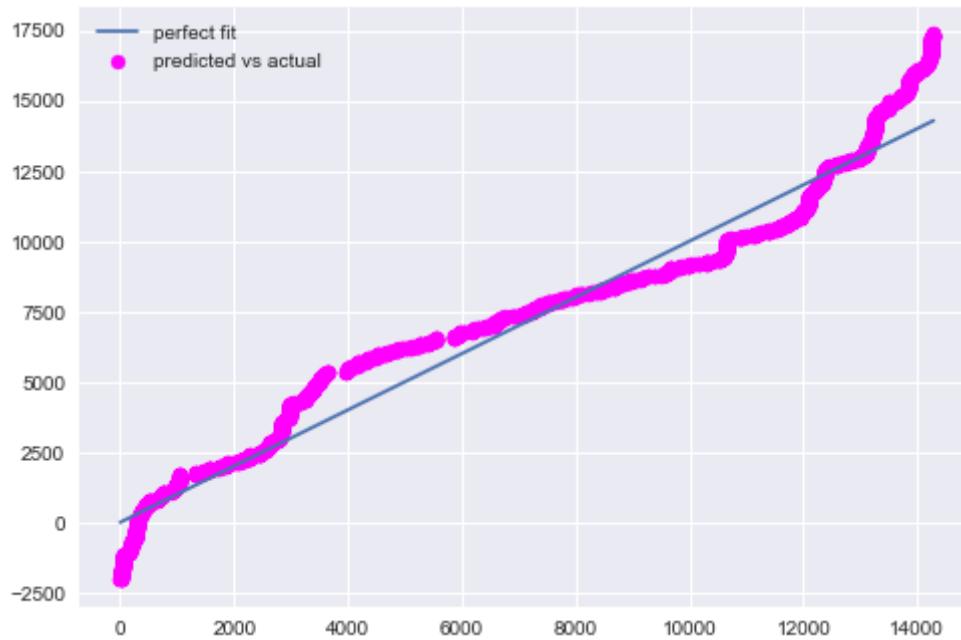
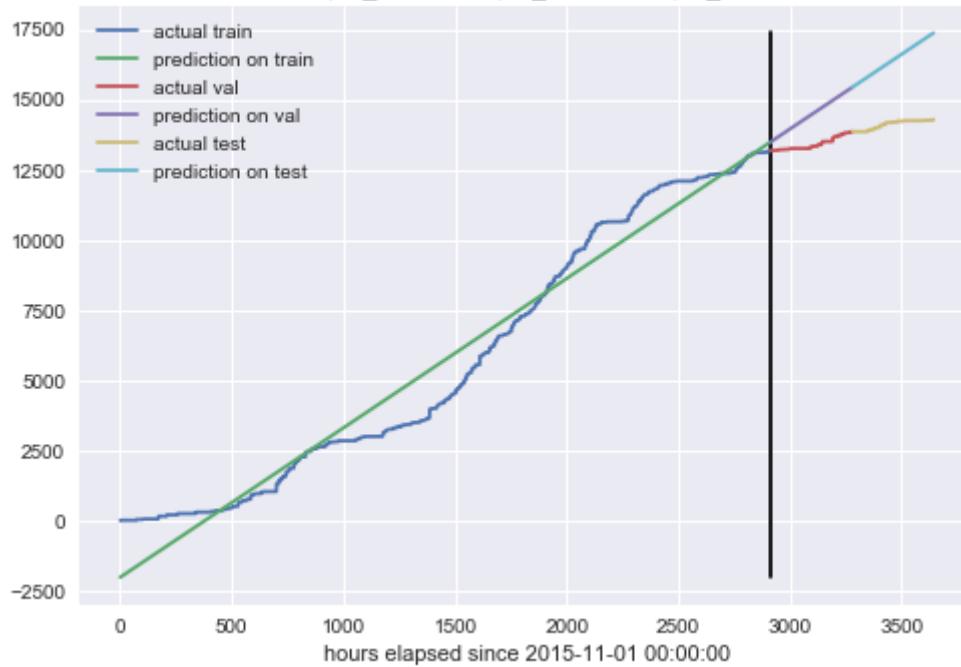
meterid 3723; r2_train: 0.980; r2_val: -1.732; r2_test: -17.282



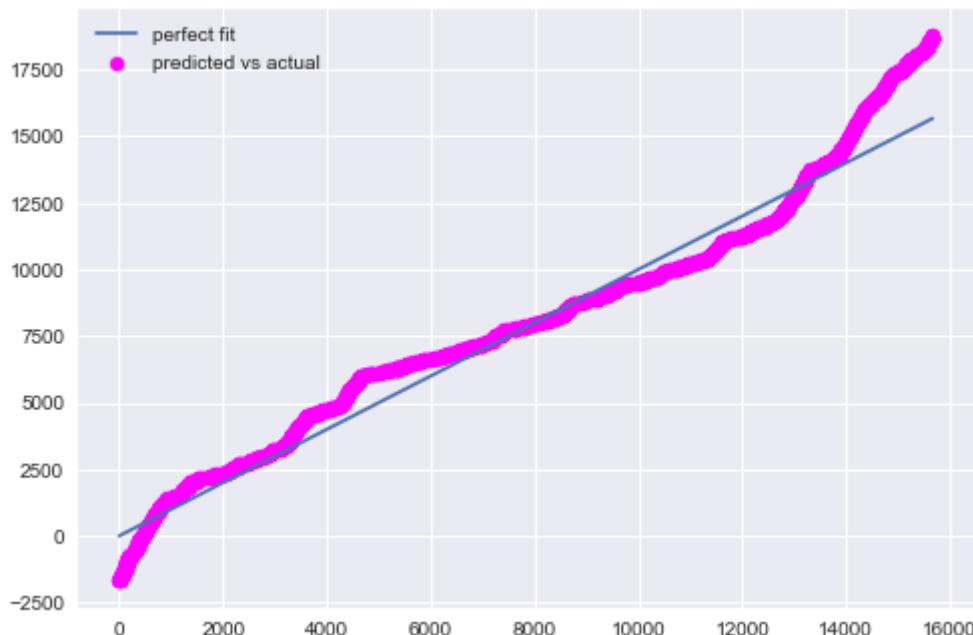
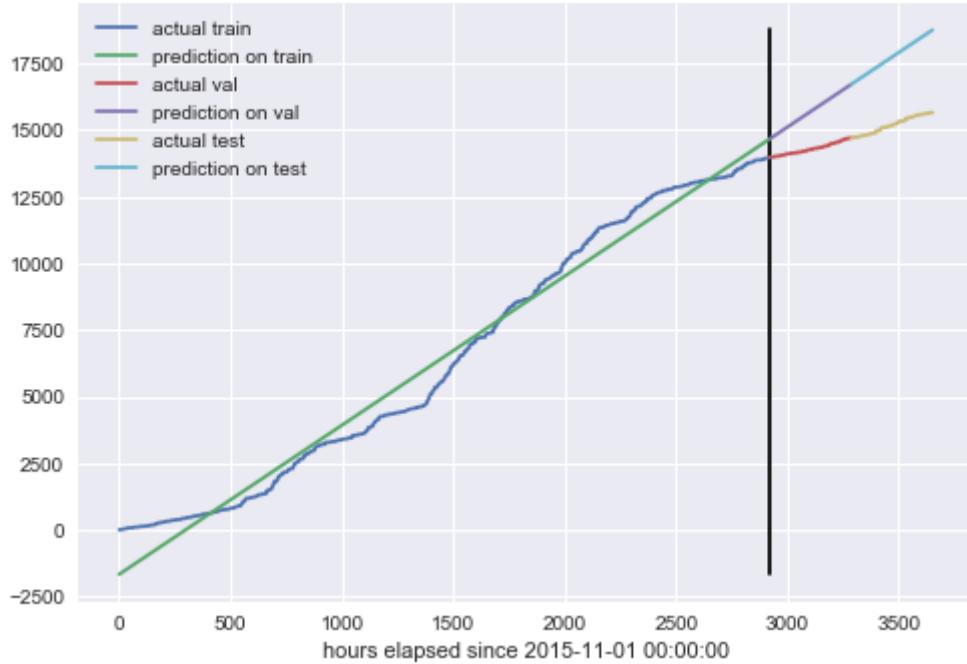
meterid 3778; r2_train: 0.983; r2_val: -19.120; r2_test: -50.304



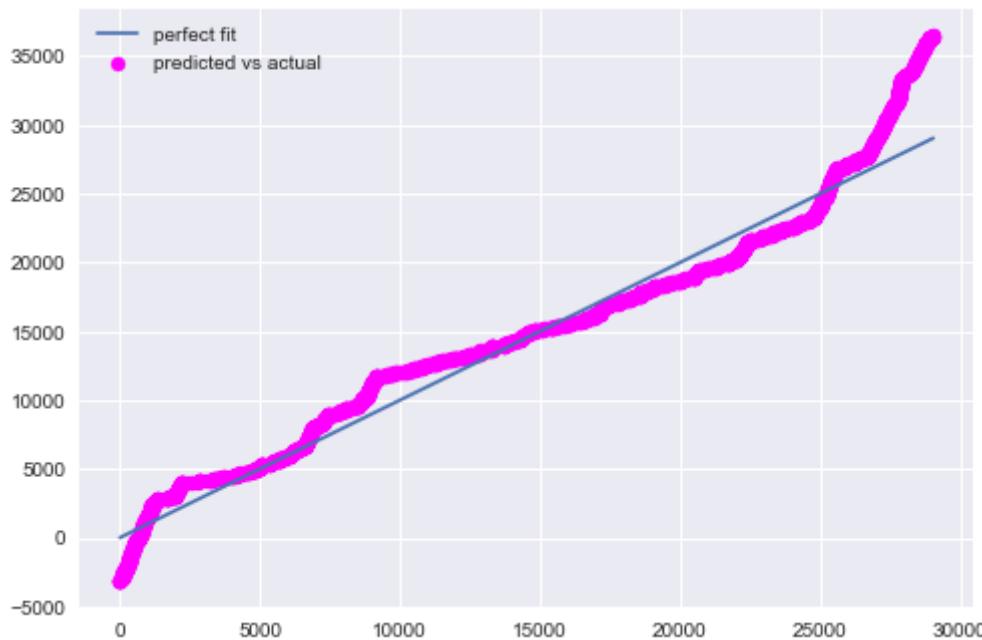
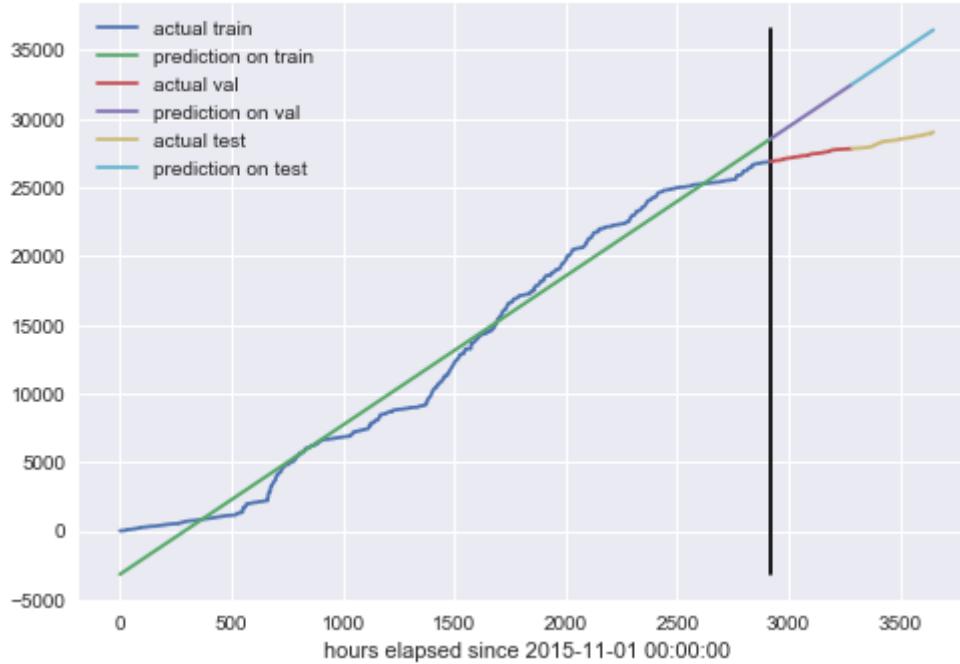
meterid 3849; r2_train: 0.965; r2_val: -25.069; r2_test: -213.019



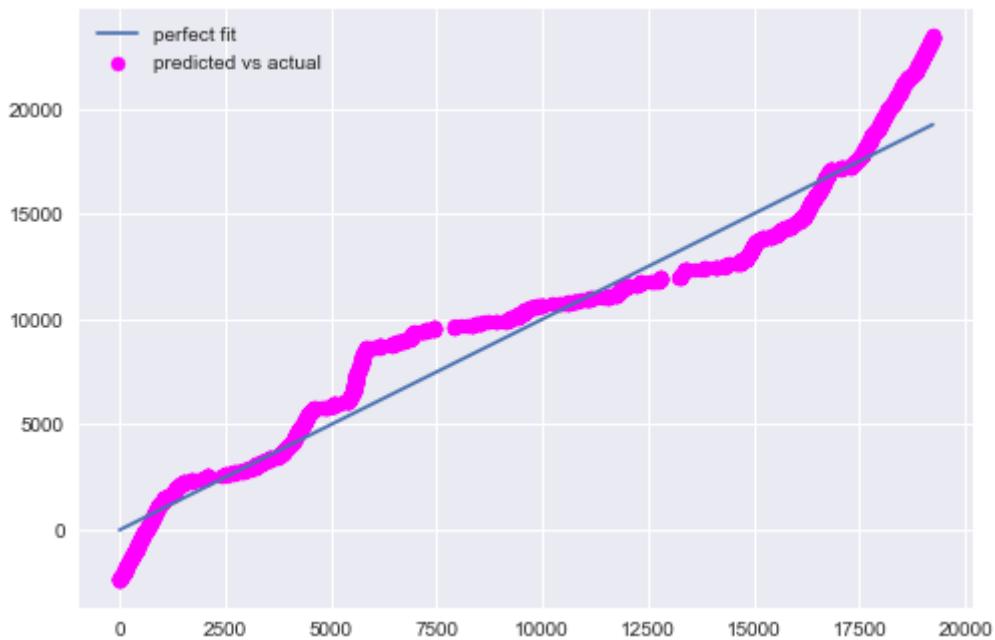
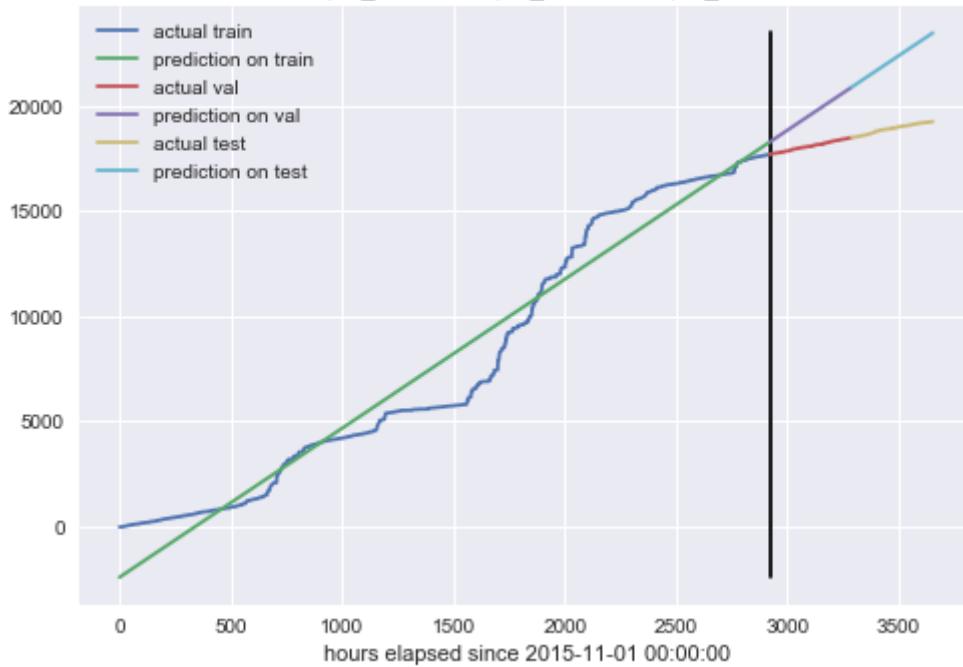
meterid 3893; r2_train: 0.982; r2_val: -44.806; r2_test: -64.665



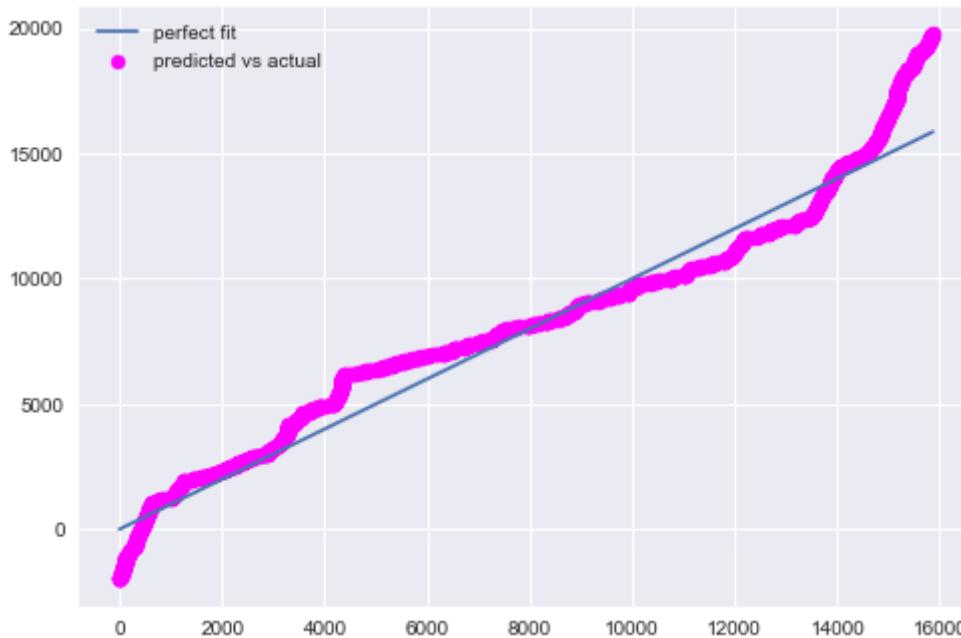
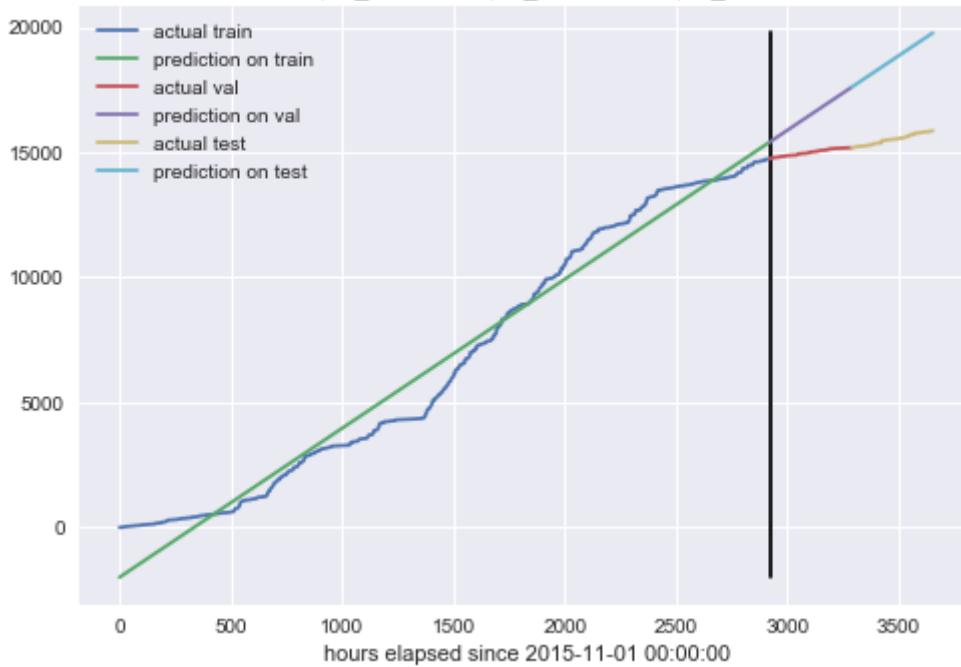
meterid 3918; r2_train: 0.982; r2_val: -121.693; r2_test: -293.831



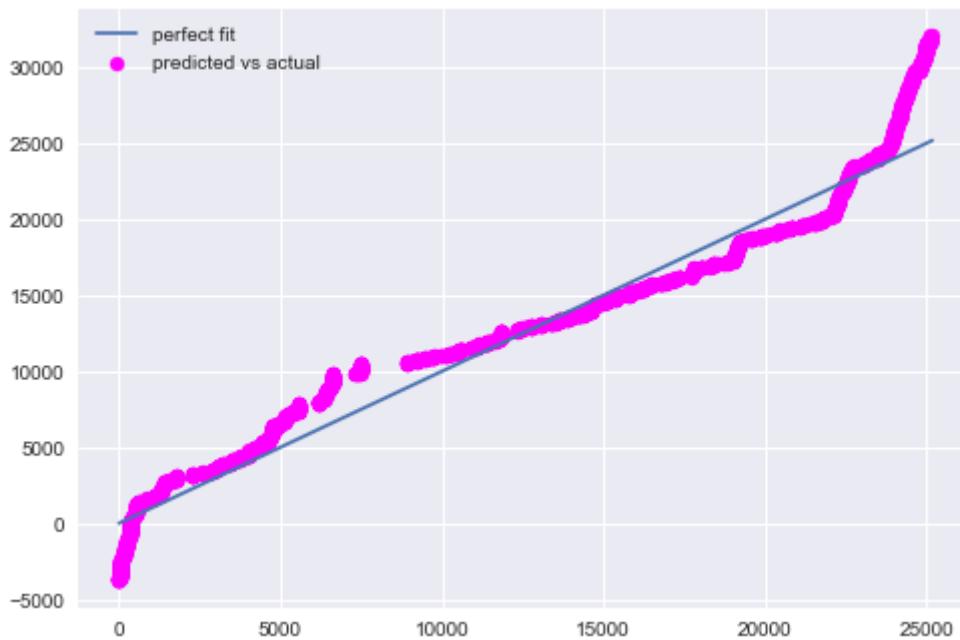
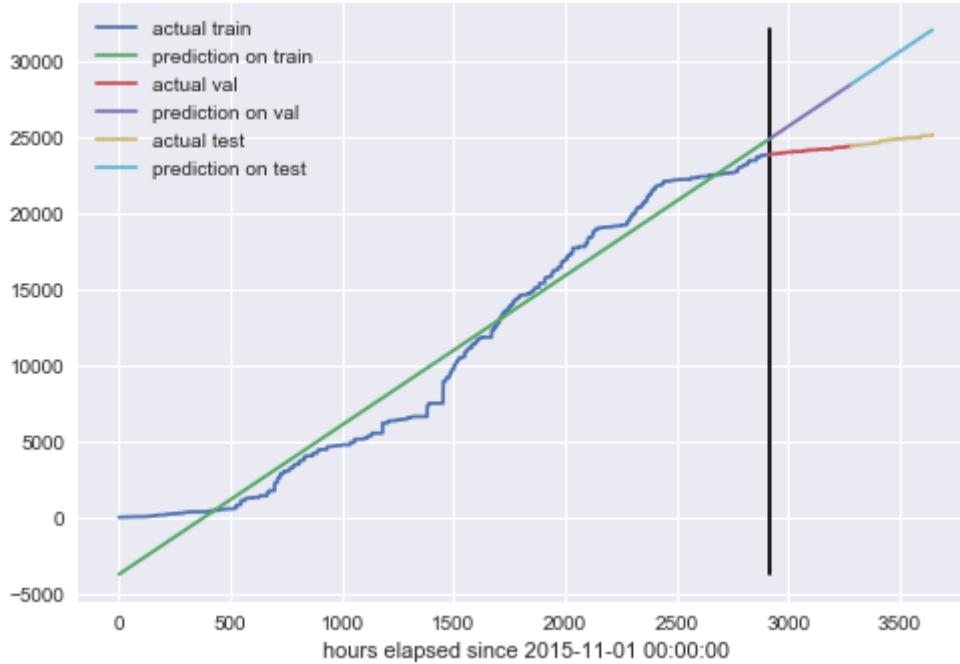
meterid 4029; r2_train: 0.959; r2_val: -45.884; r2_test: -201.732



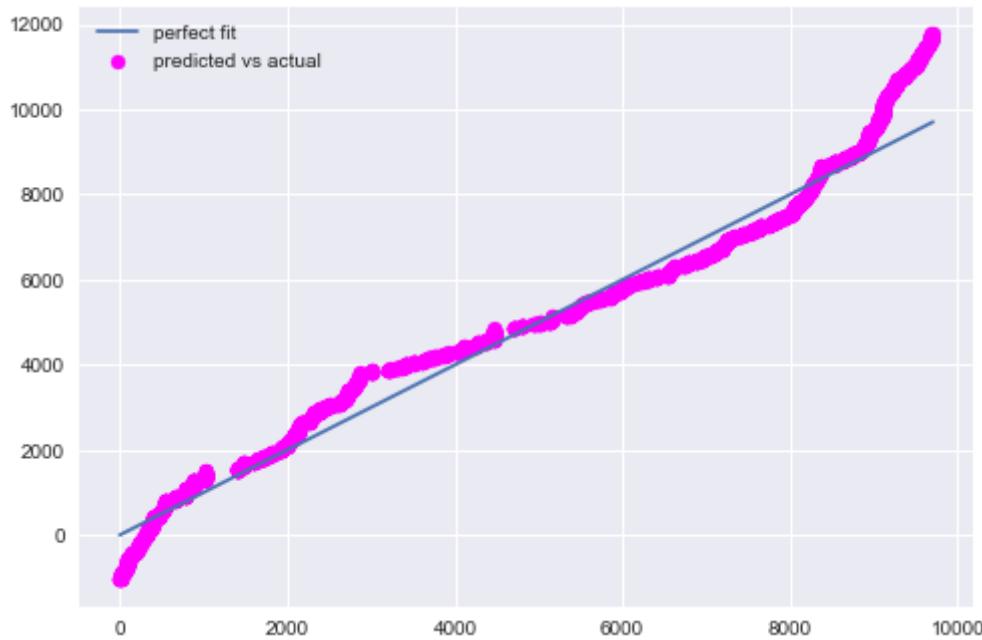
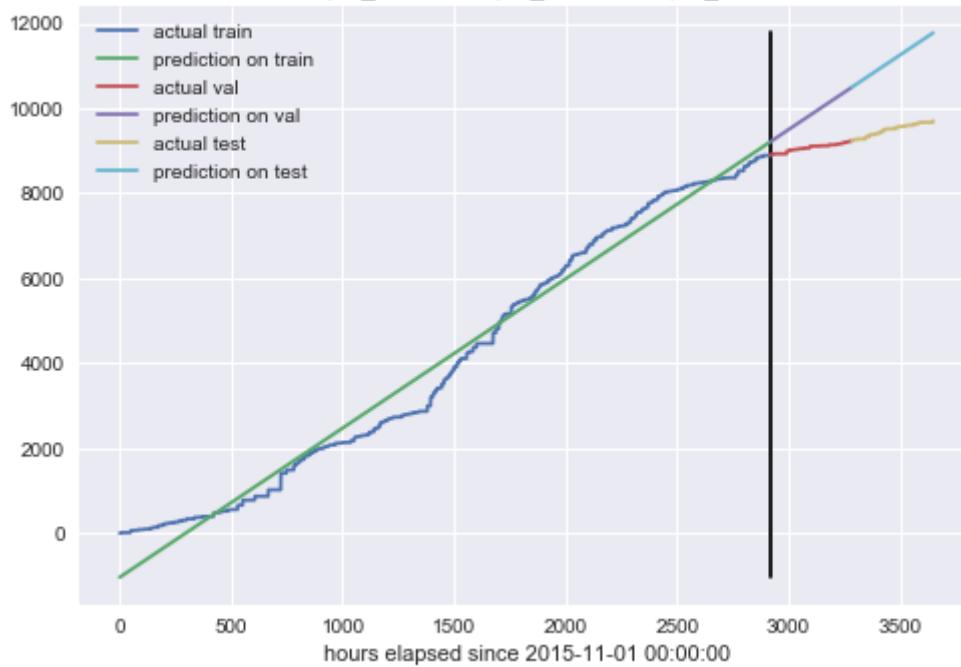
meterid 4031; r2_train: 0.975; r2_val: -137.135; r2_test: -229.650



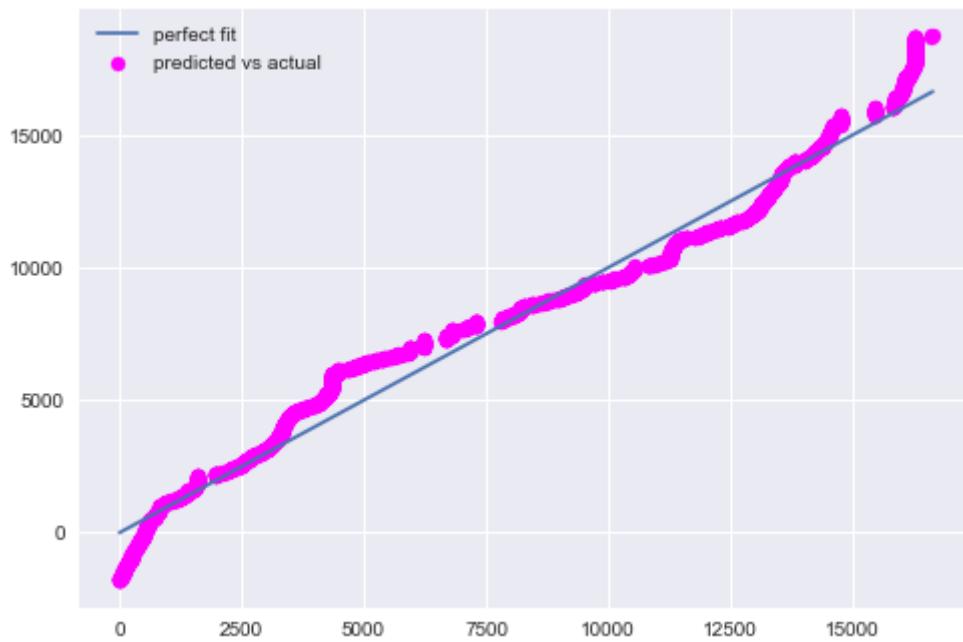
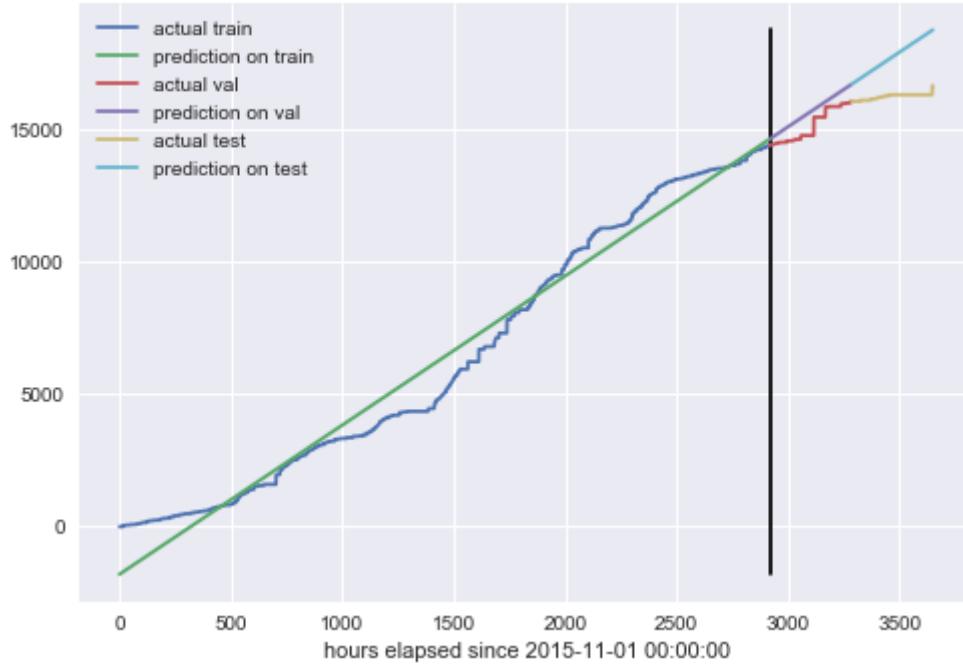
meterid 4193; r2_train: 0.971; r2_val: -324.253; r2_test: -665.626



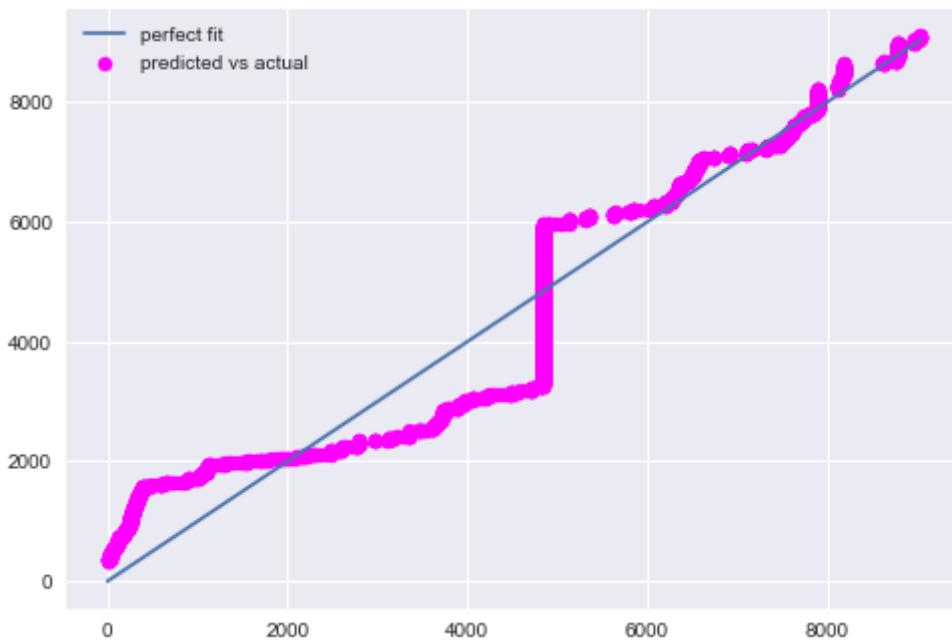
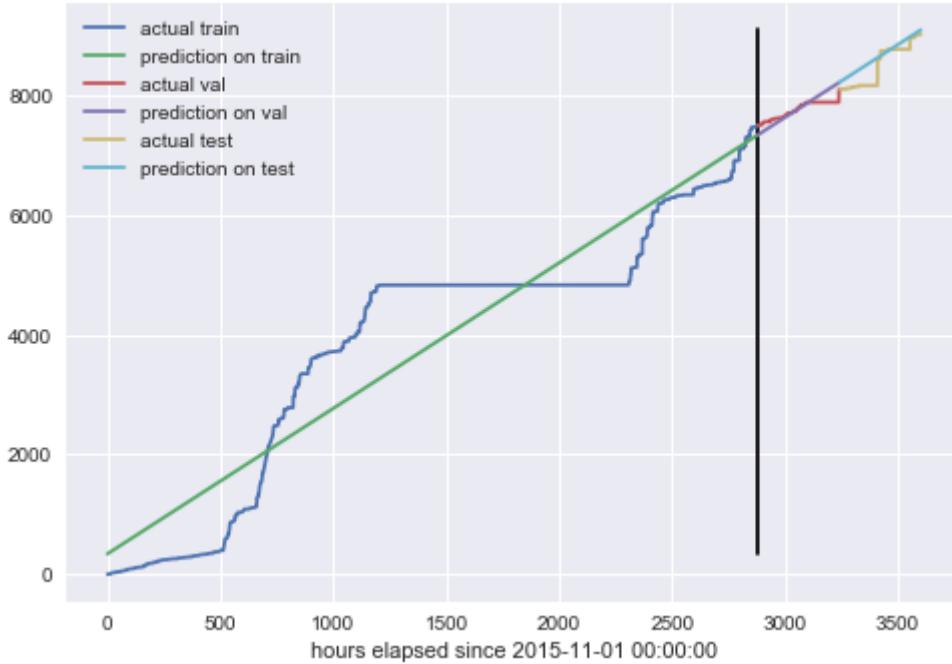
meterid 4228; r2_train: 0.982; r2_val: -81.237; r2_test: -135.866



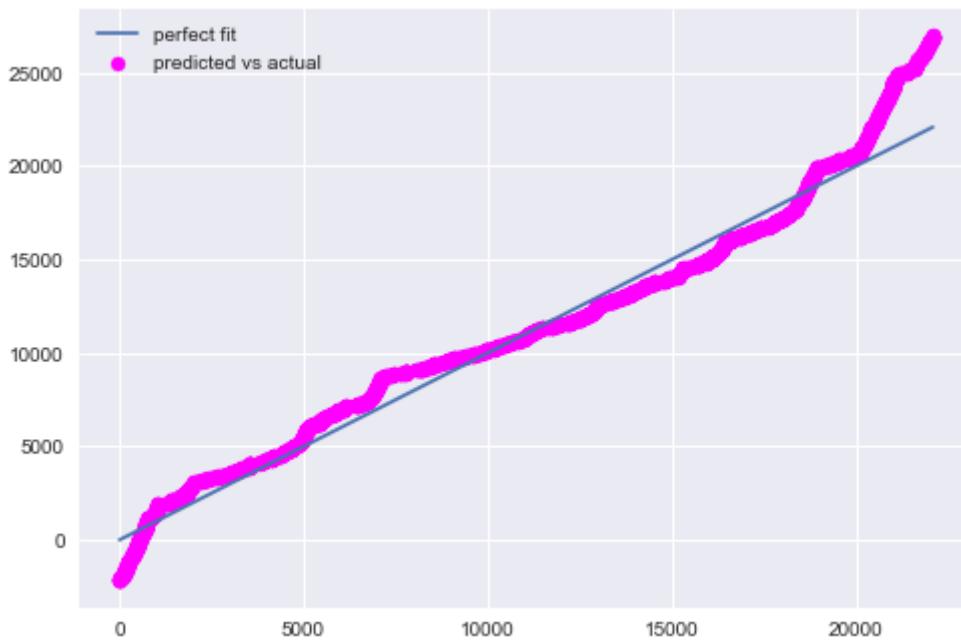
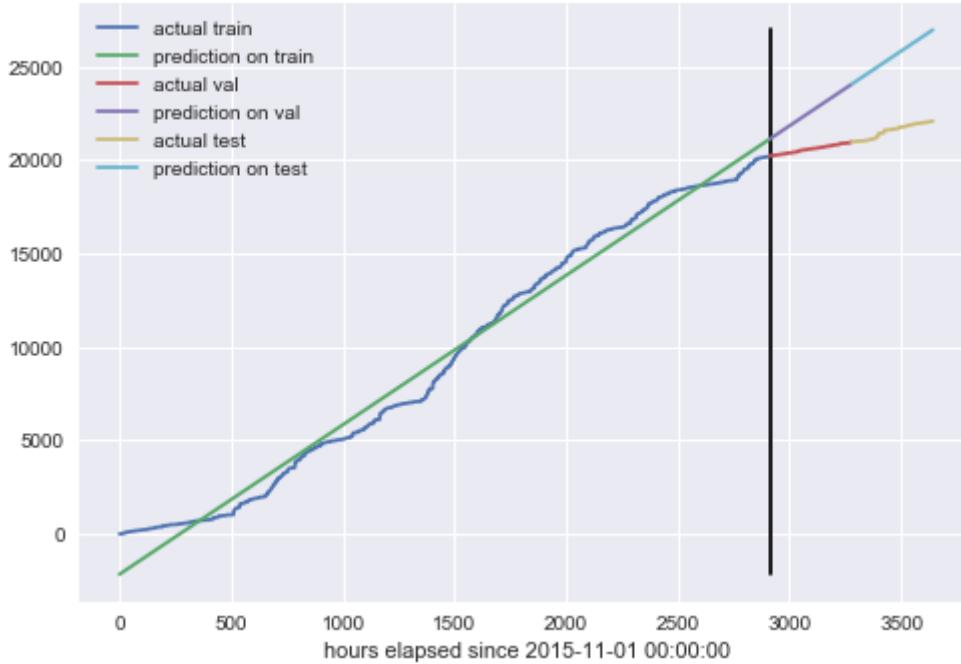
meterid 4296; r2_train: 0.976; r2_val: 0.155; r2_test: -239.766



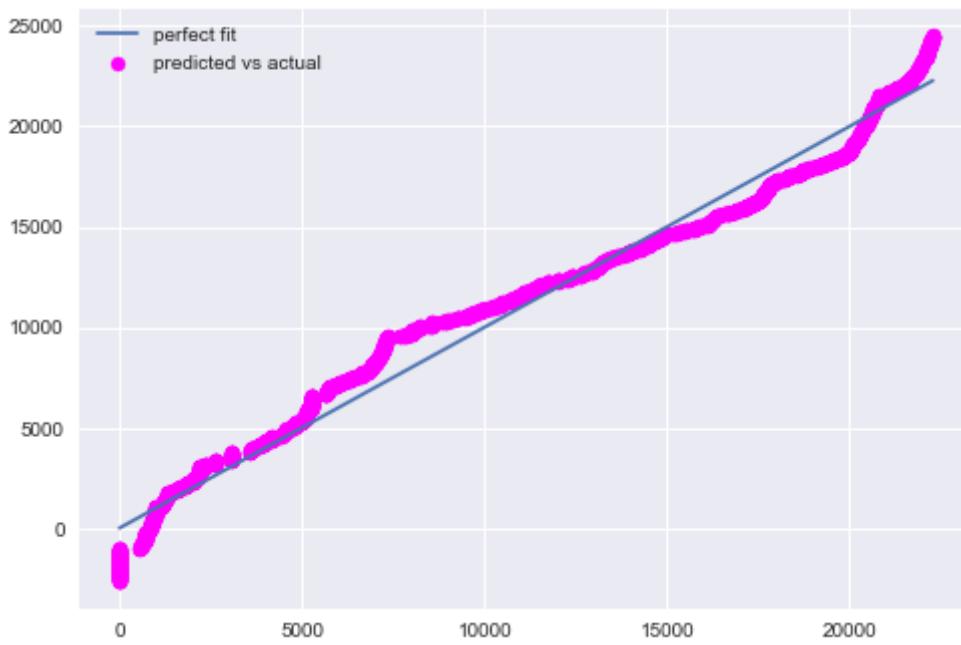
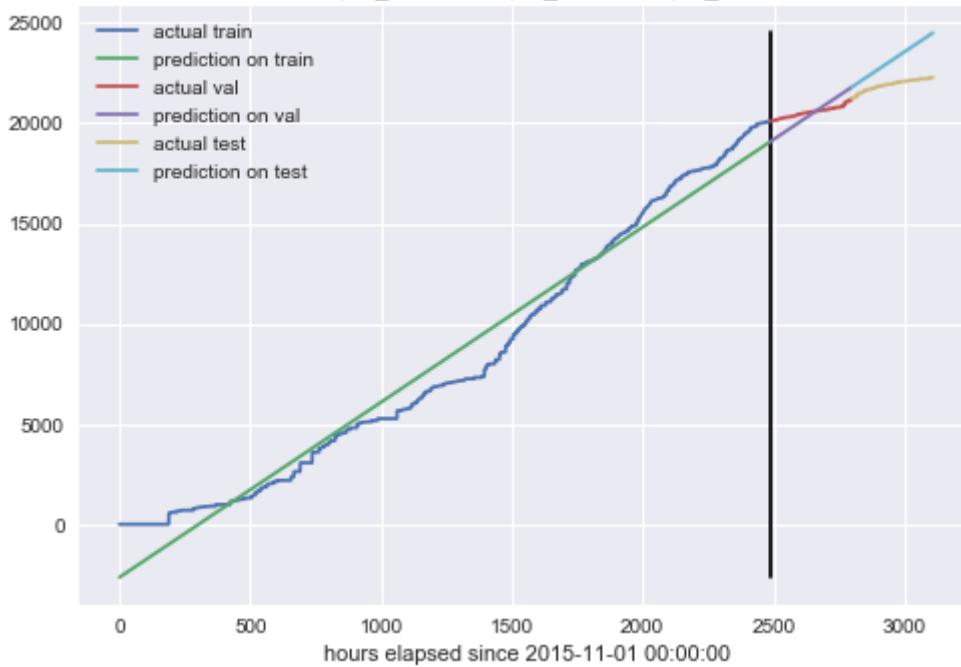
meterid 4352; r2_train: 0.875; r2_val: 0.105; r2_test: 0.617



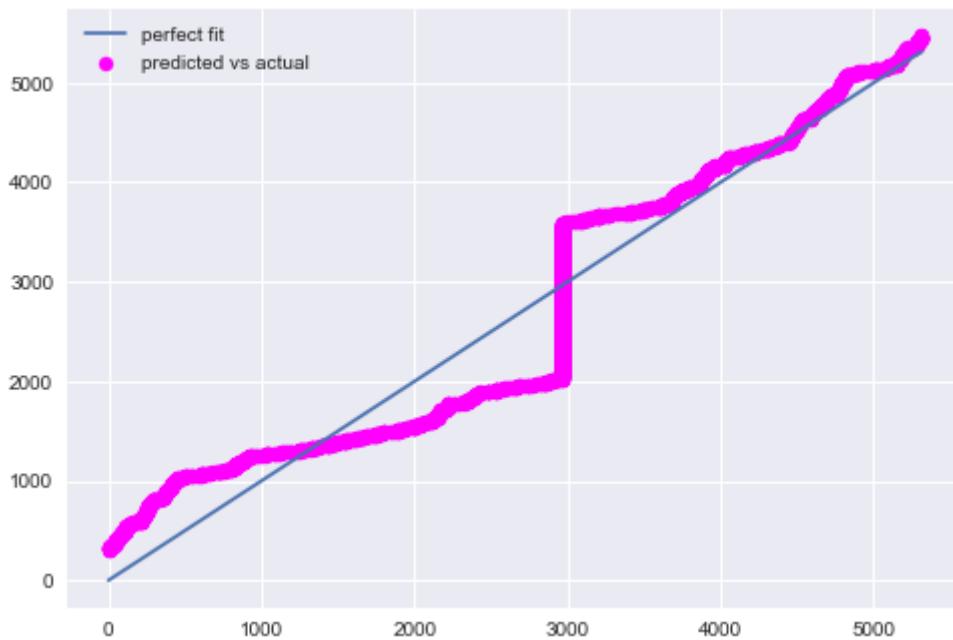
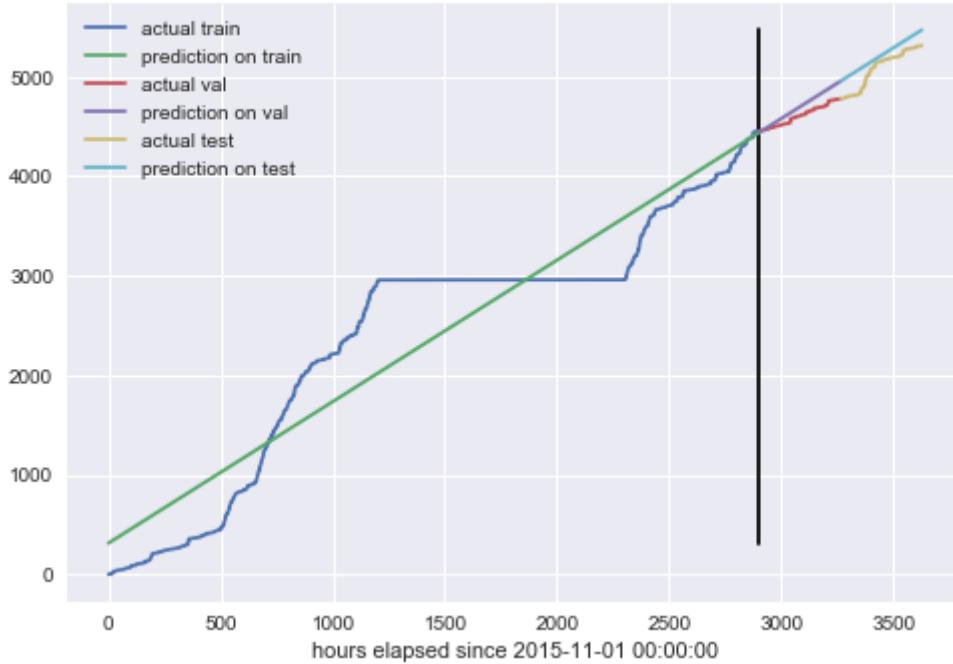
meterid 4356; r2_train: 0.986; r2_val: -92.249; r2_test: -109.050



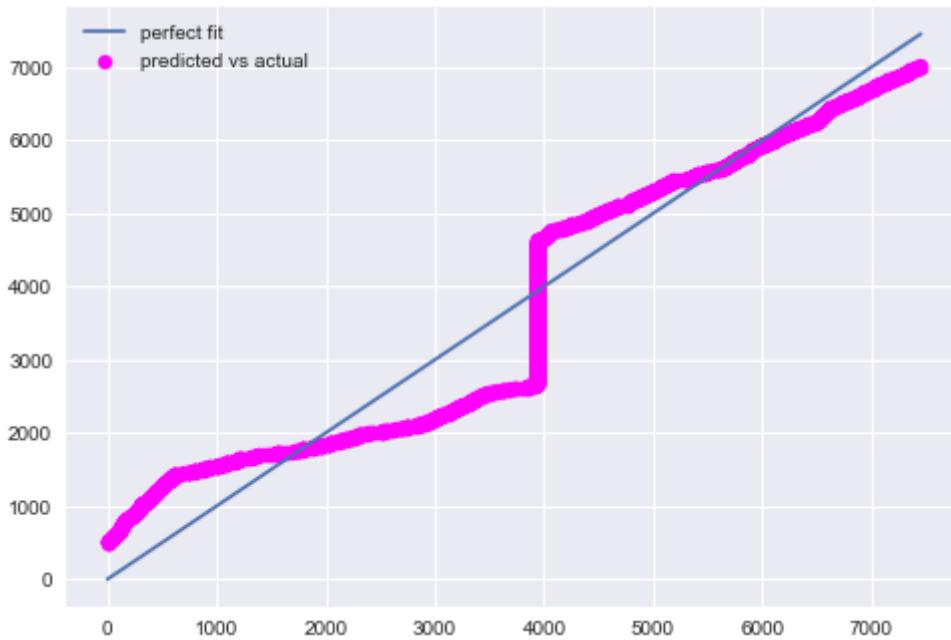
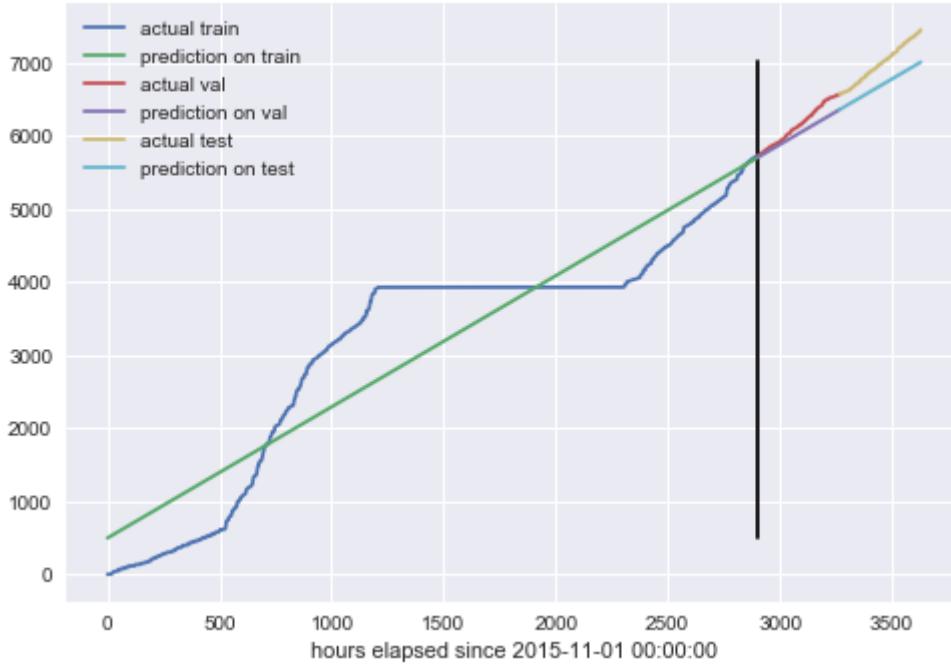
meterid 4373; r2_train: 0.974; r2_val: -3.051; r2_test: -22.153



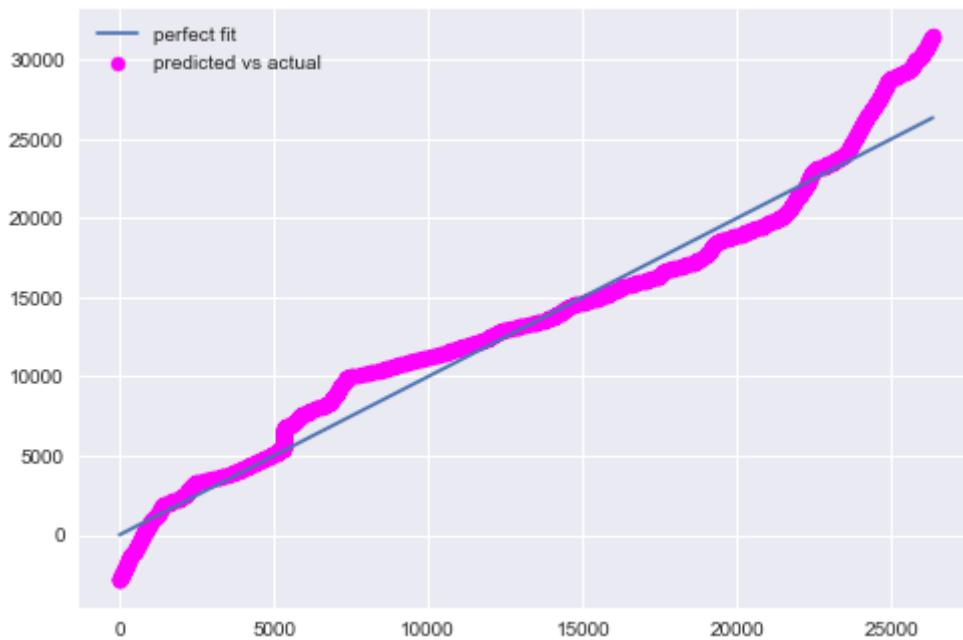
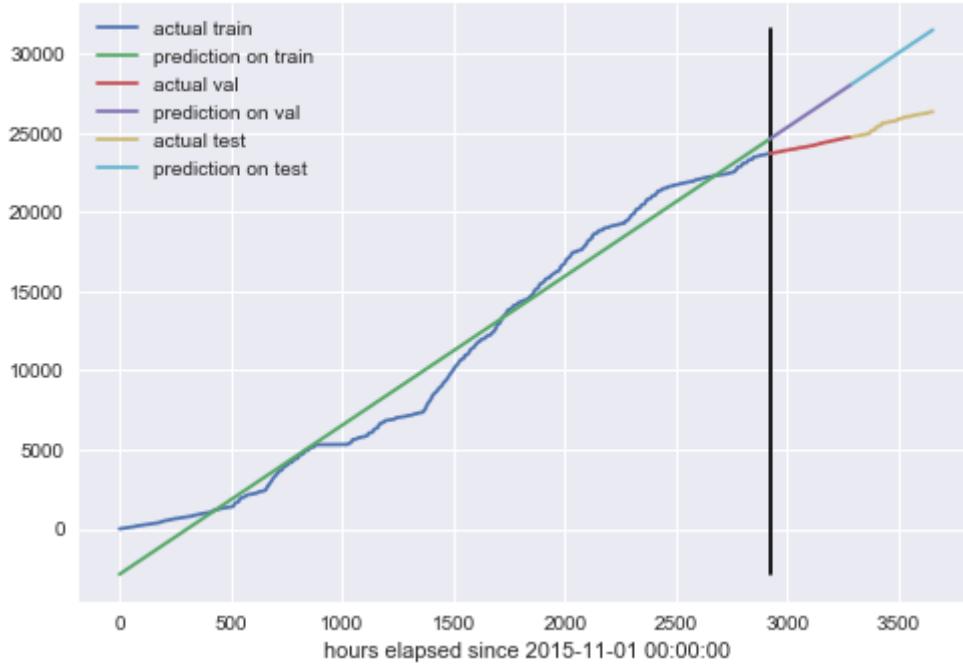
meterid 4421; r2_train: 0.886; r2_val: 0.047; r2_test: 0.398



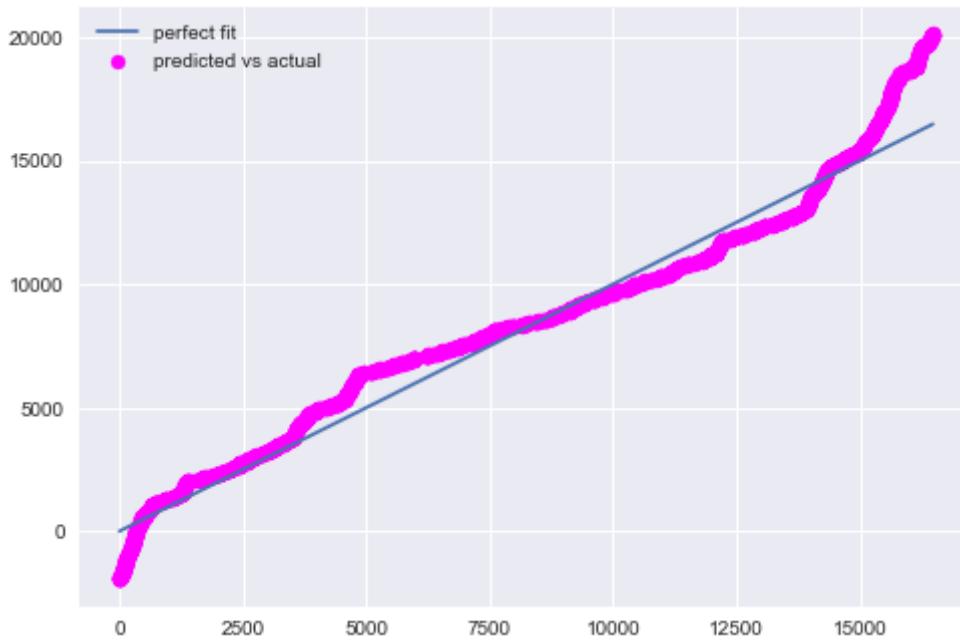
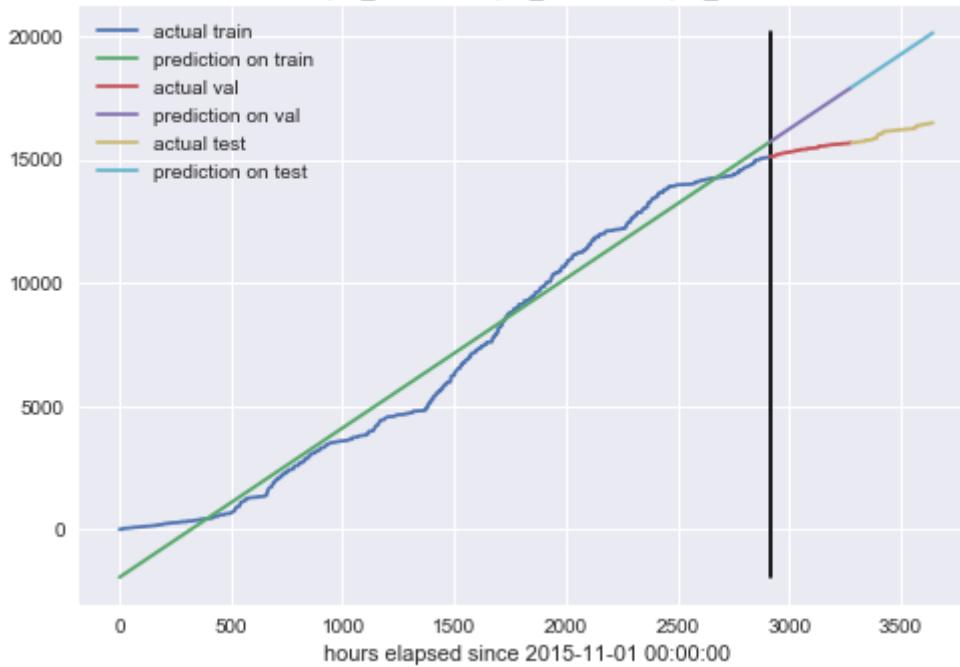
meterid 4447; r2_train: 0.854; r2_val: 0.703; r2_test: -0.383



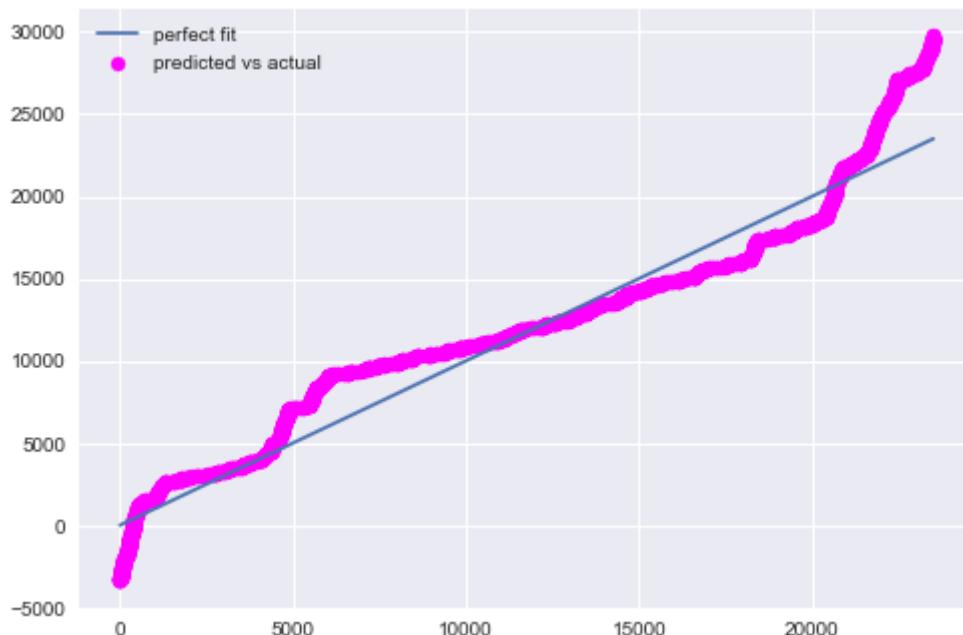
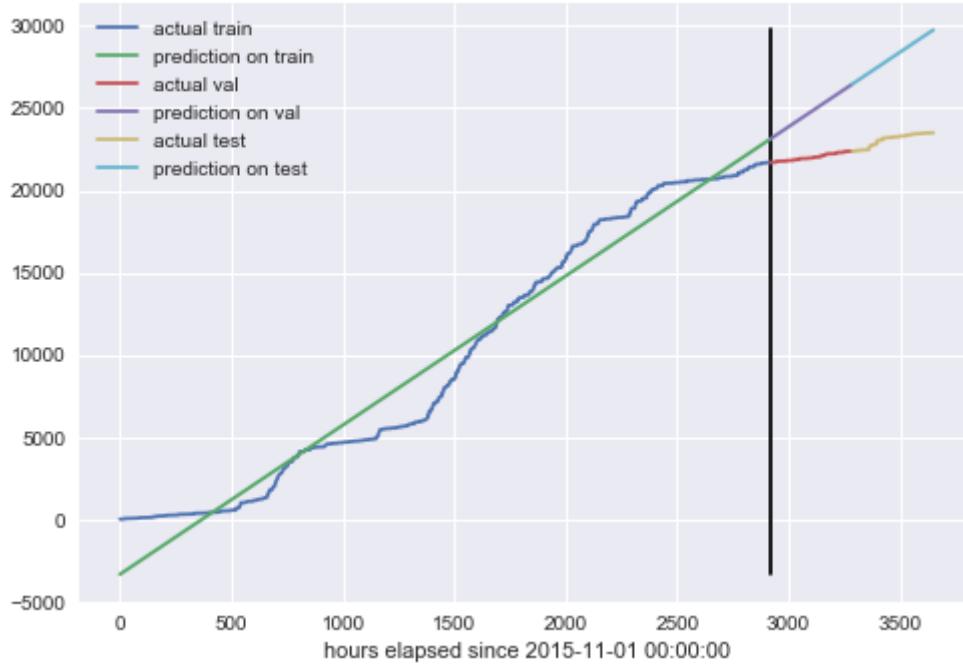
meterid 4514; r2_train: 0.979; r2_val: -51.598; r2_test: -64.250



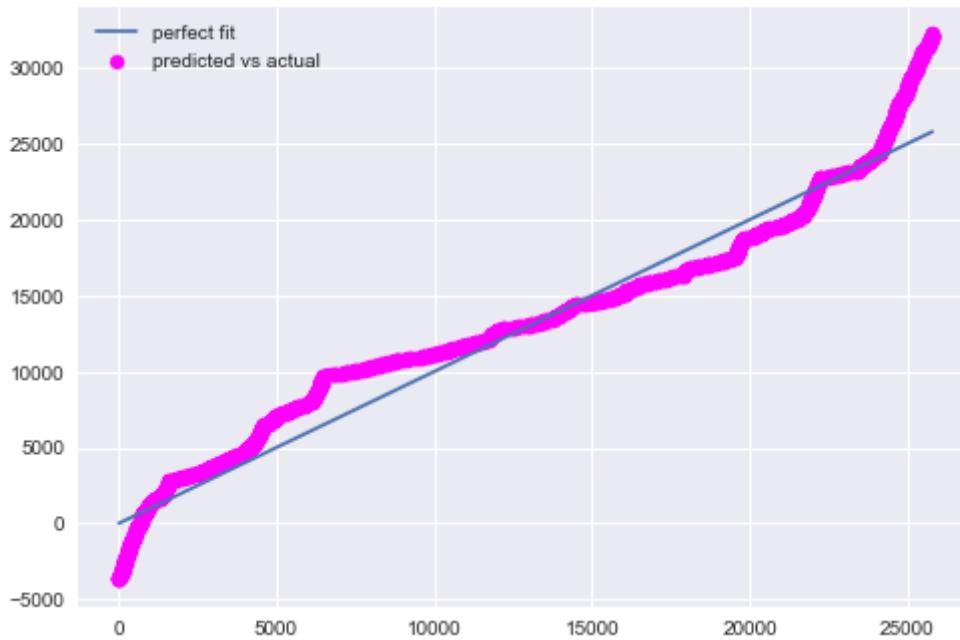
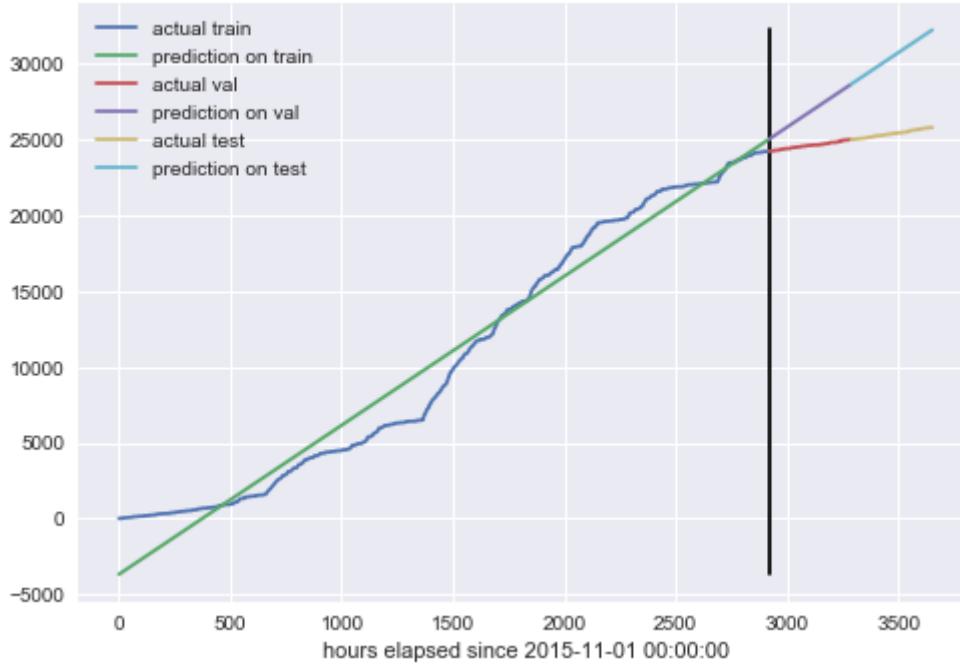
meterid 4732; r2_train: 0.980; r2_val: -79.268; r2_test: -134.735



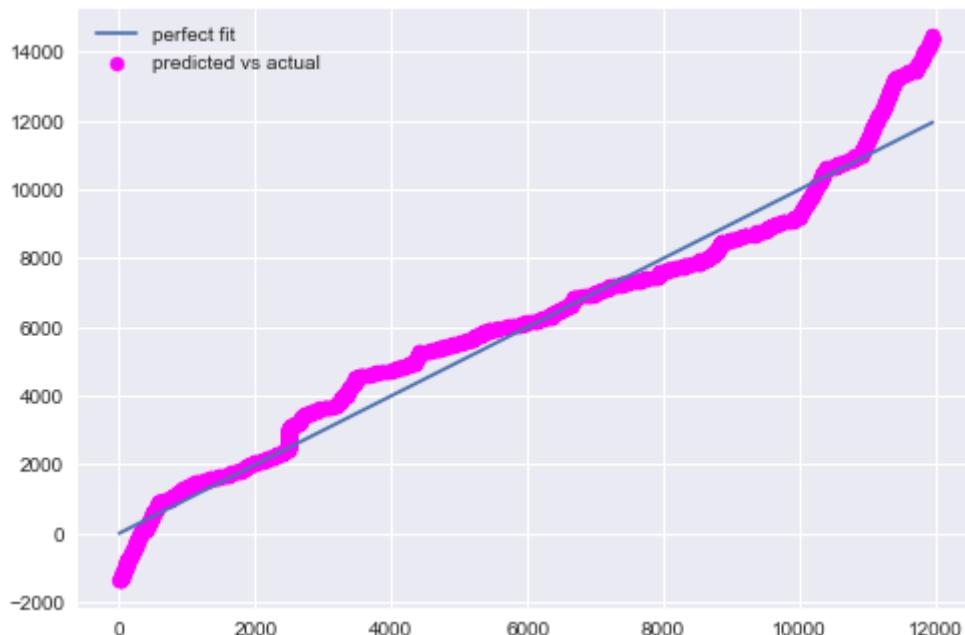
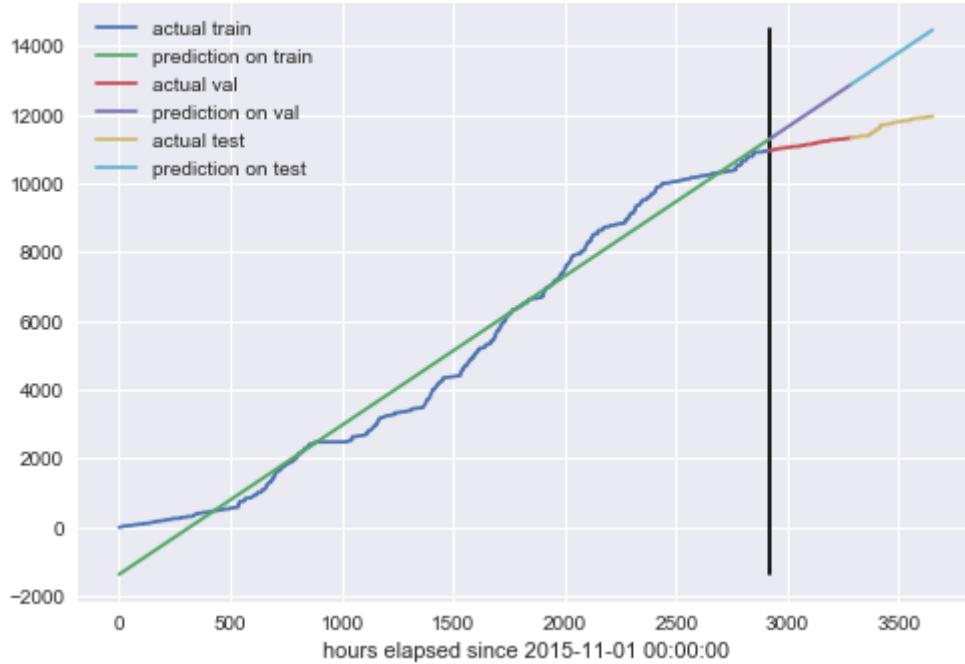
meterid 4767; r2_train: 0.967; r2_val: -177.077; r2_test: -167.353



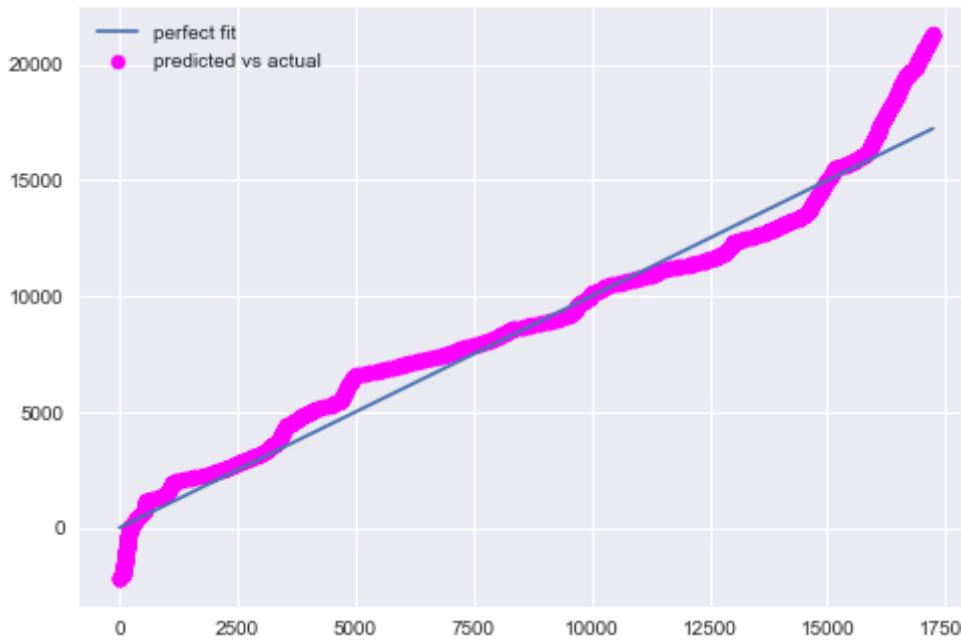
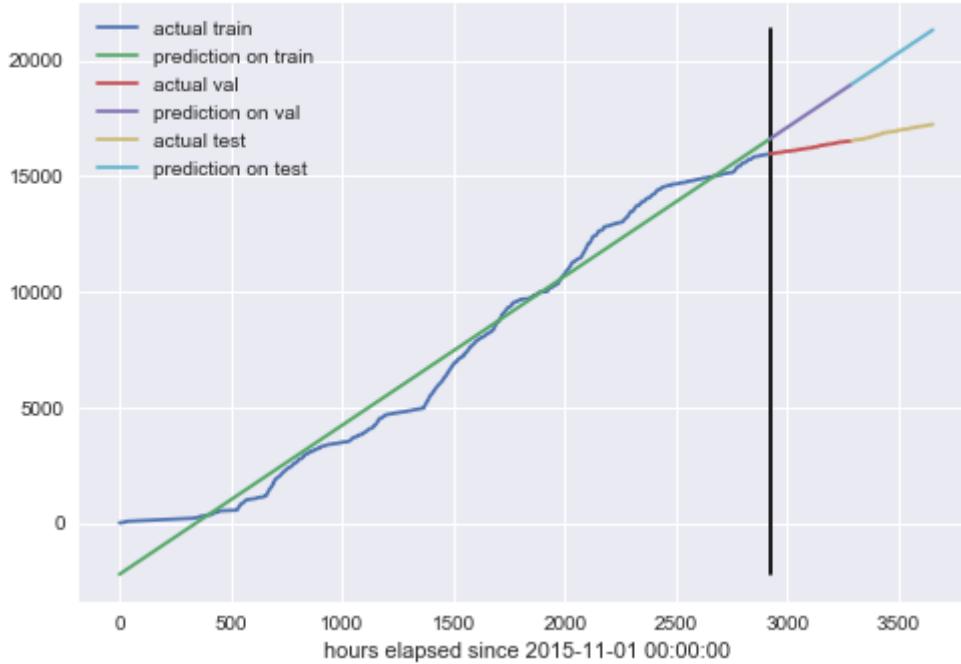
meterid 4998; r2_train: 0.970; r2_val: -119.063; r2_test: -443.851



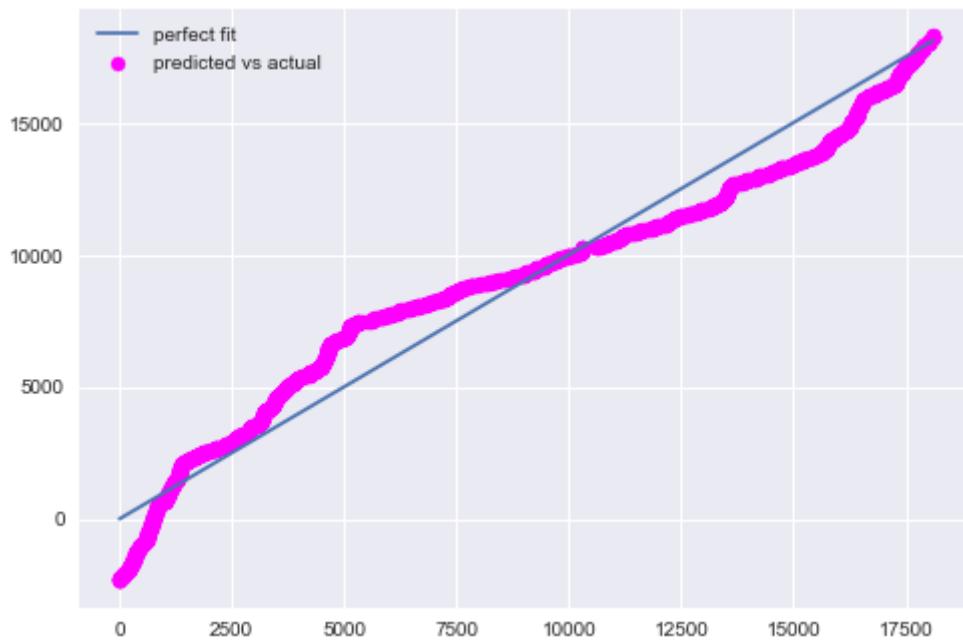
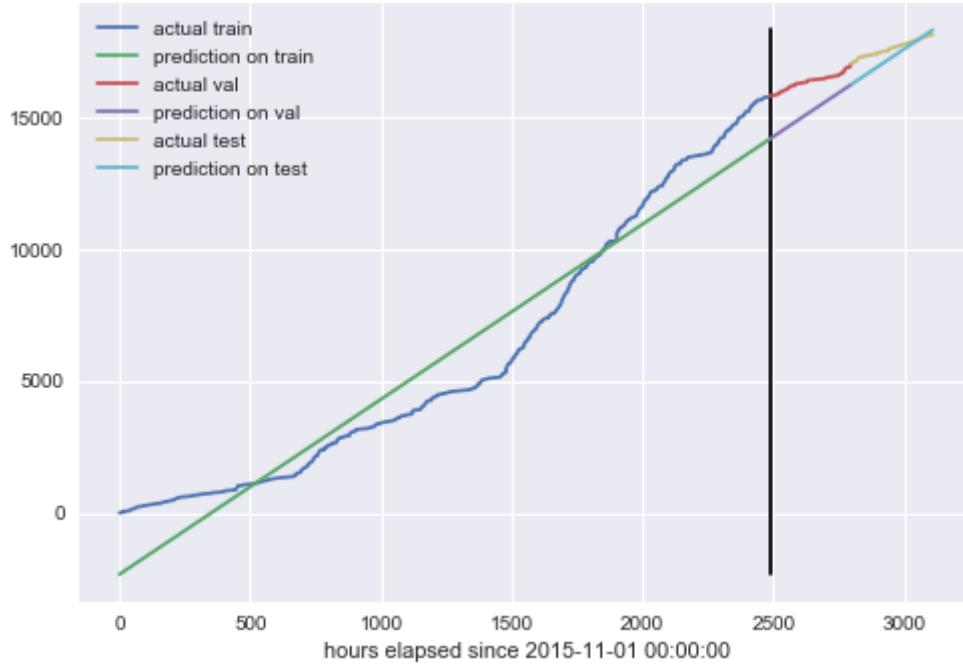
meterid 5129; r2_train: 0.979; r2_val: -84.391; r2_test: -90.079



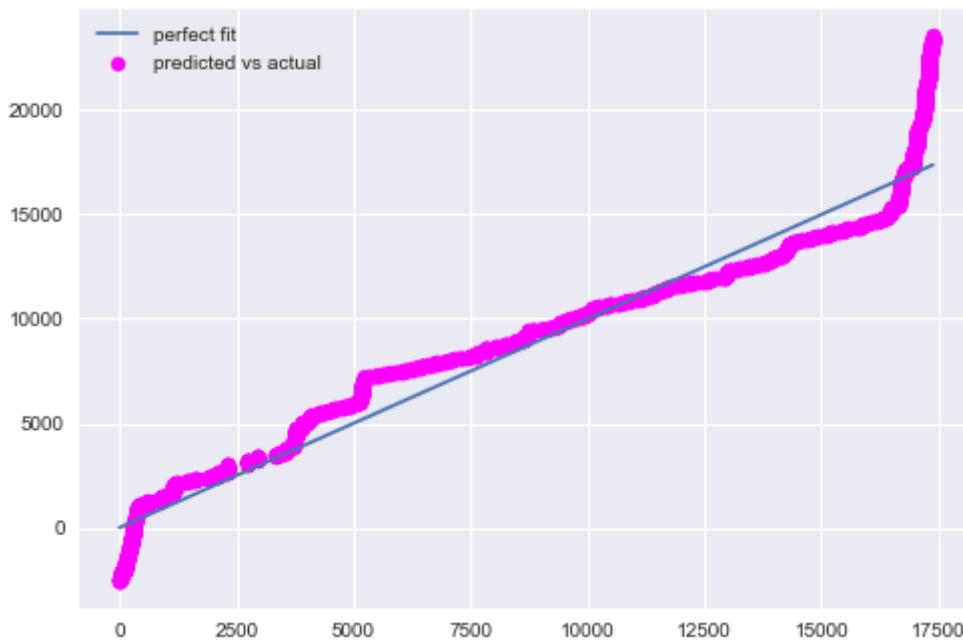
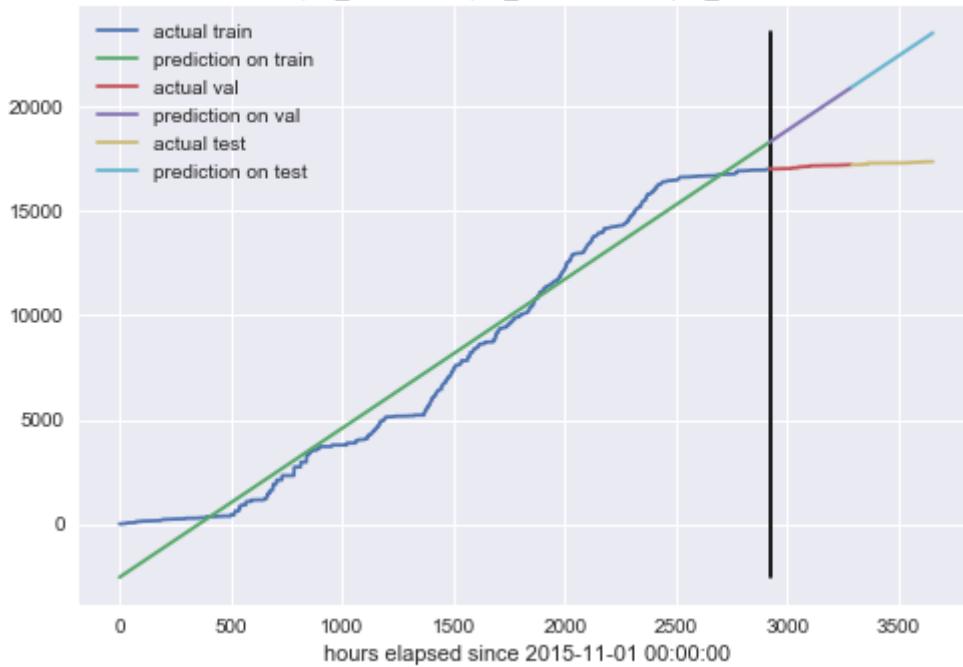
meterid 5131; r2_train: 0.980; r2_val: -90.658; r2_test: -233.410



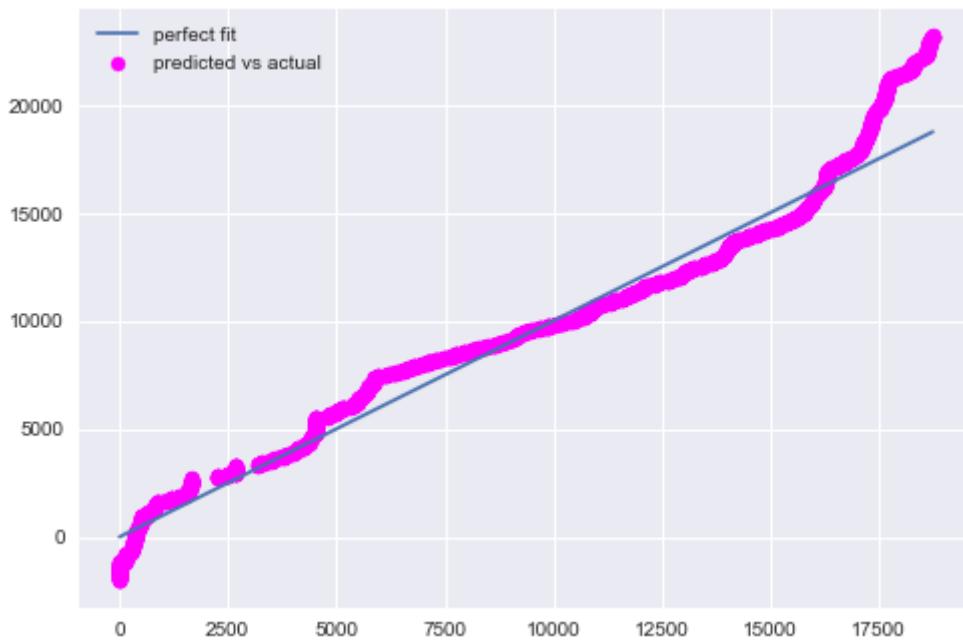
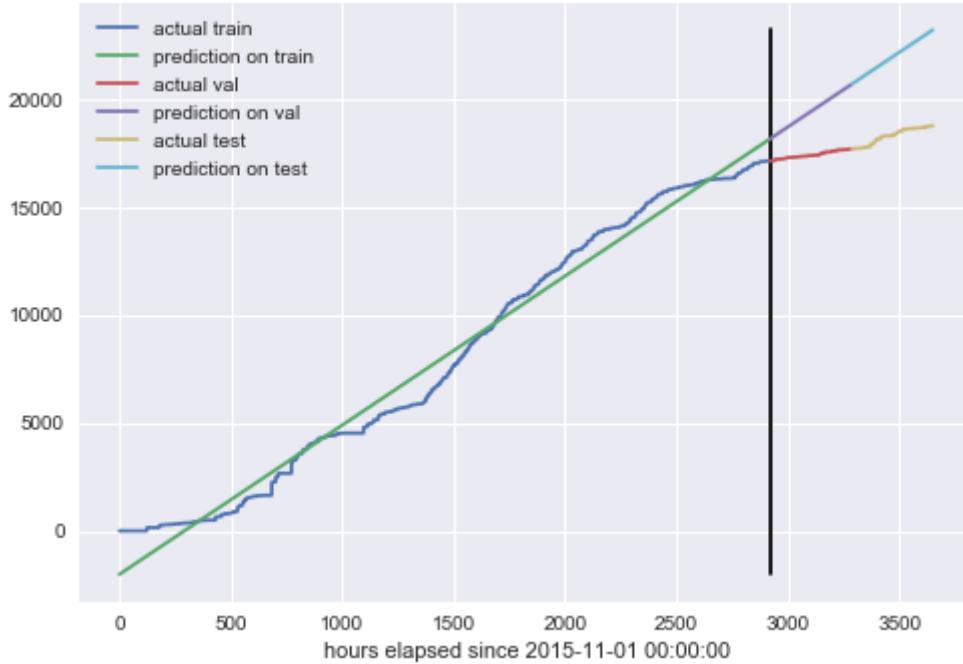
meterid 5193; r2_train: 0.939; r2_val: -15.311; r2_test: -1.332



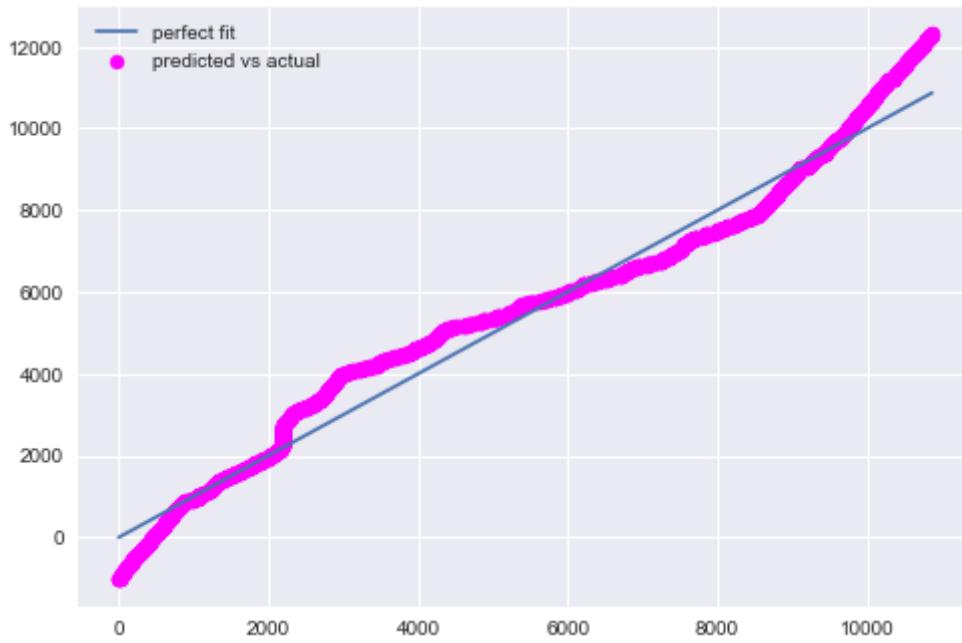
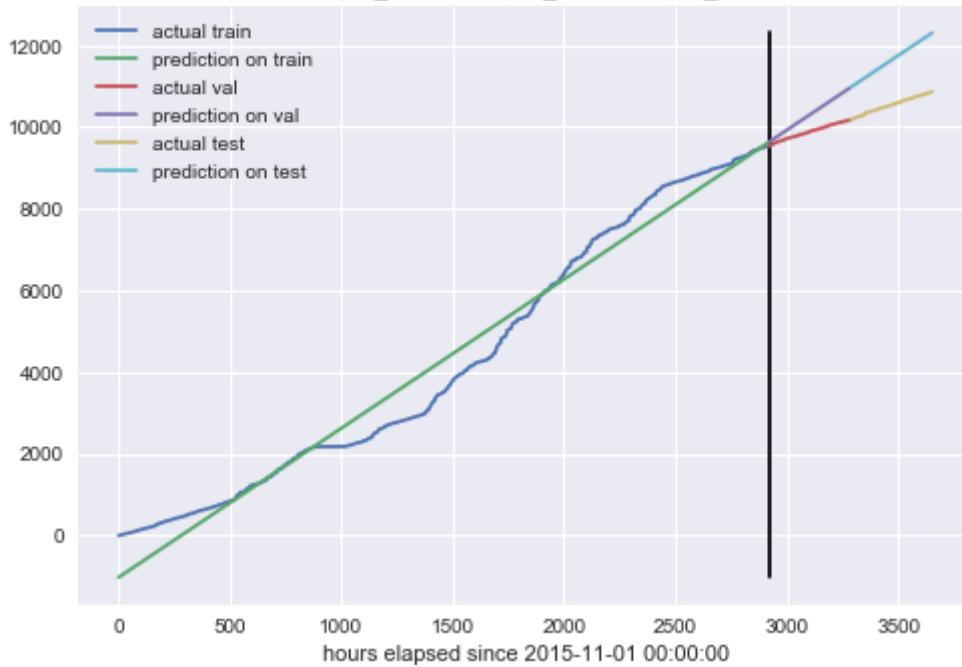
meterid 5275; r2_train: 0.974; r2_val: -1350.418; r2_test: -13220.266



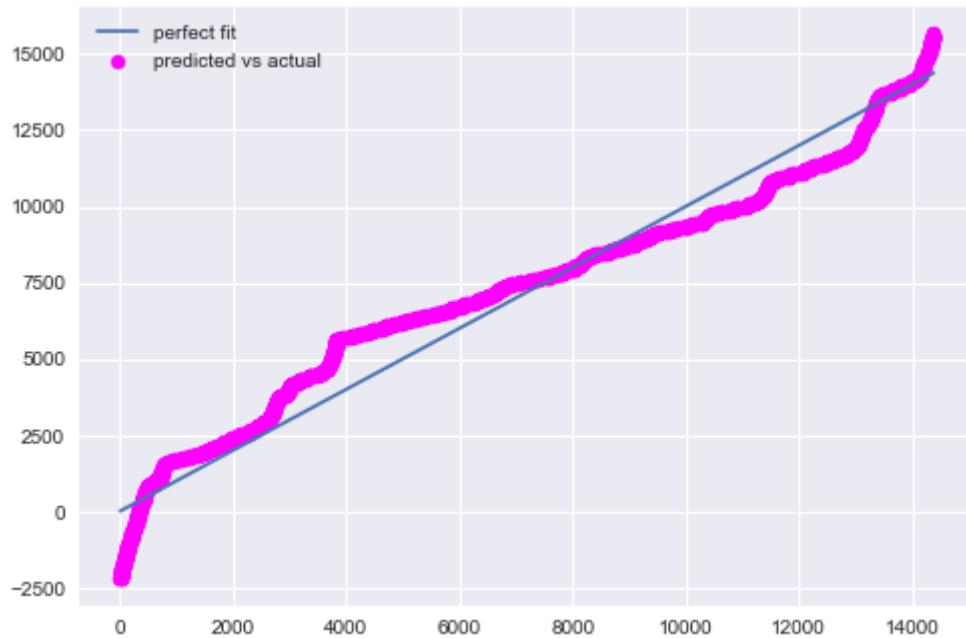
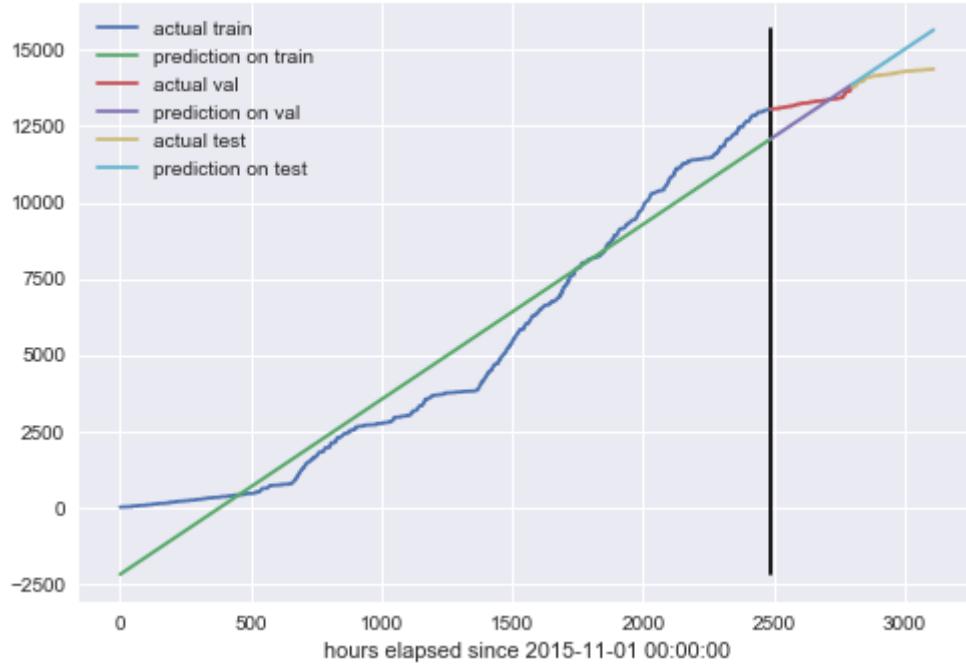
meterid 5395; r2_train: 0.984; r2_val: -160.669; r2_test: -96.470



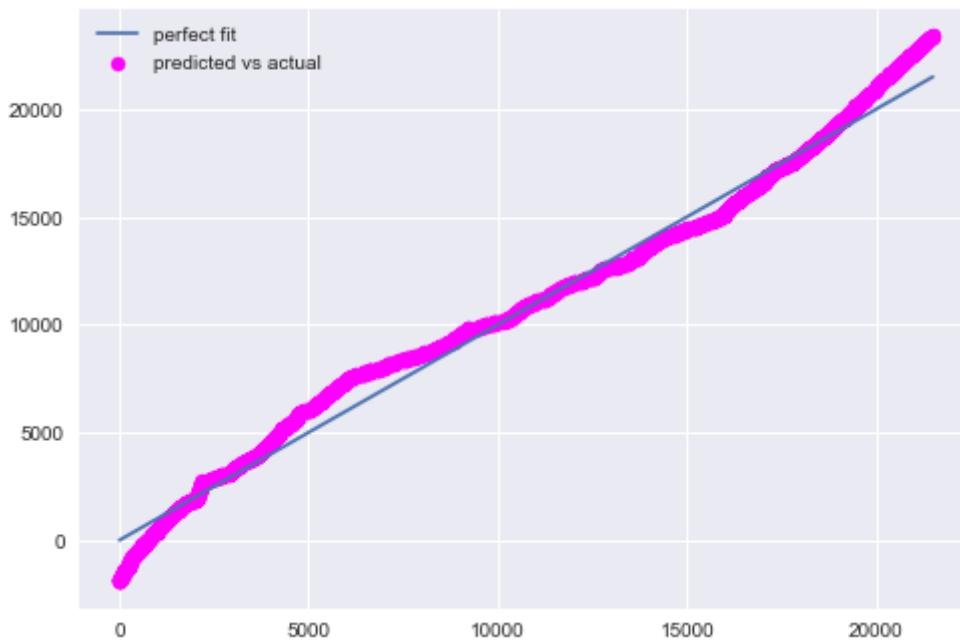
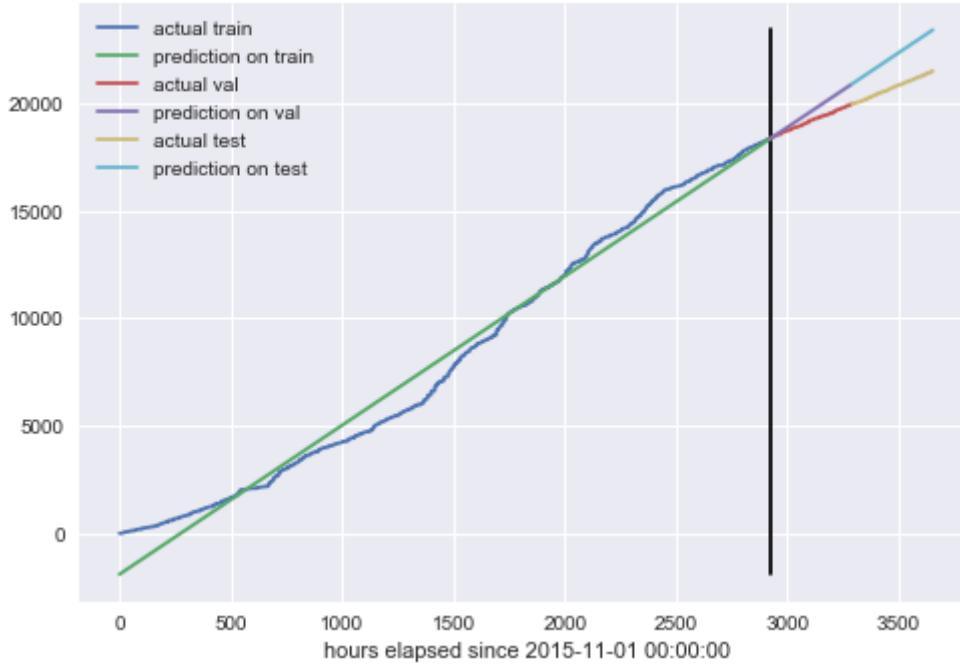
meterid 5403; r2_train: 0.975; r2_val: -5.710; r2_test: -31.677

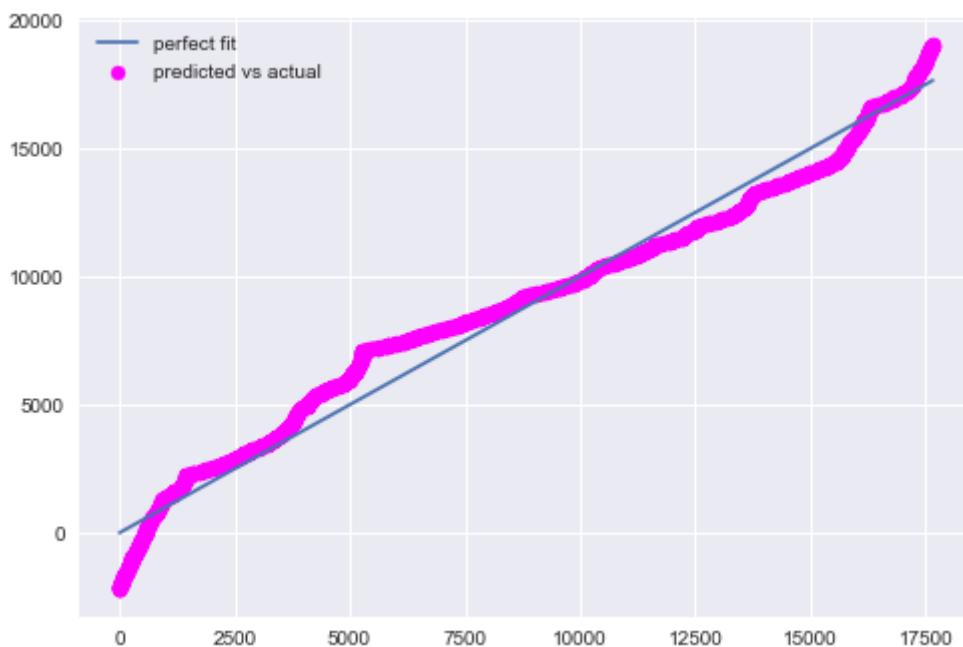
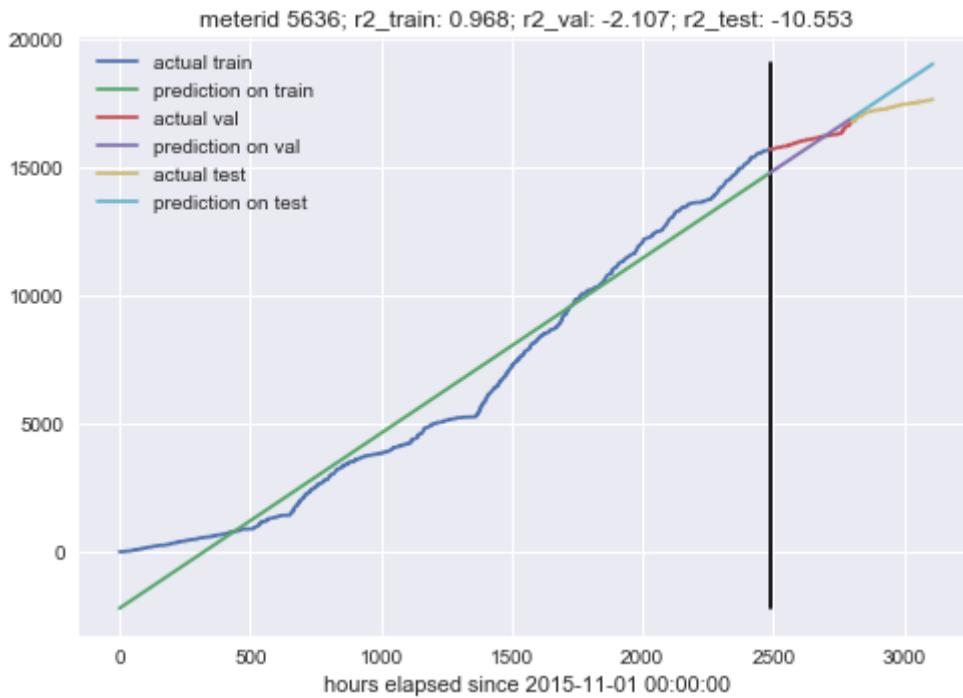


meterid 5439; r2_train: 0.951; r2_val: -7.777; r2_test: -19.416

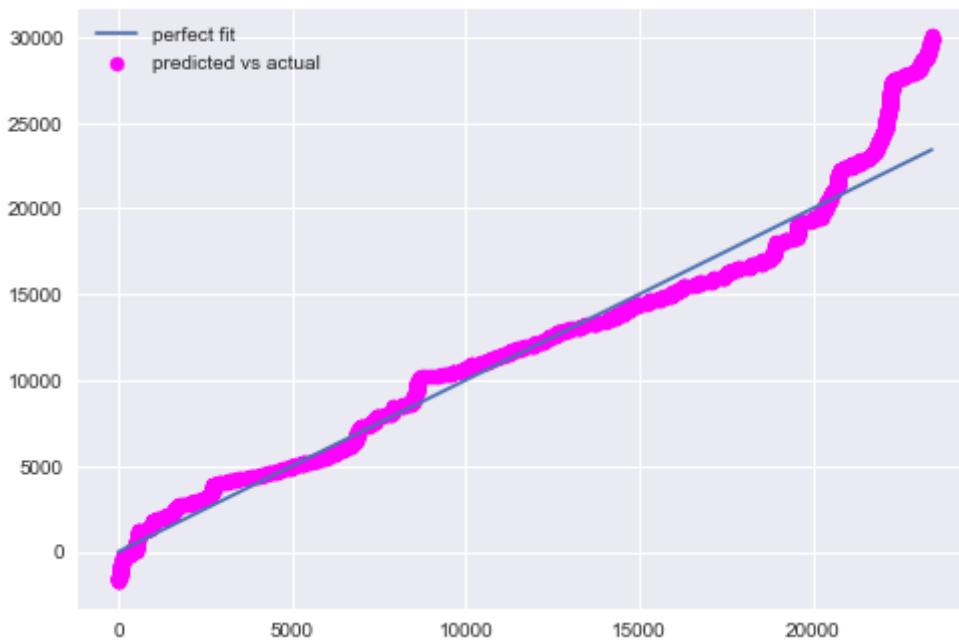
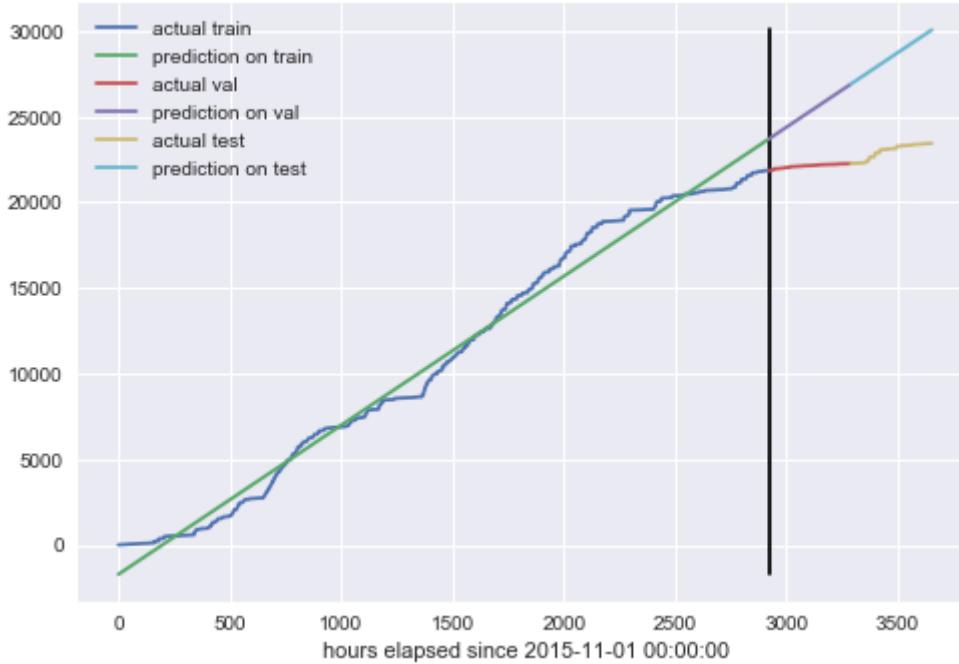


meterid 5484; r2_train: 0.985; r2_val: -0.337; r2_test: -9.481

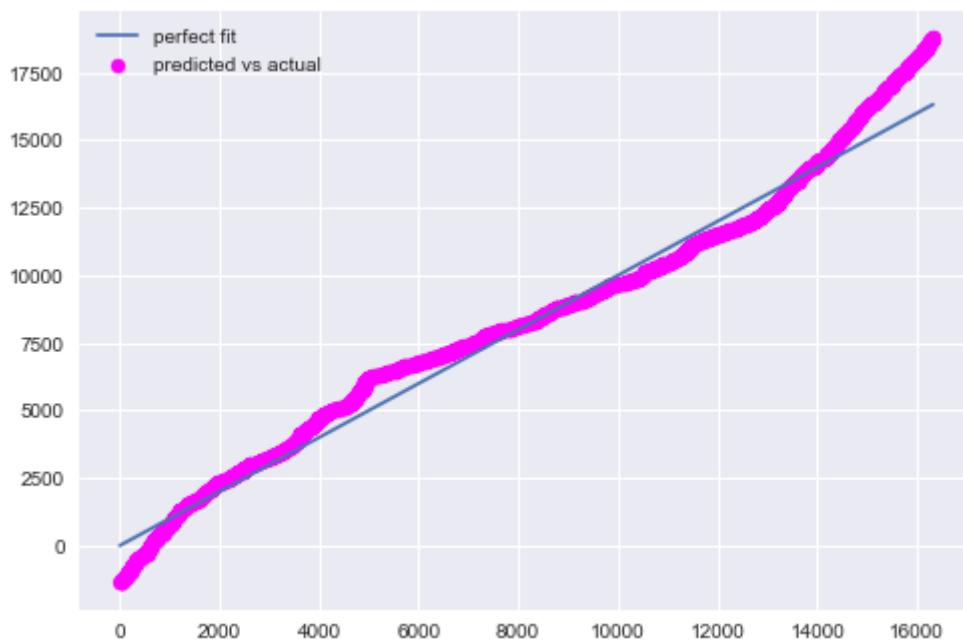
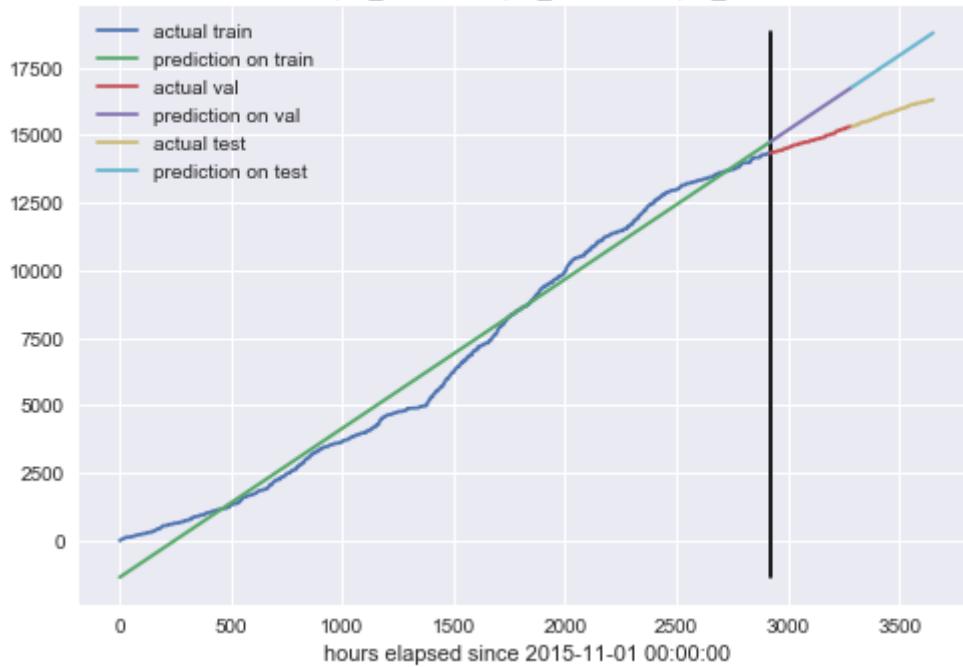




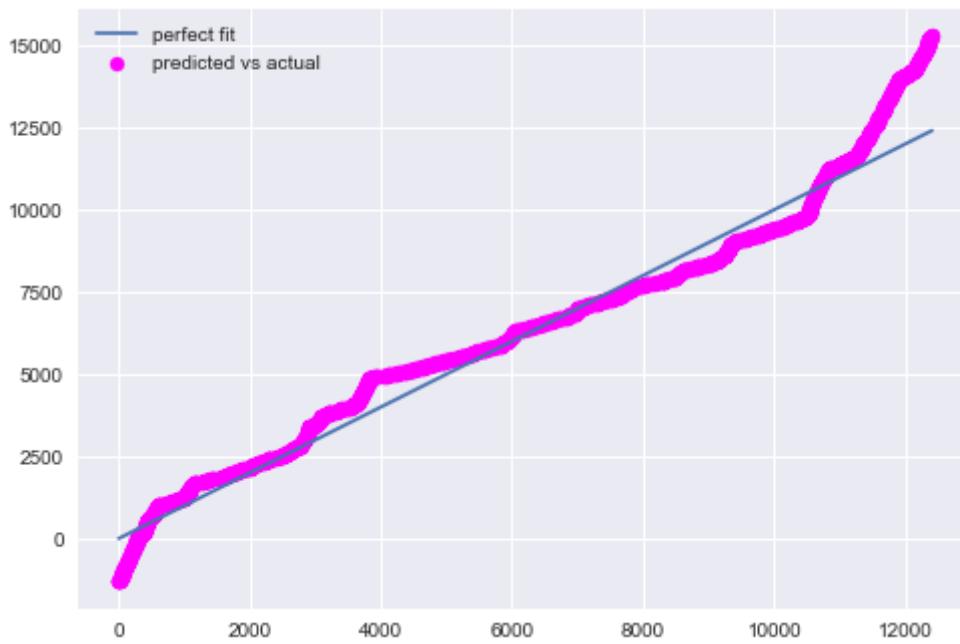
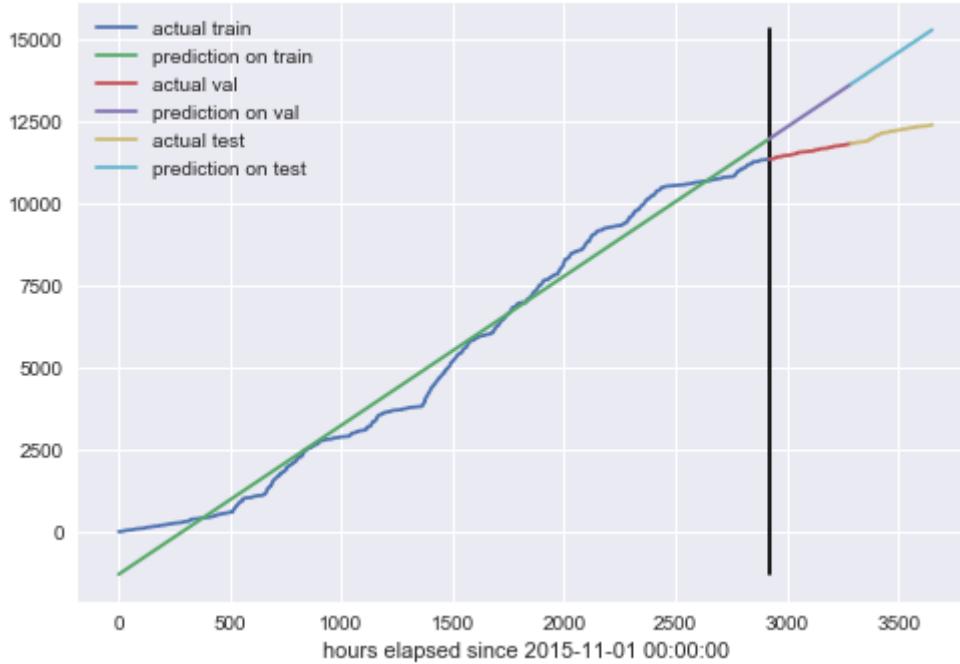
meterid 5785; r2_train: 0.987; r2_val: -941.191; r2_test: -169.561



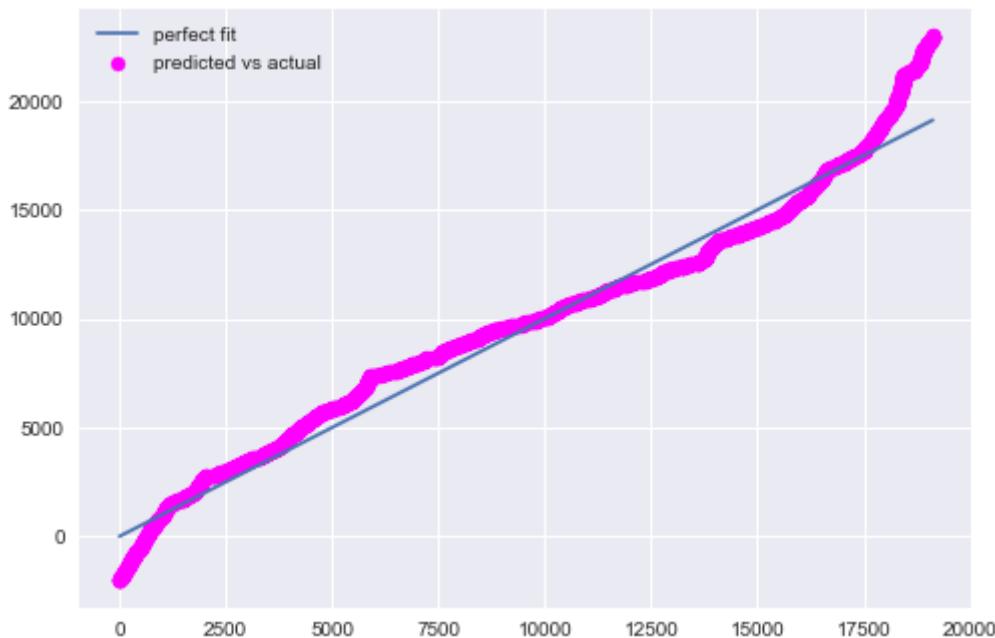
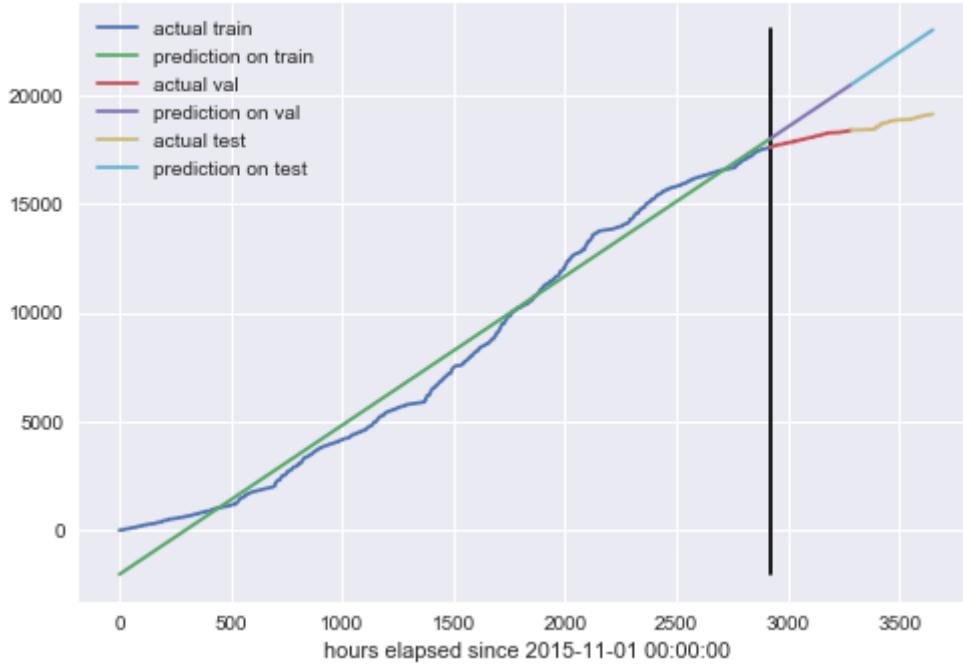
meterid 5810; r2_train: 0.986; r2_val: -11.135; r2_test: -41.971



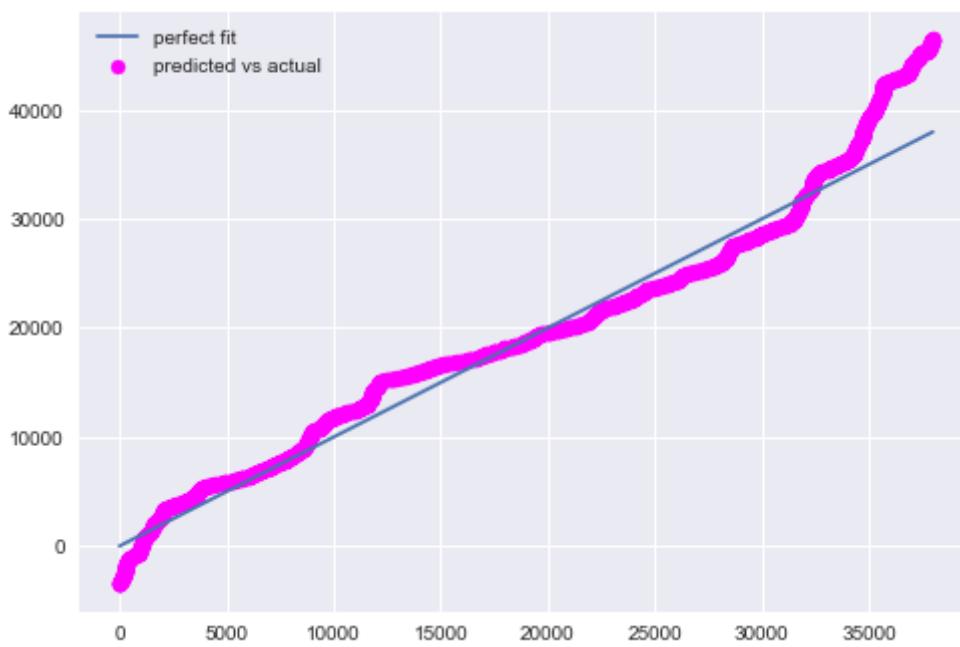
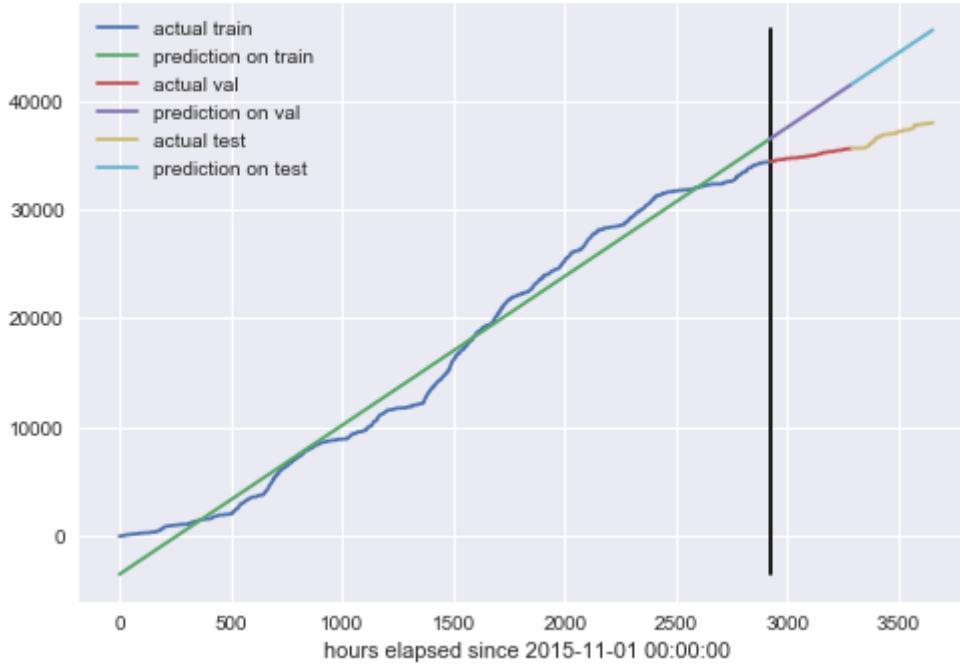
meterid 5814; r2_train: 0.984; r2_val: -86.514; r2_test: -160.723



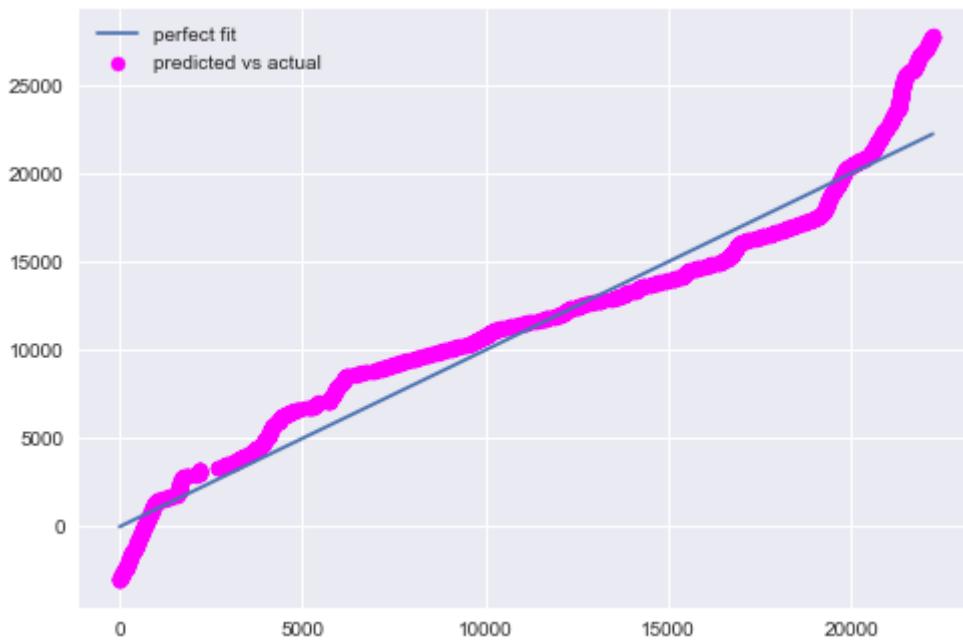
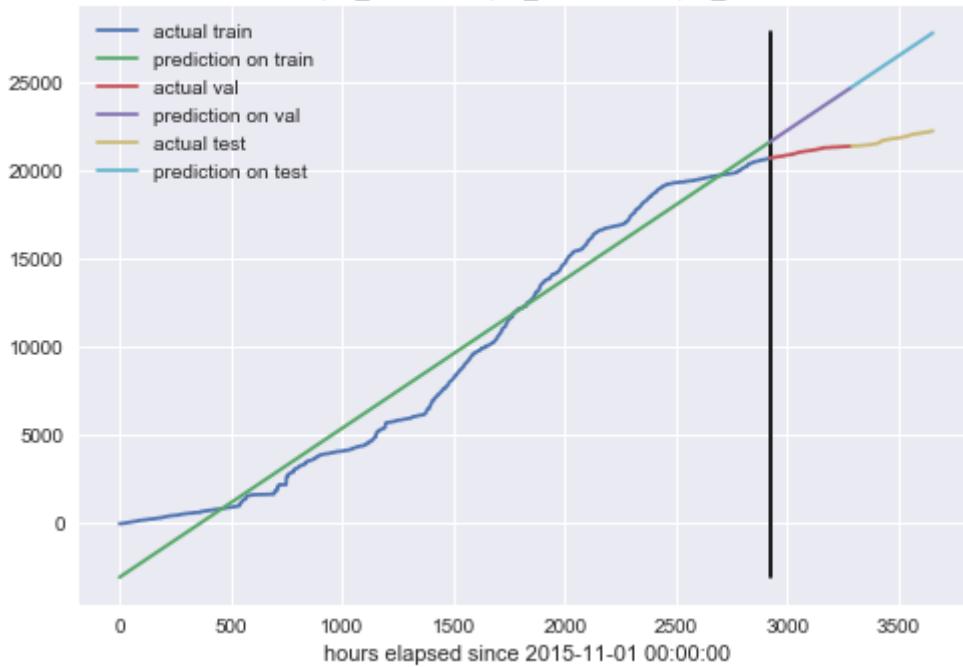
meterid 5892; r2_train: 0.983; r2_val: -31.125; r2_test: -150.717



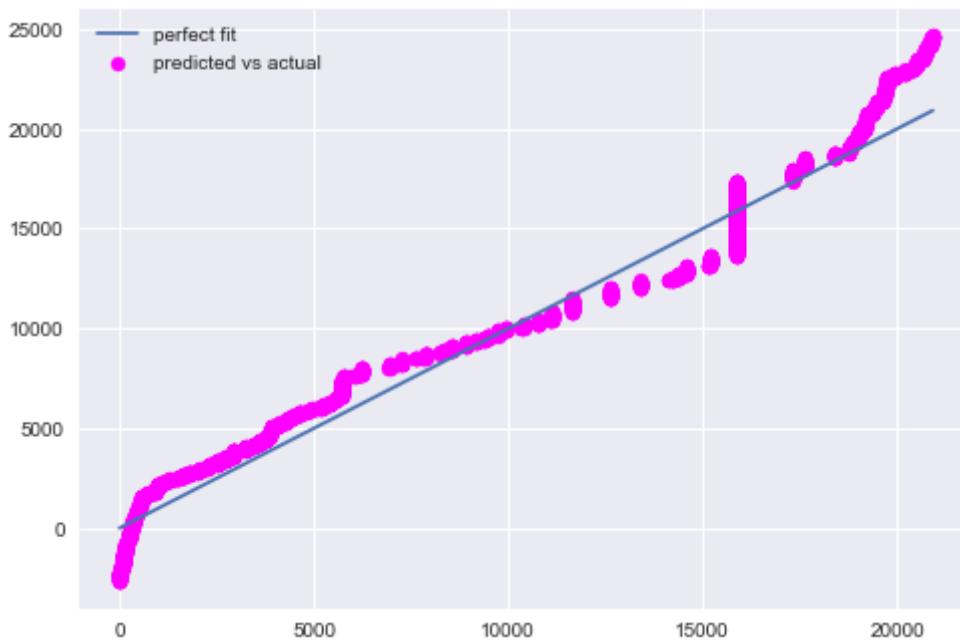
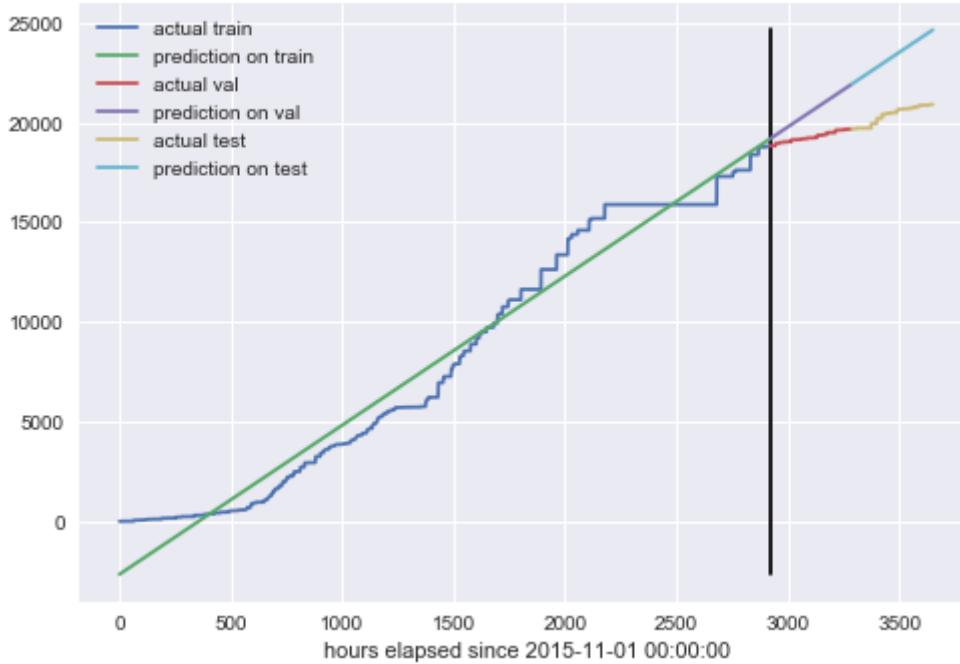
meterid 5972; r2_train: 0.985; r2_val: -135.841; r2_test: -81.013



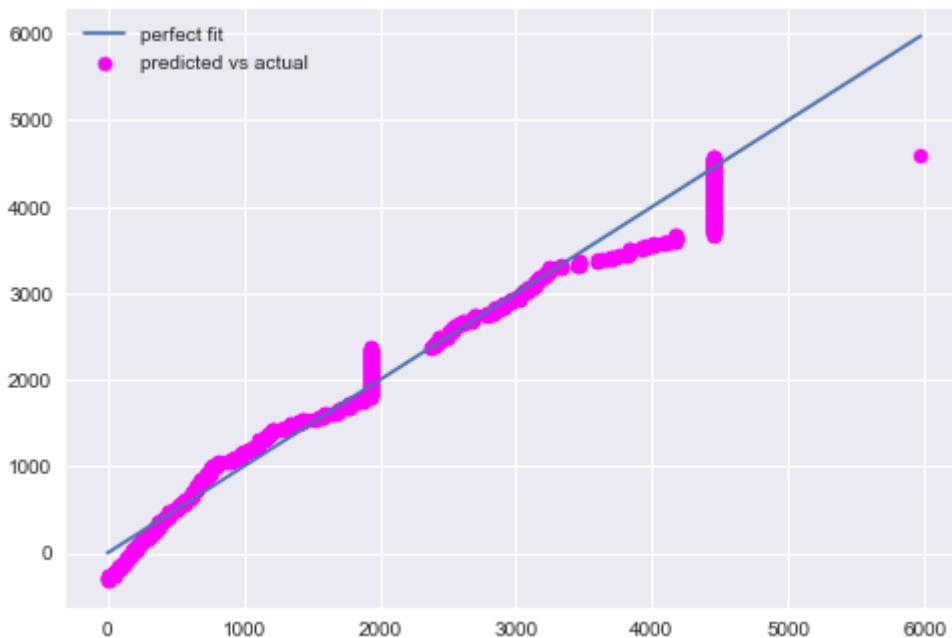
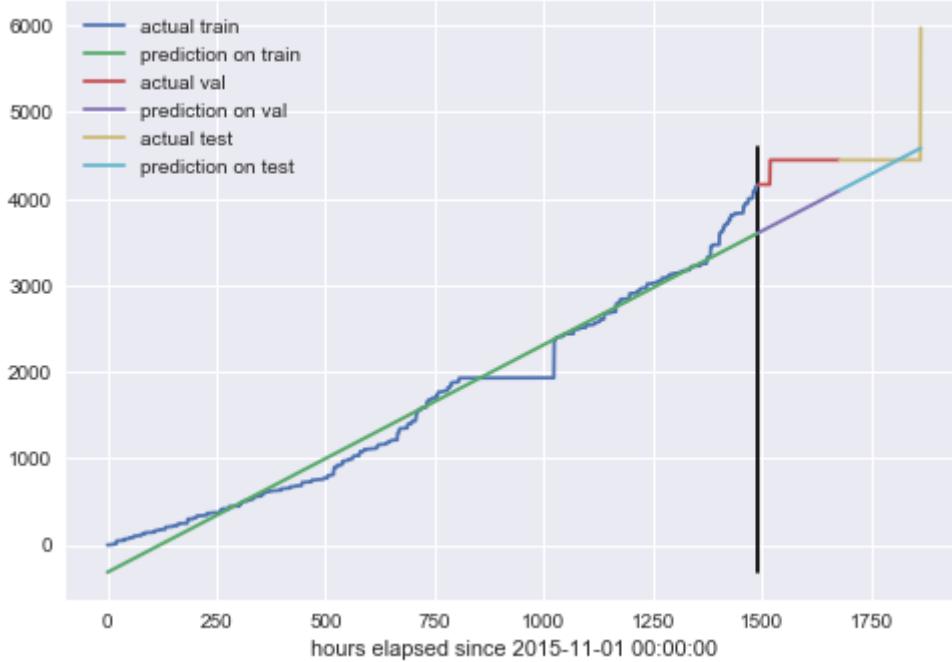
meterid 6412; r2_train: 0.971; r2_val: -101.640; r2_test: -263.575



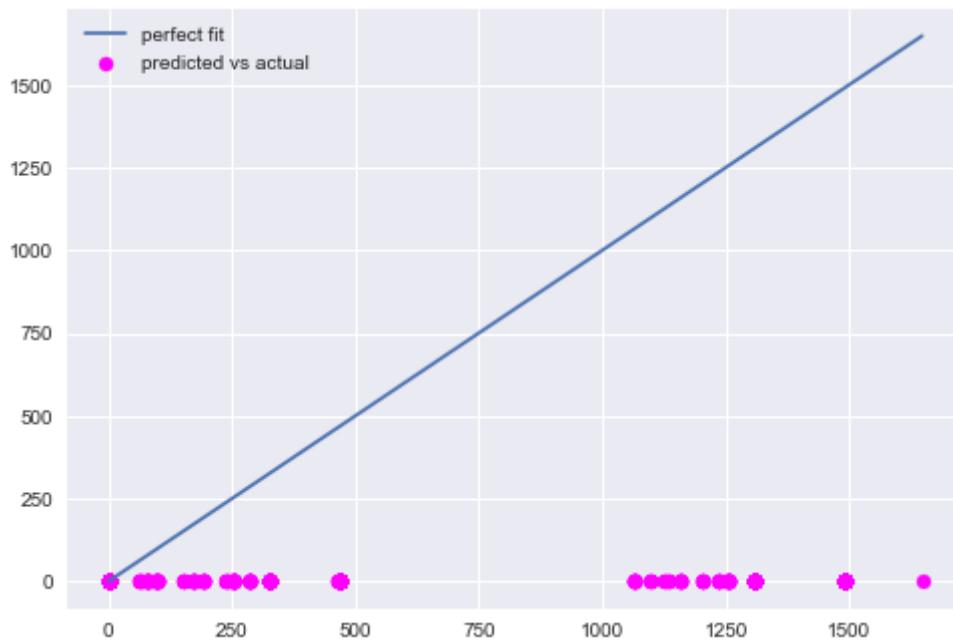
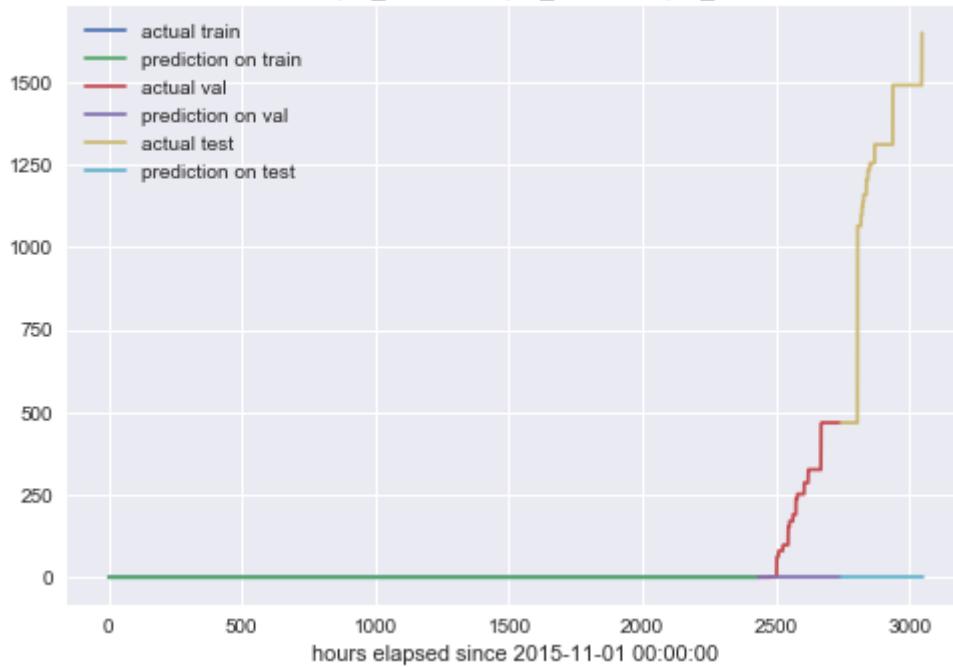
meterid 6505; r2_train: 0.971; r2_val: -28.973; r2_test: -42.797



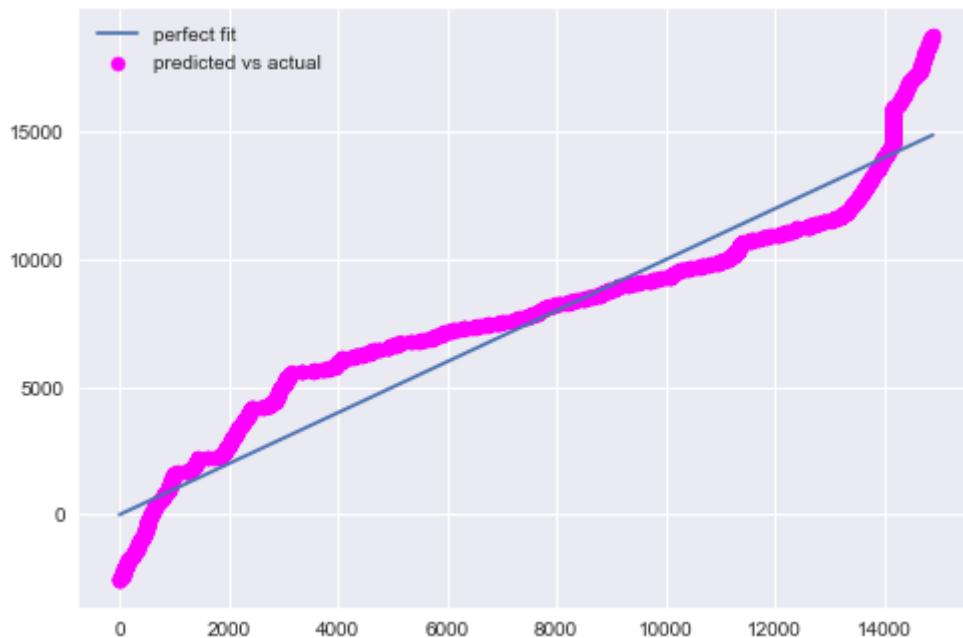
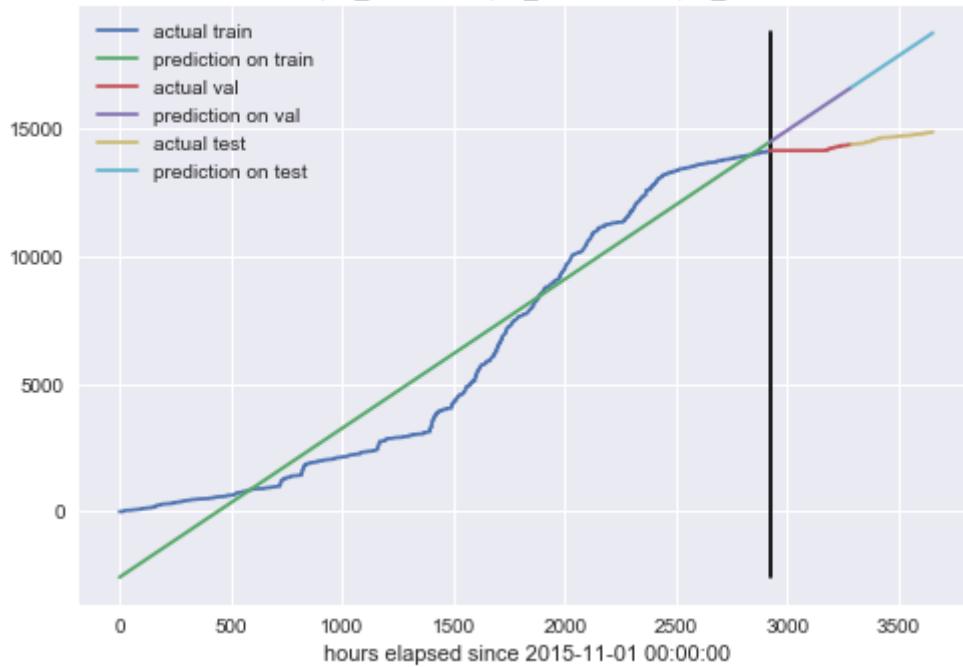
meterid 6578; r2_train: 0.978; r2_val: -31.605; r2_test: -2.369



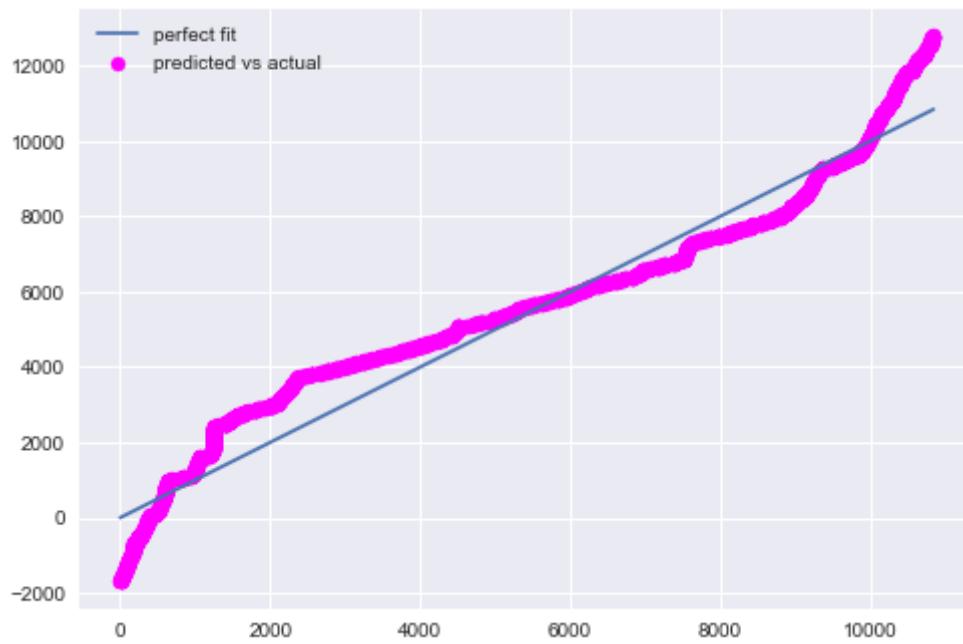
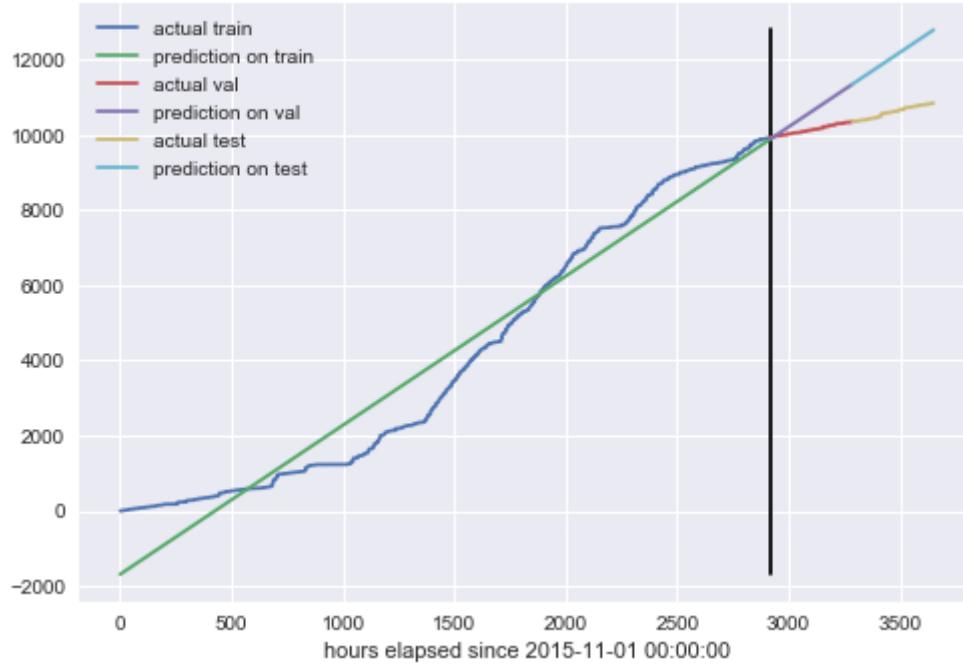
meterid 6685; r2_train: 1.000; r2_val: -1.819; r2_test: -9.507



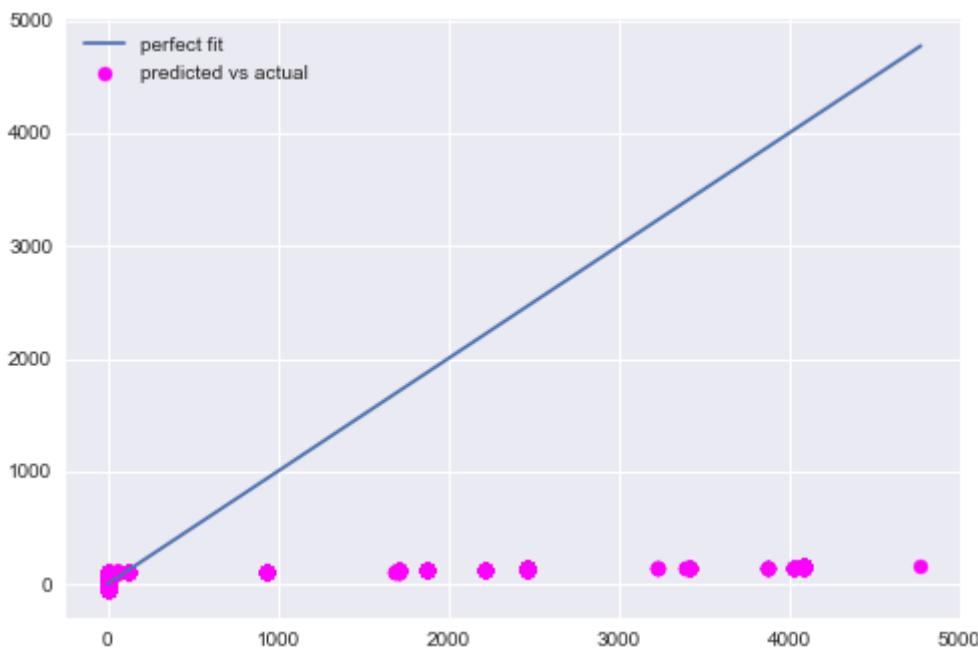
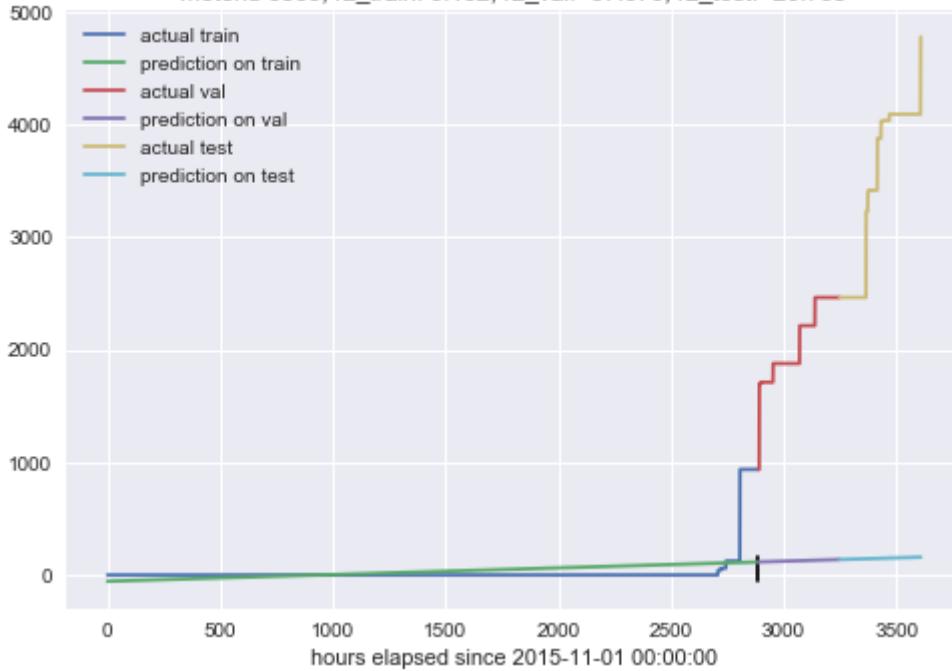
meterid 6830; r2_train: 0.942; r2_val: -347.577; r2_test: -466.768



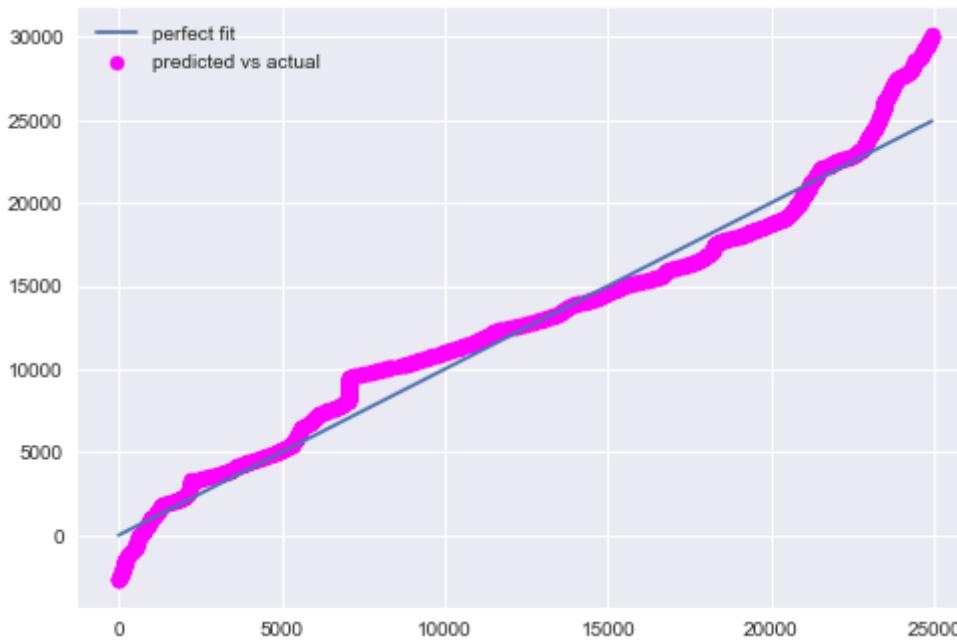
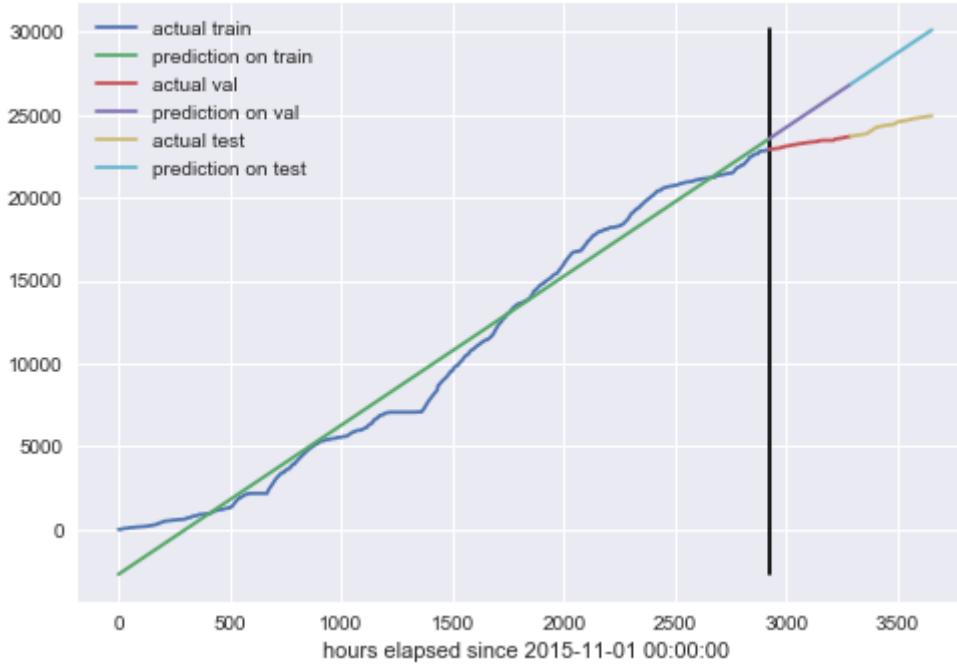
meterid 6836; r2_train: 0.955; r2_val: -19.658; r2_test: -86.188



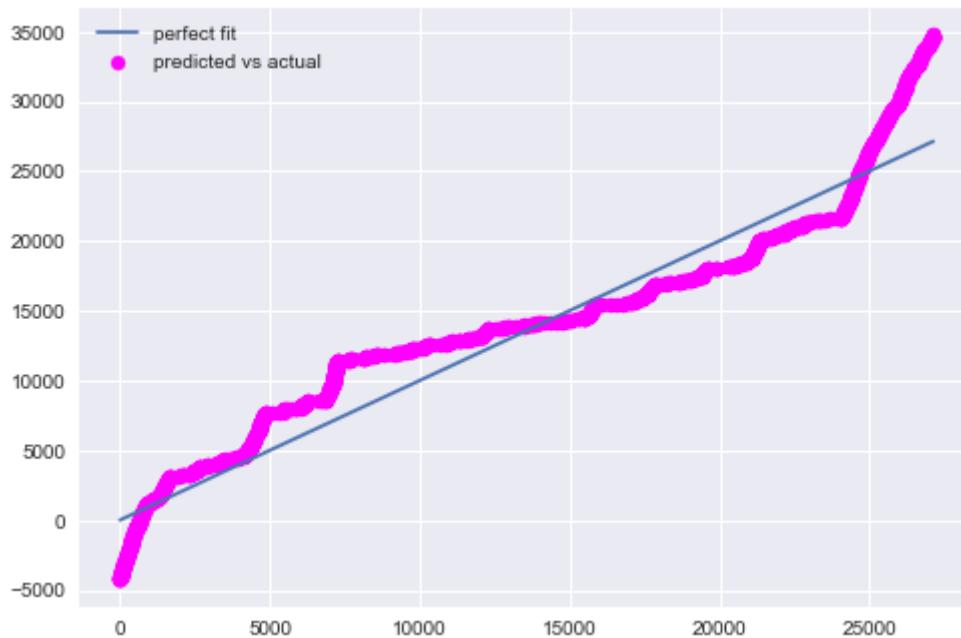
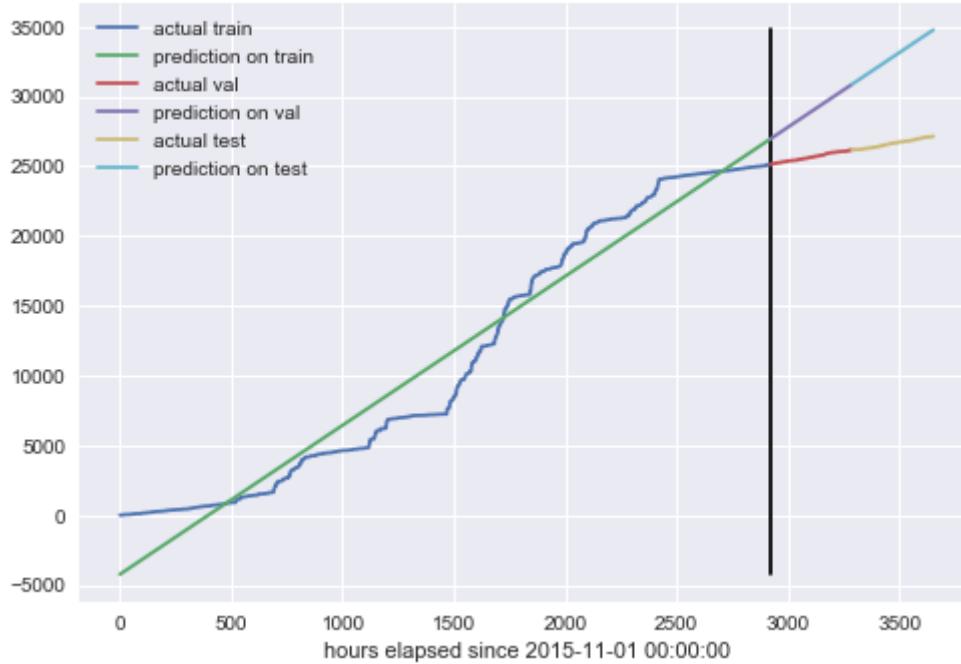
meterid 6863; r2_train: 0.102; r2_val: -37.070; r2_test: -20.738



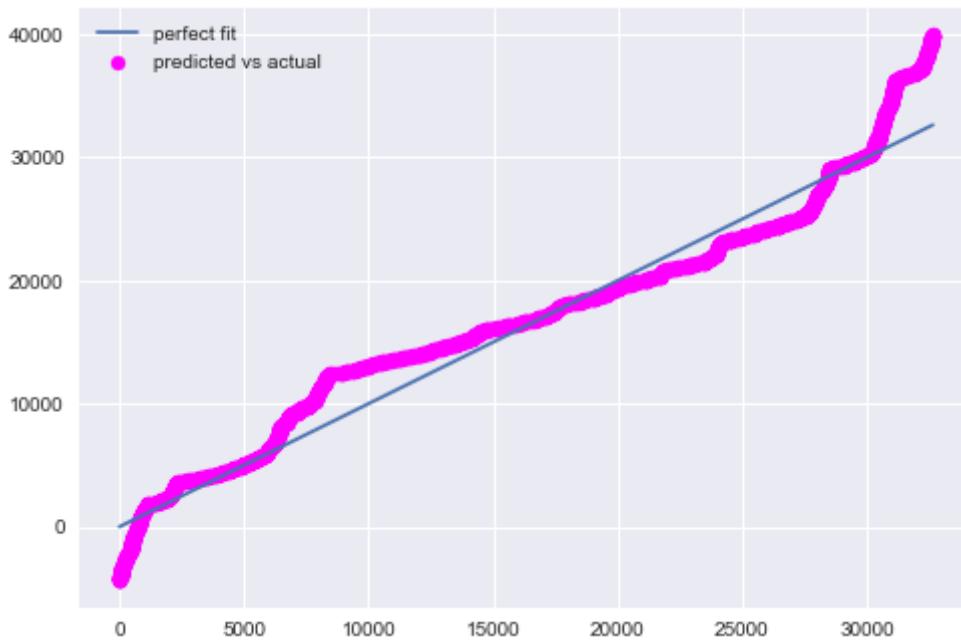
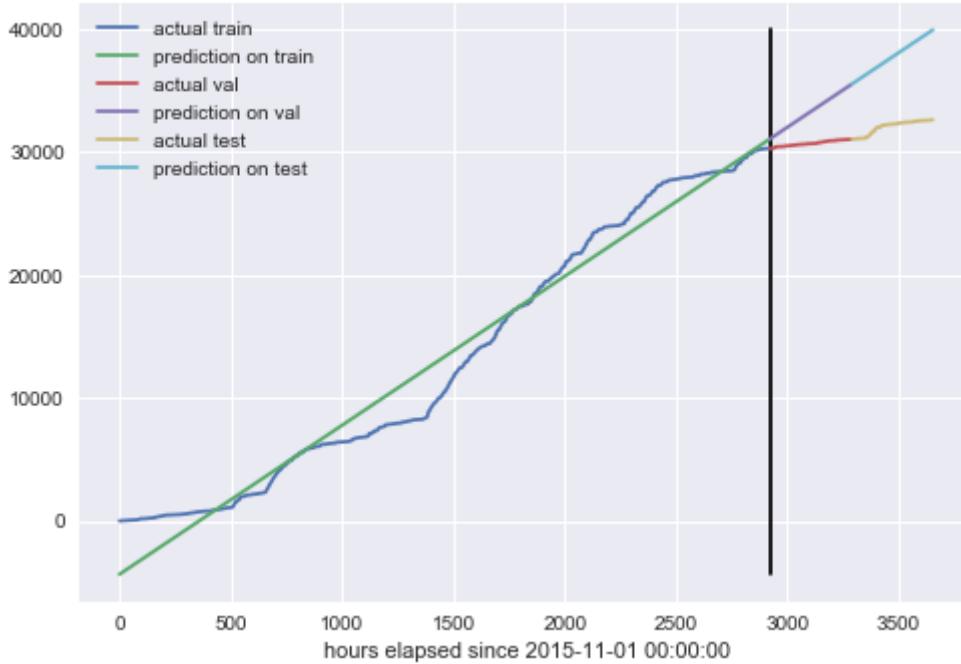
meterid 6910; r2_train: 0.982; r2_val: -84.391; r2_test: -108.968



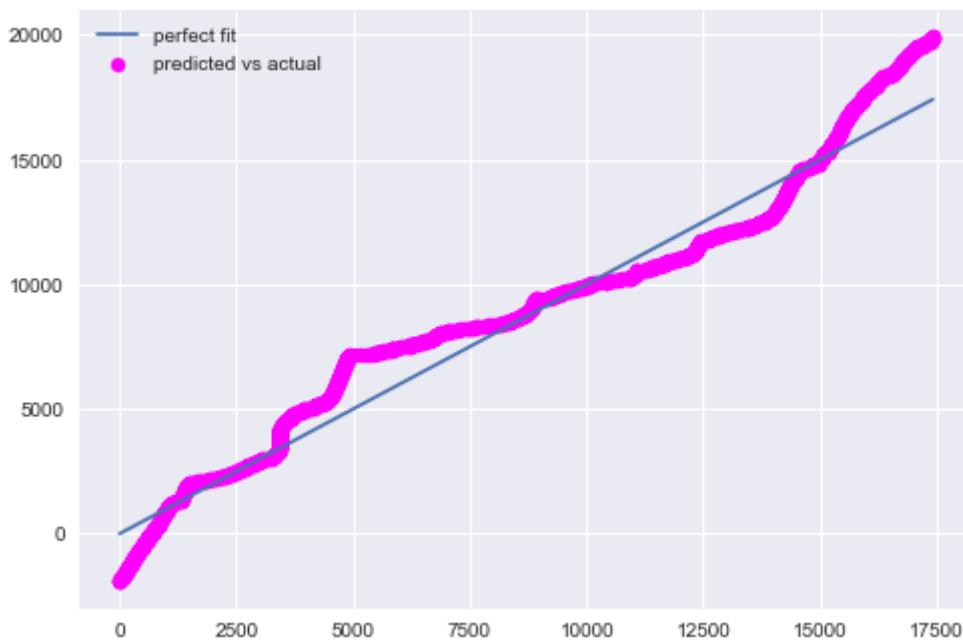
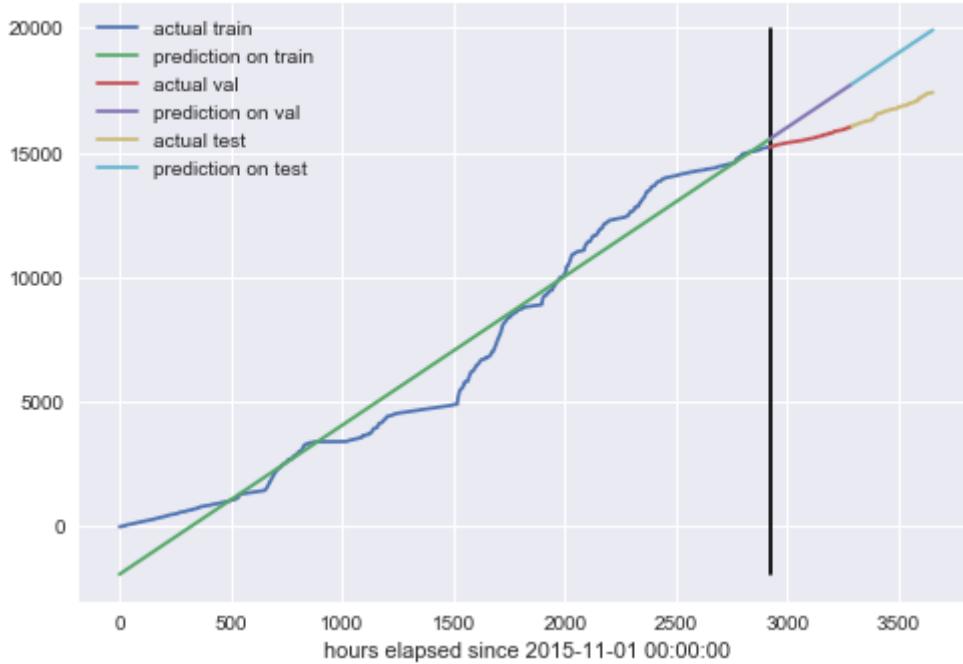
meterid 7016; r2_train: 0.958; r2_val: -119.023; r2_test: -417.982



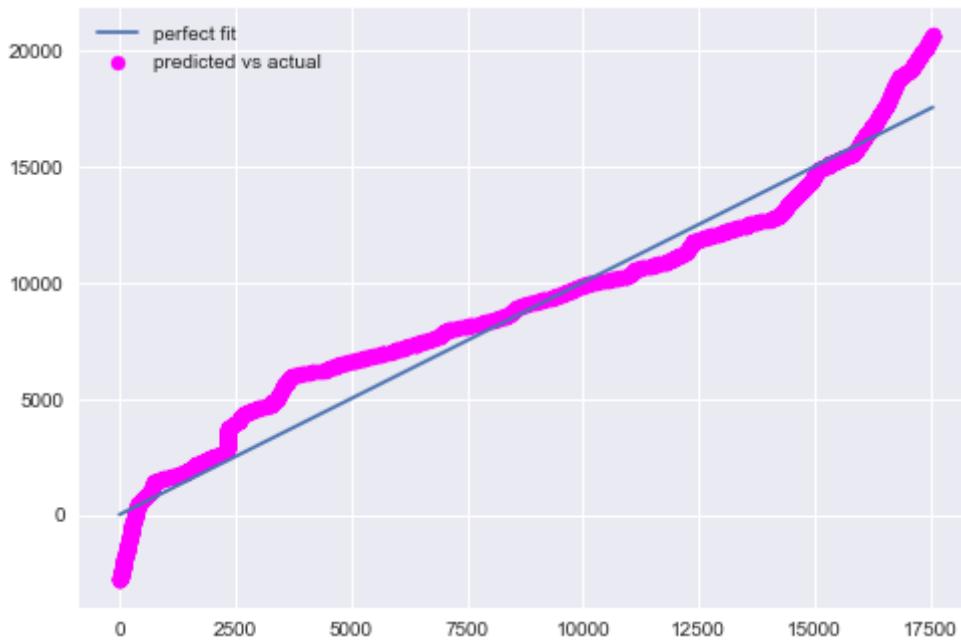
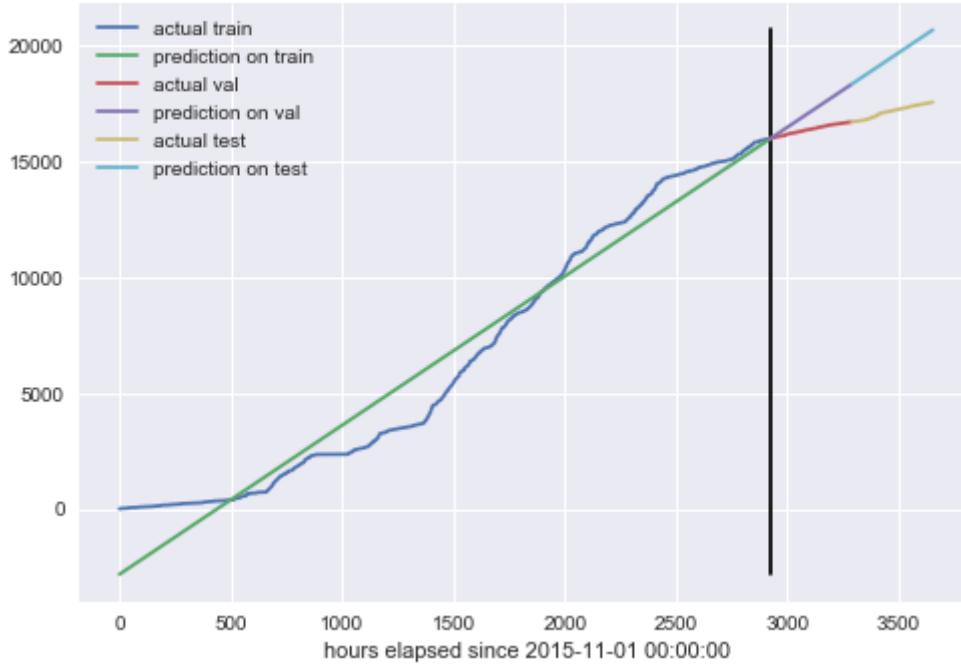
meterid 7017; r2_train: 0.972; r2_val: -167.343; r2_test: -108.003



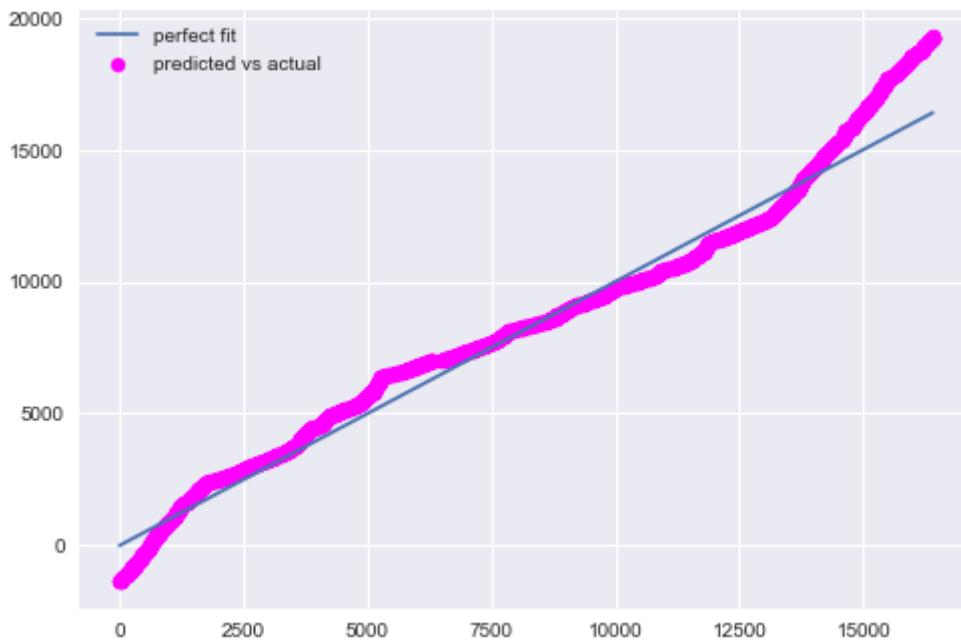
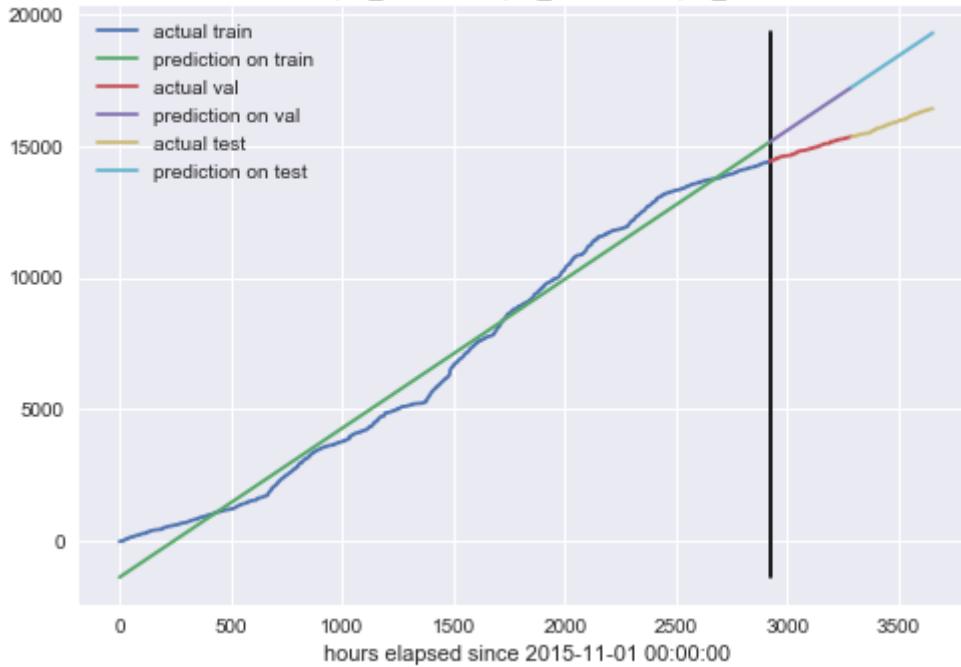
meterid 7030; r2_train: 0.968; r2_val: -23.889; r2_test: -27.928



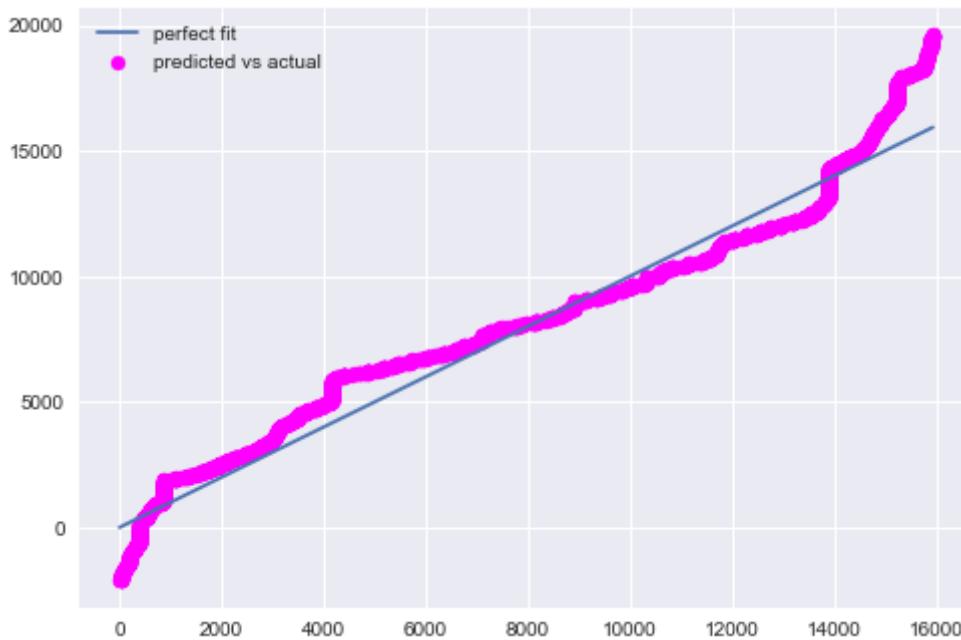
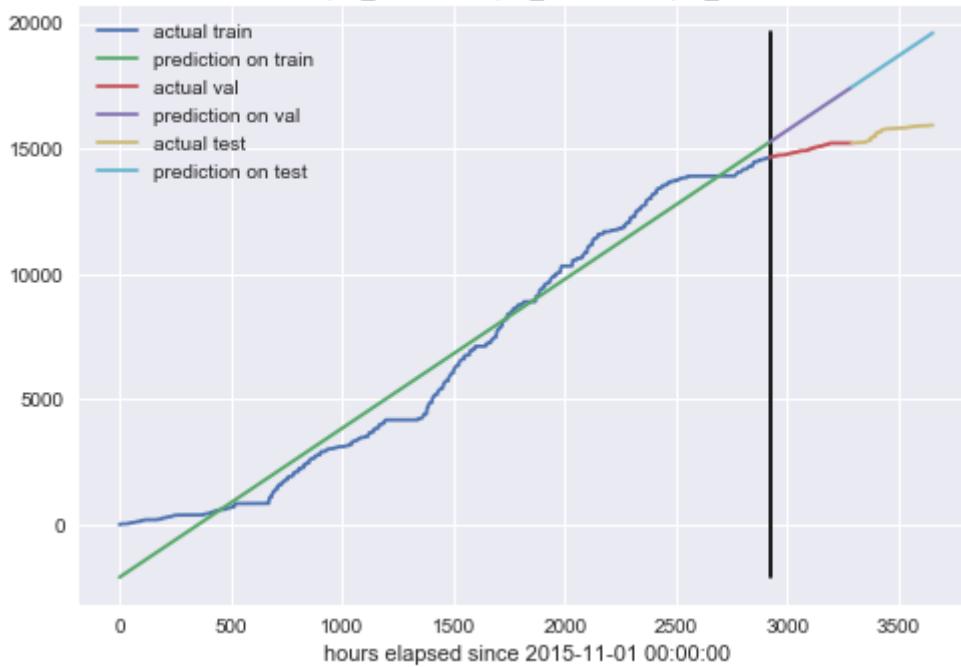
meterid 7117; r2_train: 0.958; r2_val: -16.901; r2_test: -78.029

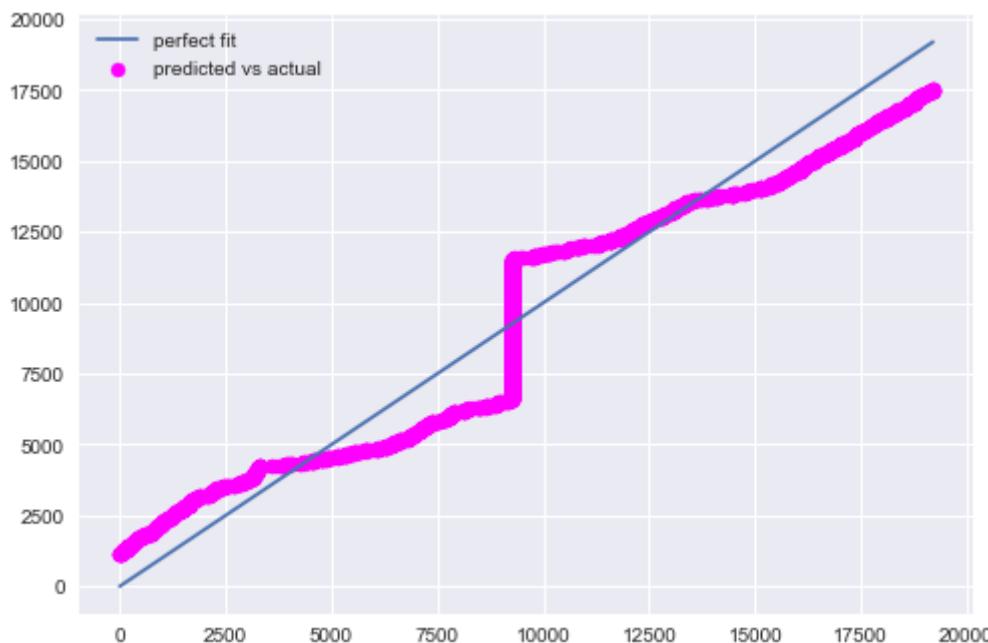
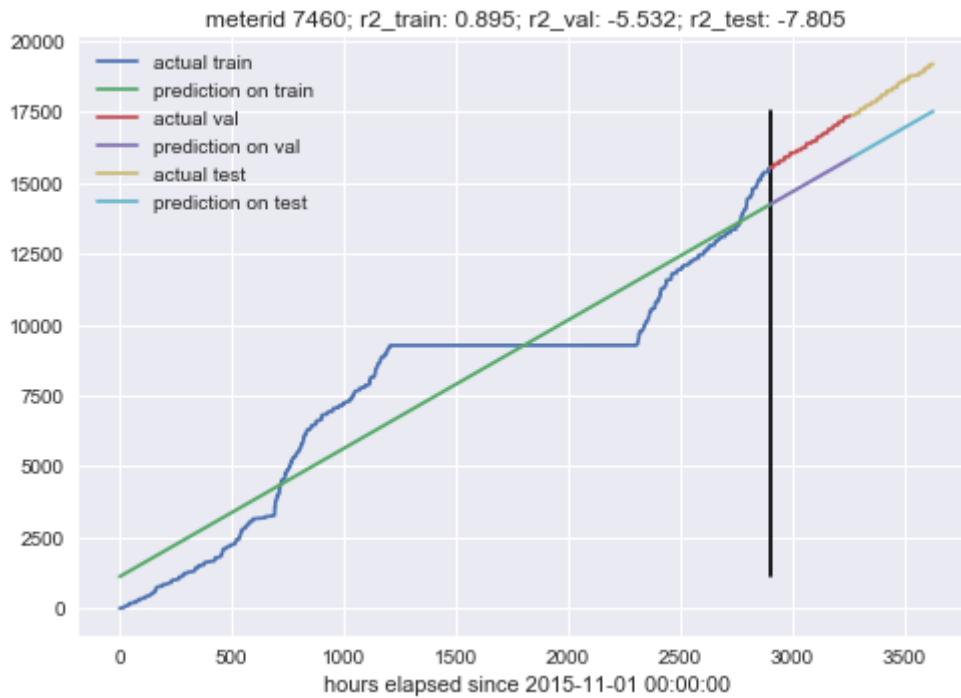


meterid 7287; r2_train: 0.987; r2_val: -23.405; r2_test: -52.410

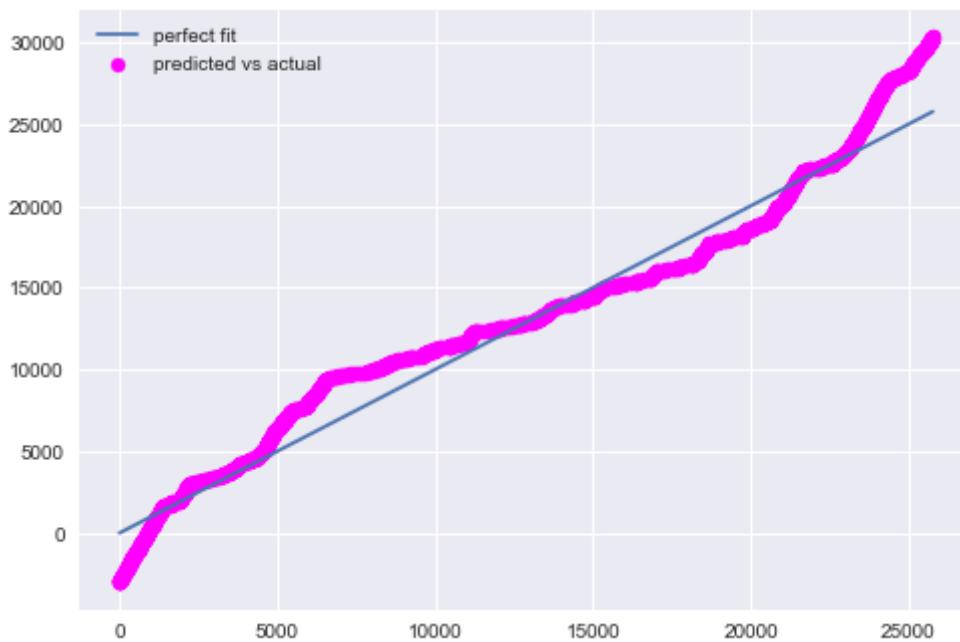
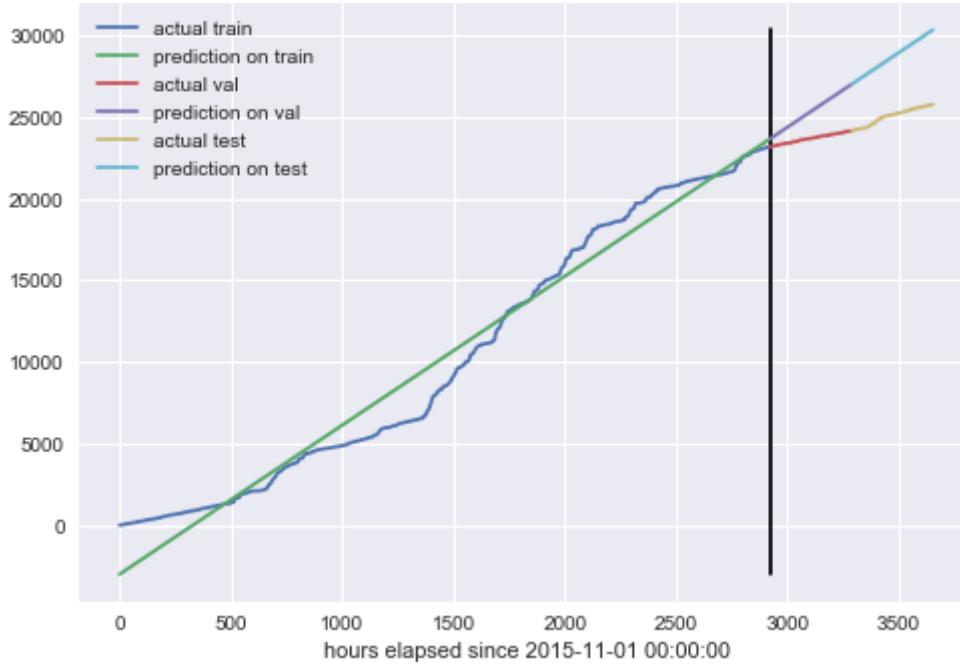


meterid 7429; r2_train: 0.974; r2_val: -54.653; r2_test: -125.851

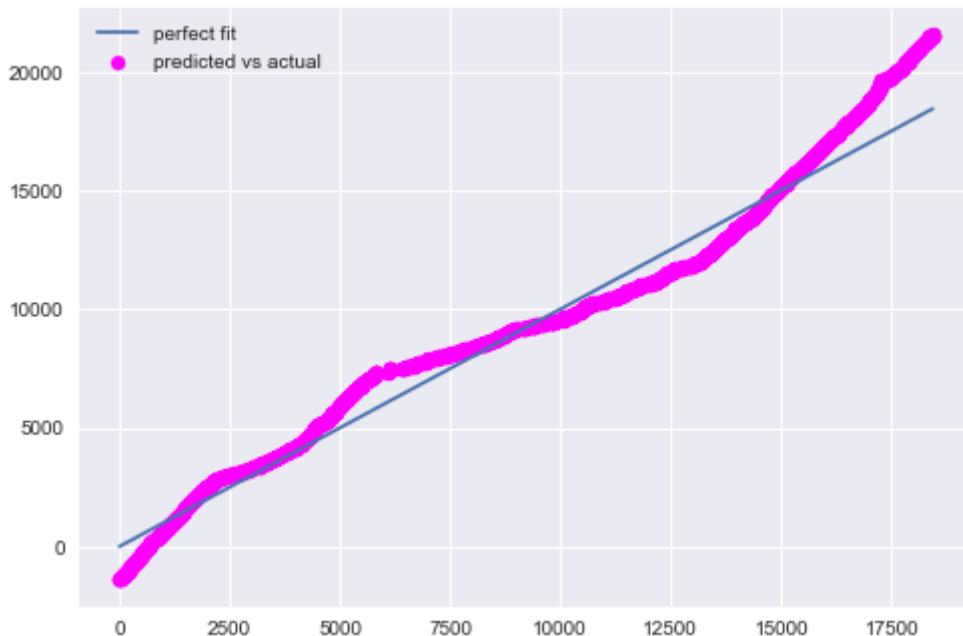
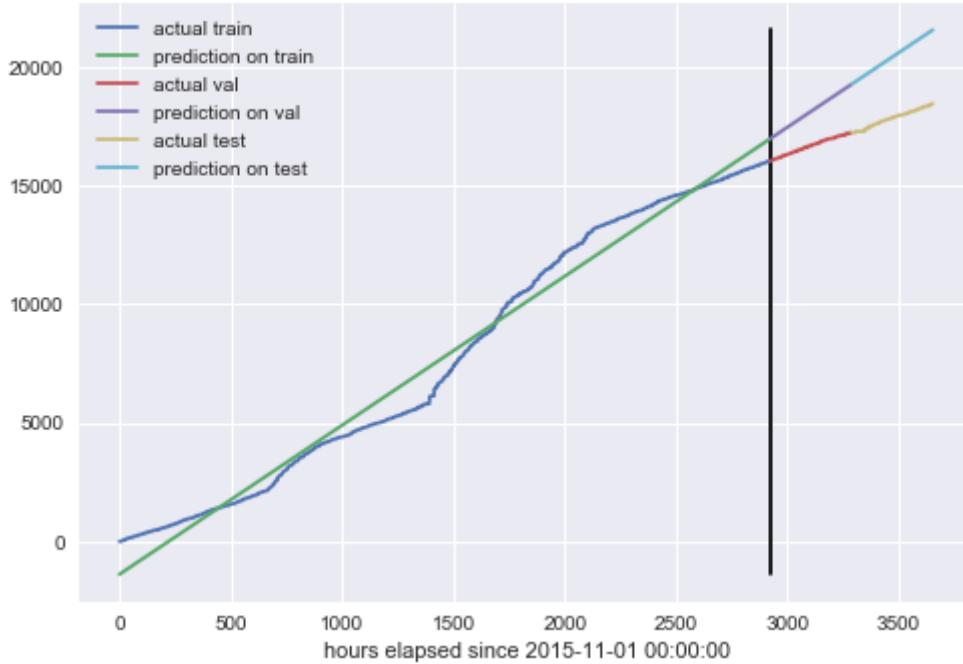




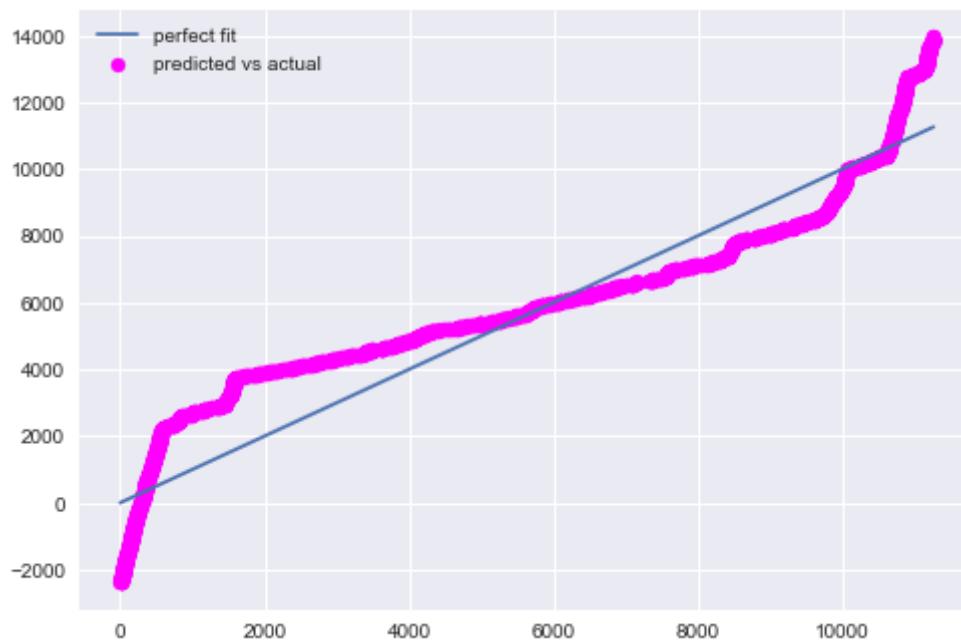
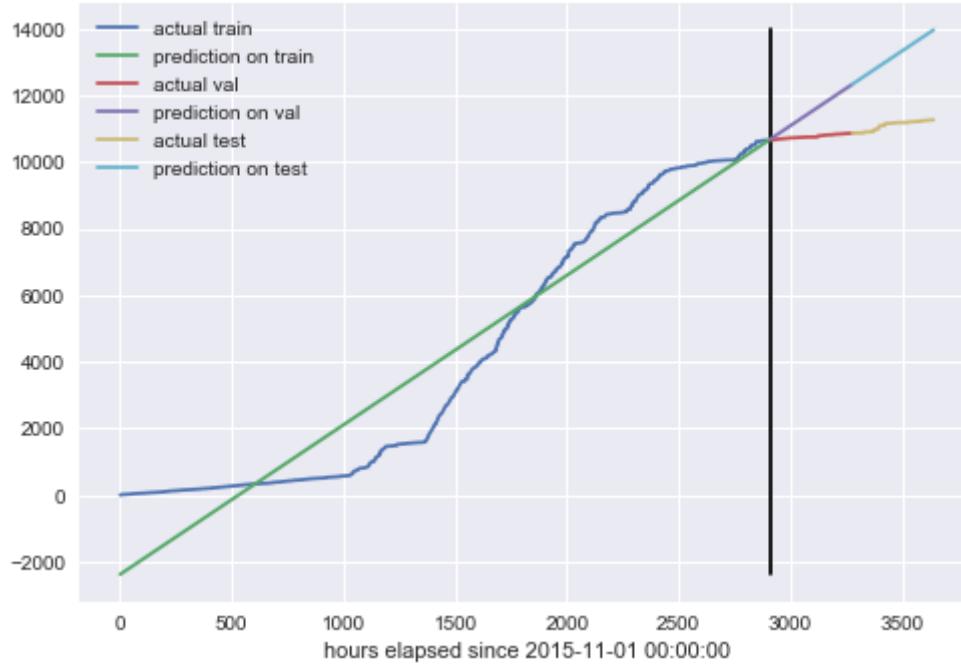
meterid 7674; r2_train: 0.972; r2_val: -40.066; r2_test: -51.355



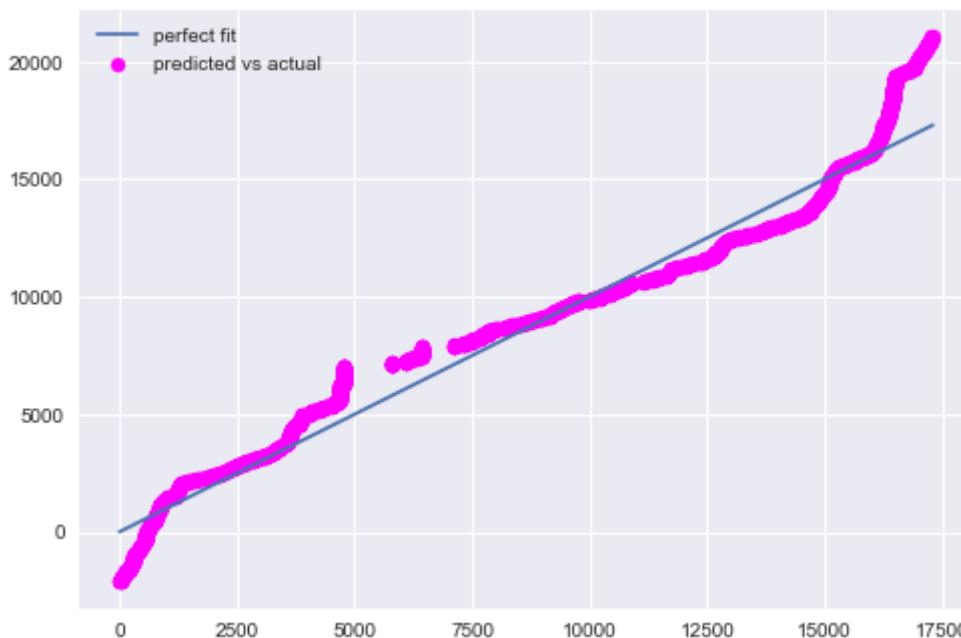
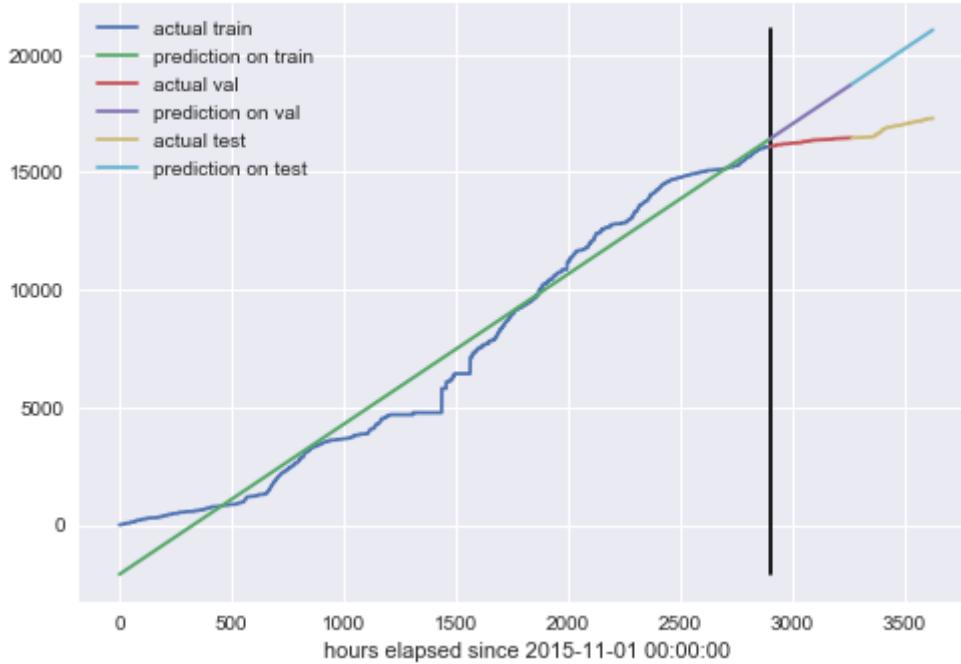
meterid 7682; r2_train: 0.982; r2_val: -16.603; r2_test: -49.515



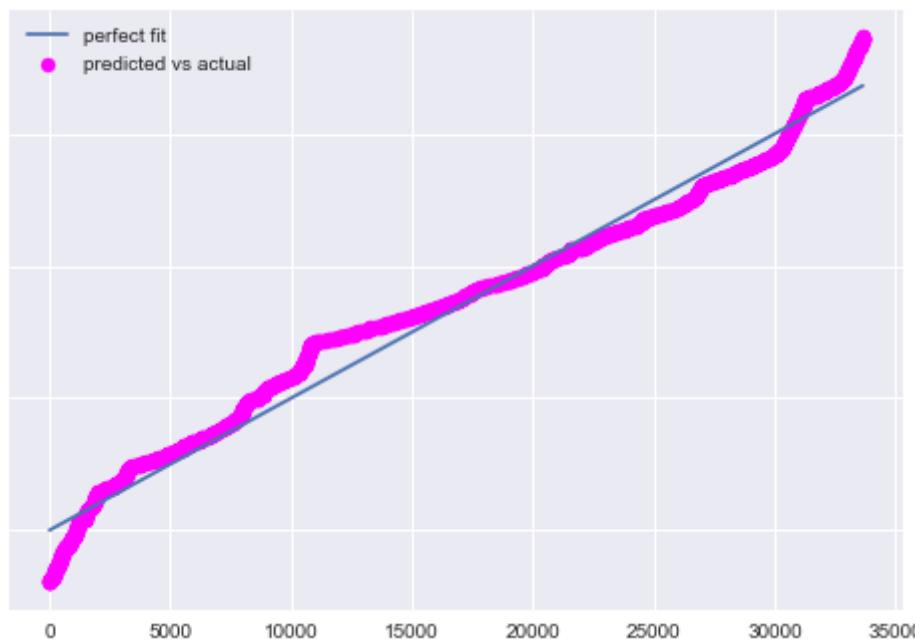
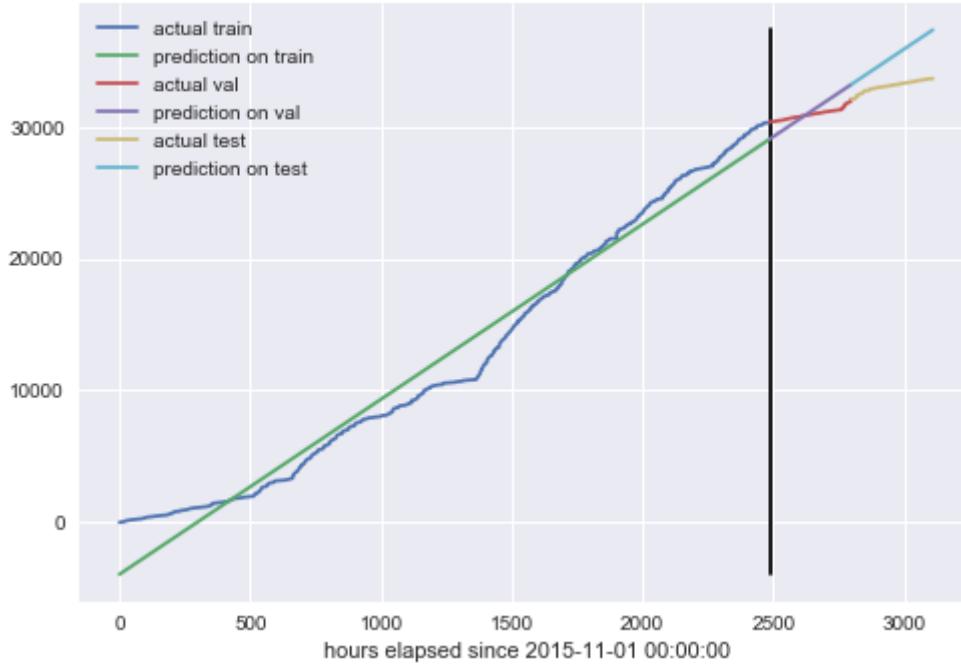
meterid 7739; r2_train: 0.919; r2_val: -205.358; r2_test: -212.357



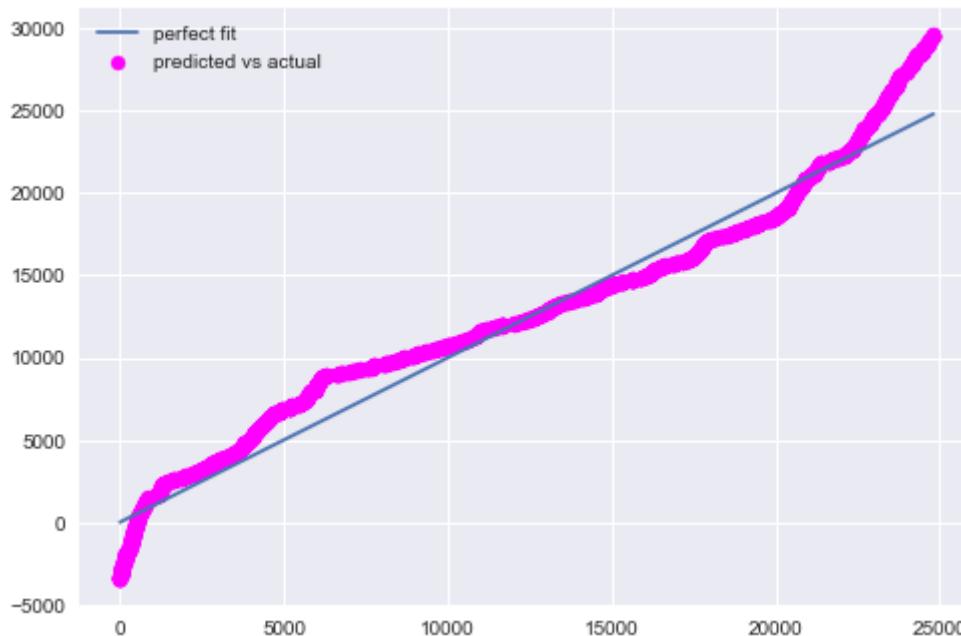
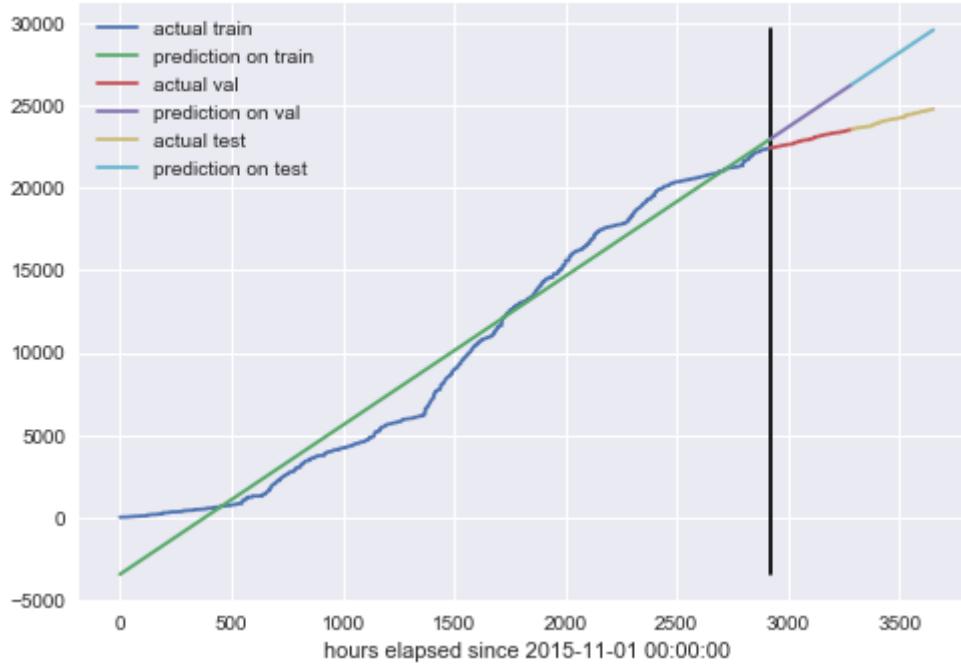
meterid 7741; r2_train: 0.973; r2_val: -167.557; r2_test: -110.263



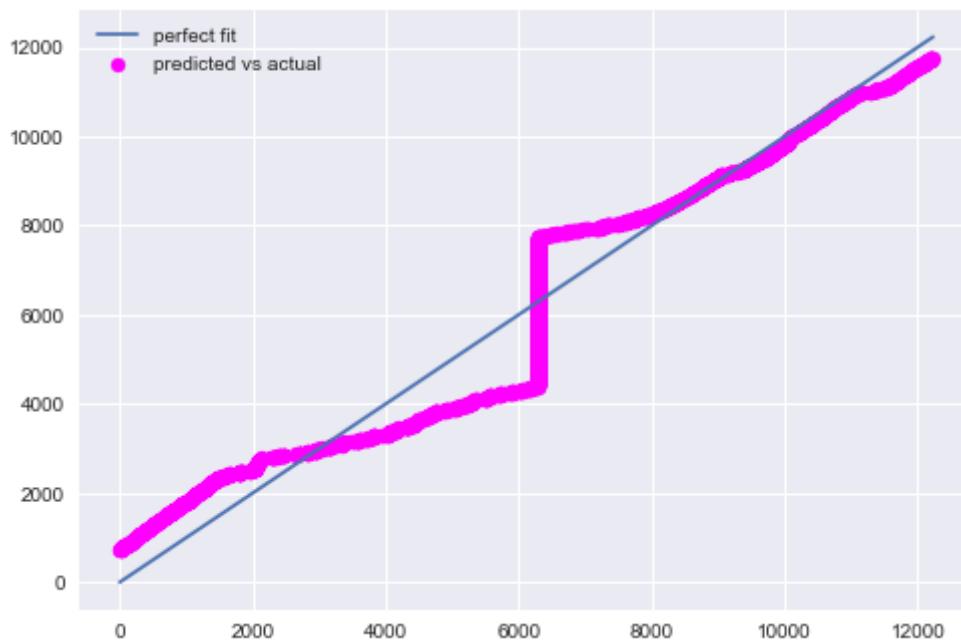
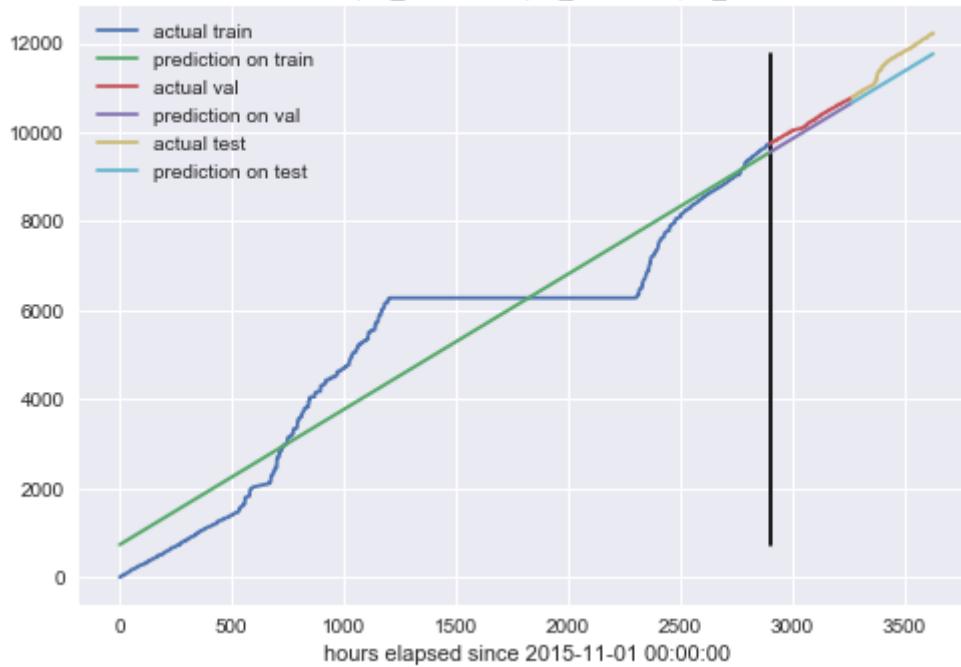
meterid 7794; r2_train: 0.975; r2_val: -3.530; r2_test: -30.835



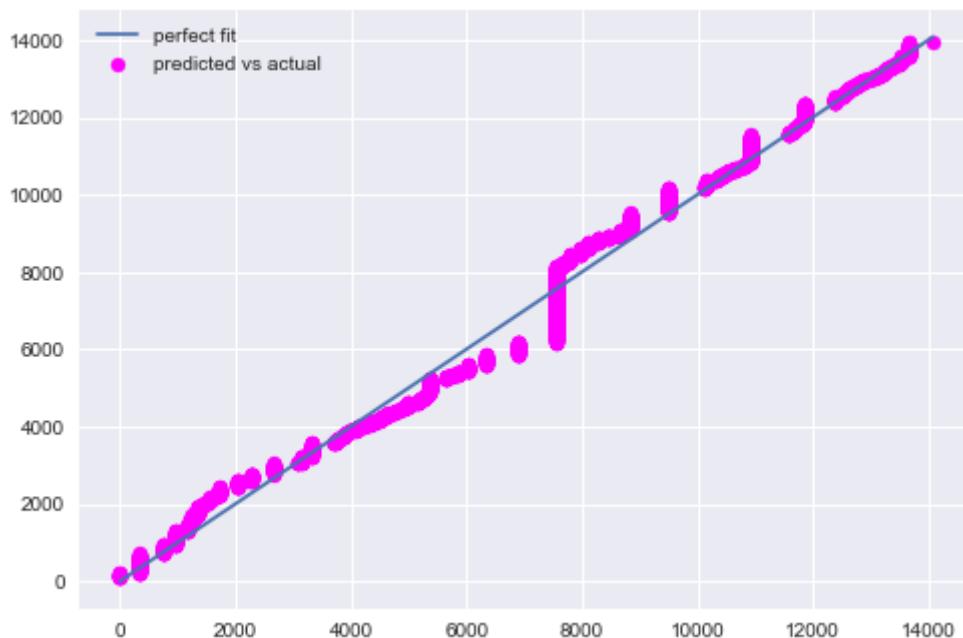
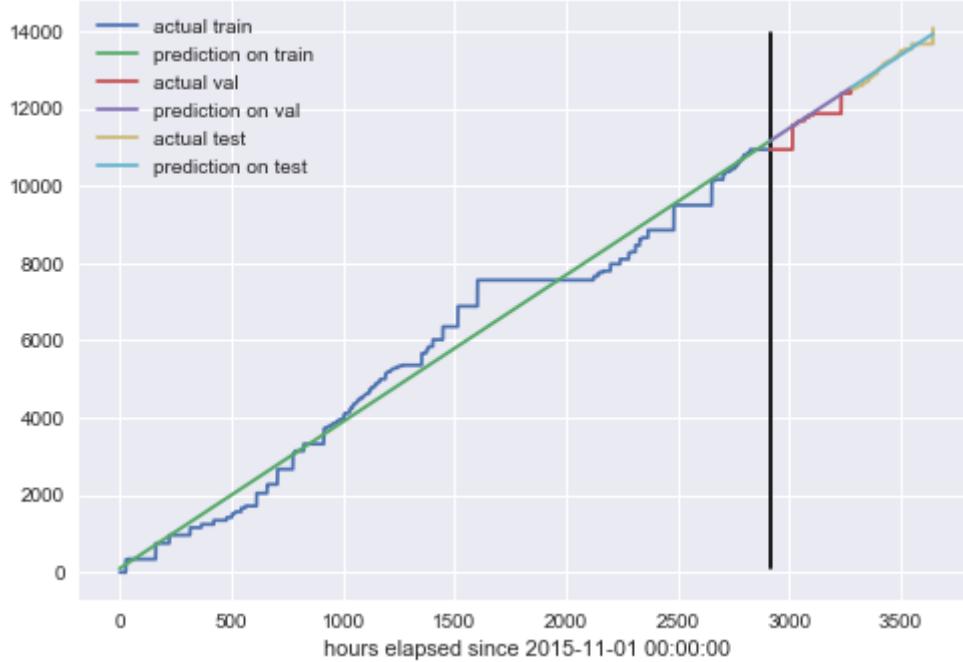
meterid 7900; r2_train: 0.972; r2_val: -25.629; r2_test: -105.616



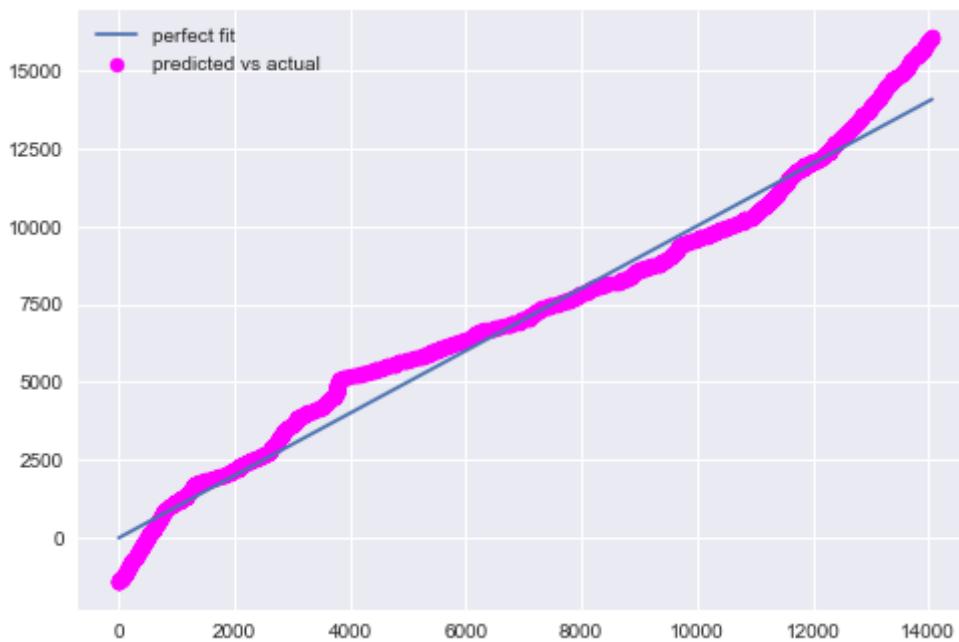
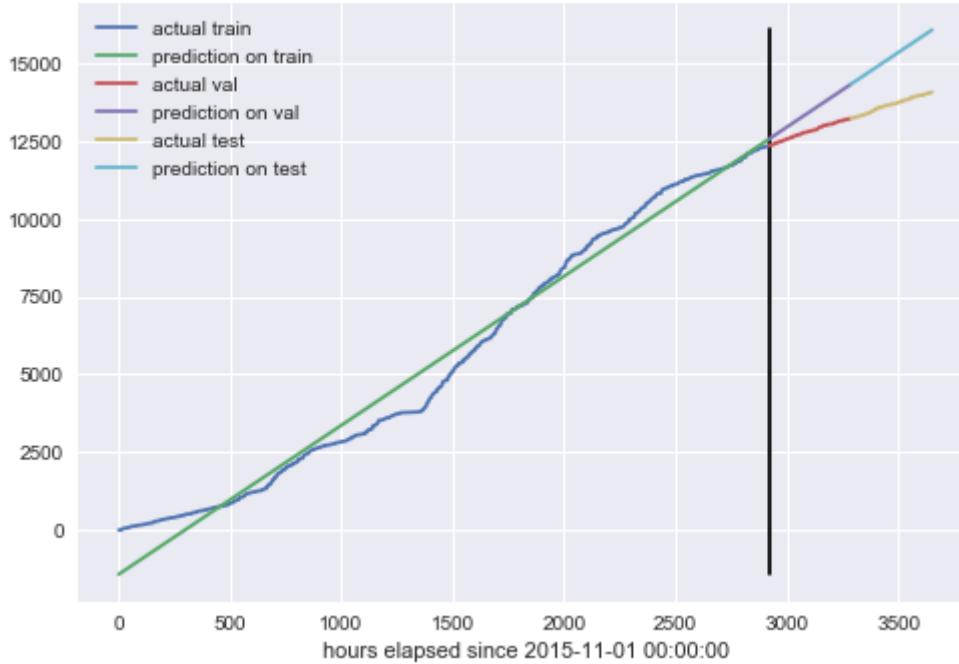
meterid 7919; r2_train: 0.899; r2_val: 0.746; r2_test: 0.322

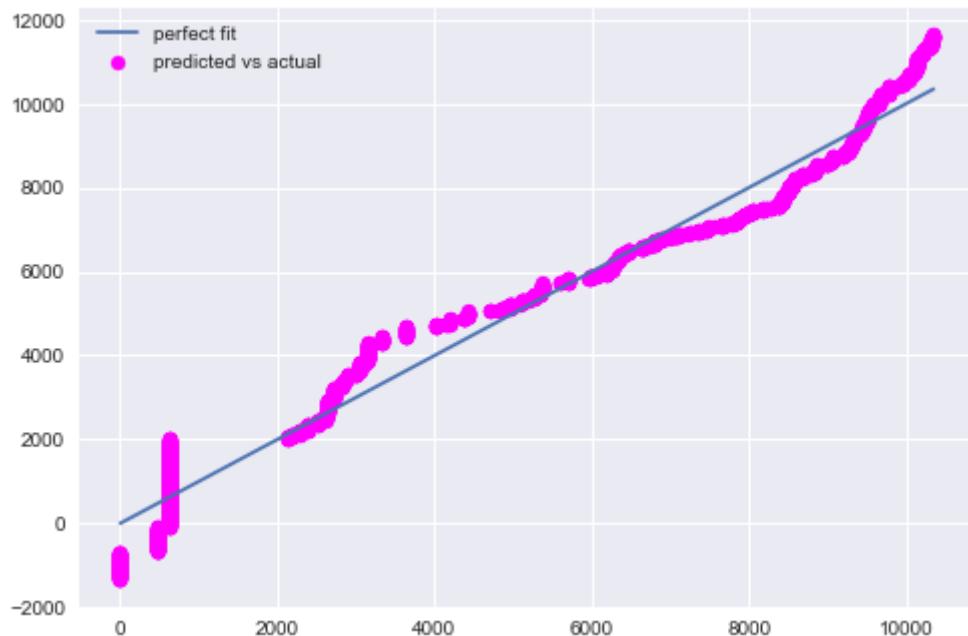
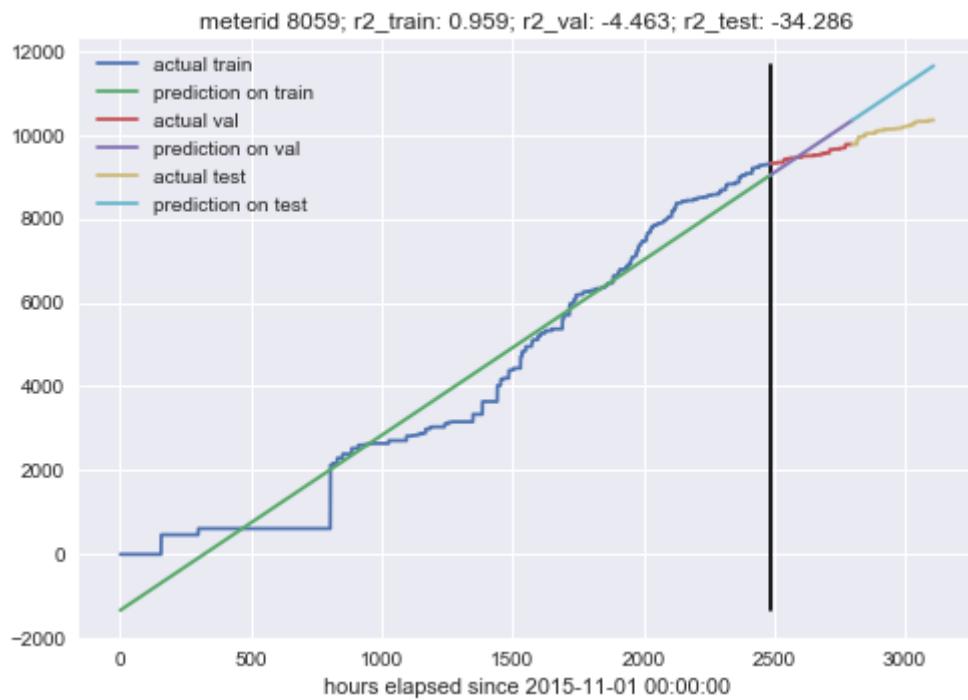


meterid 7965; r2_train: 0.979; r2_val: 0.643; r2_test: 0.953

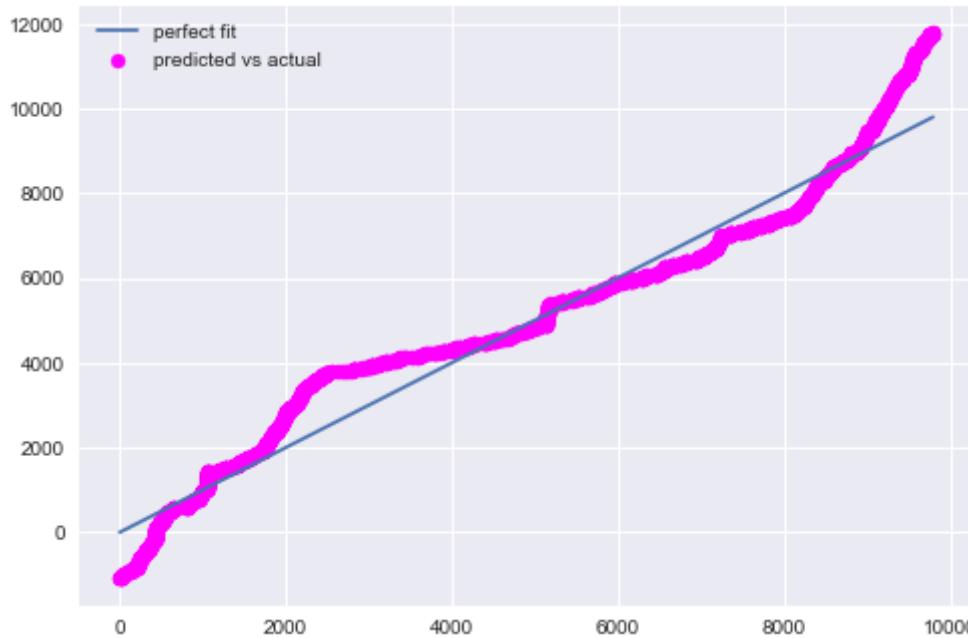
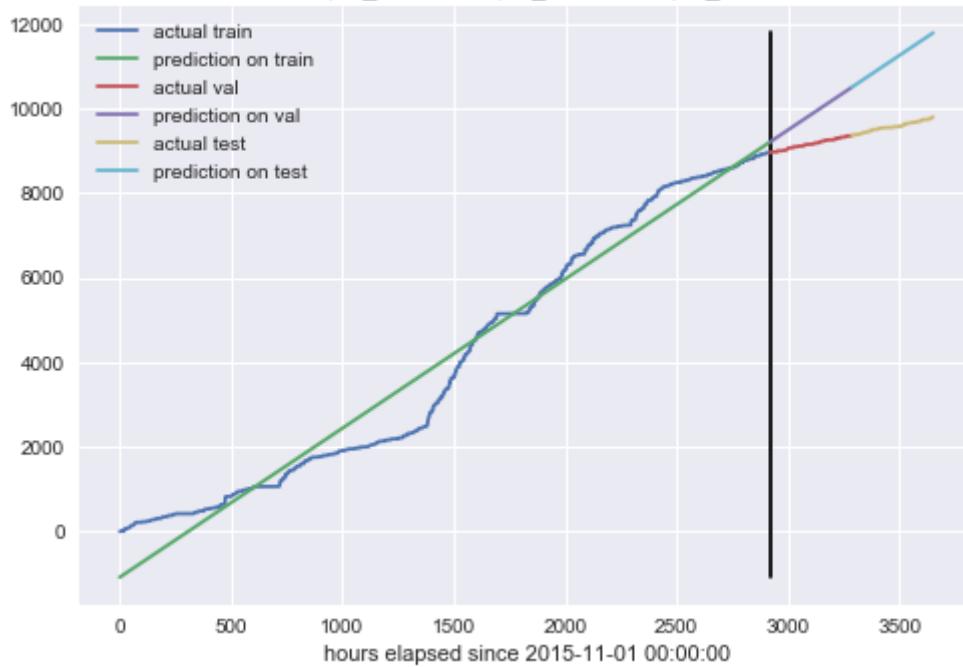


meterid 7989; r2_train: 0.981; r2_val: -5.954; r2_test: -36.778

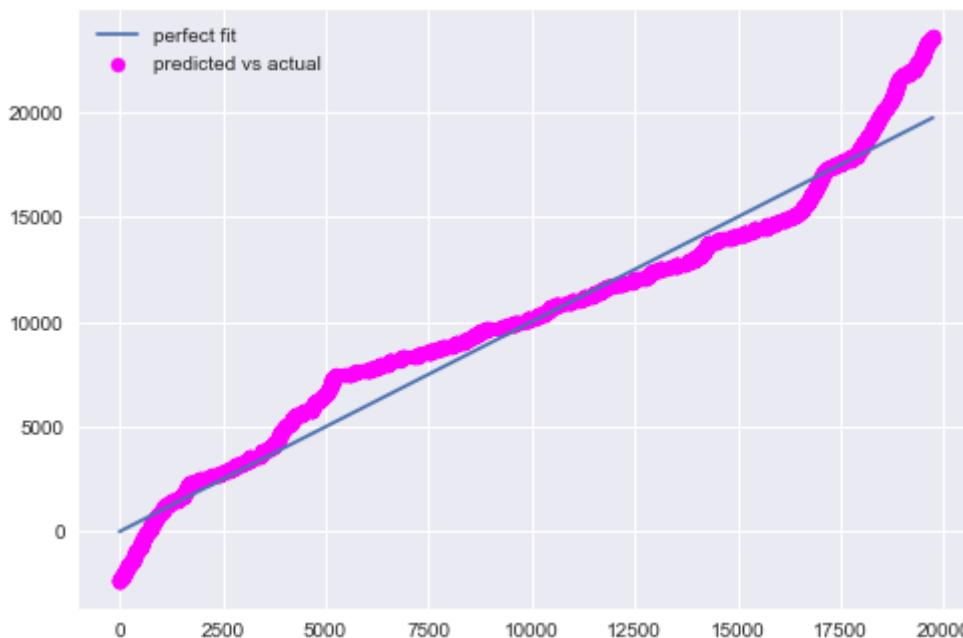
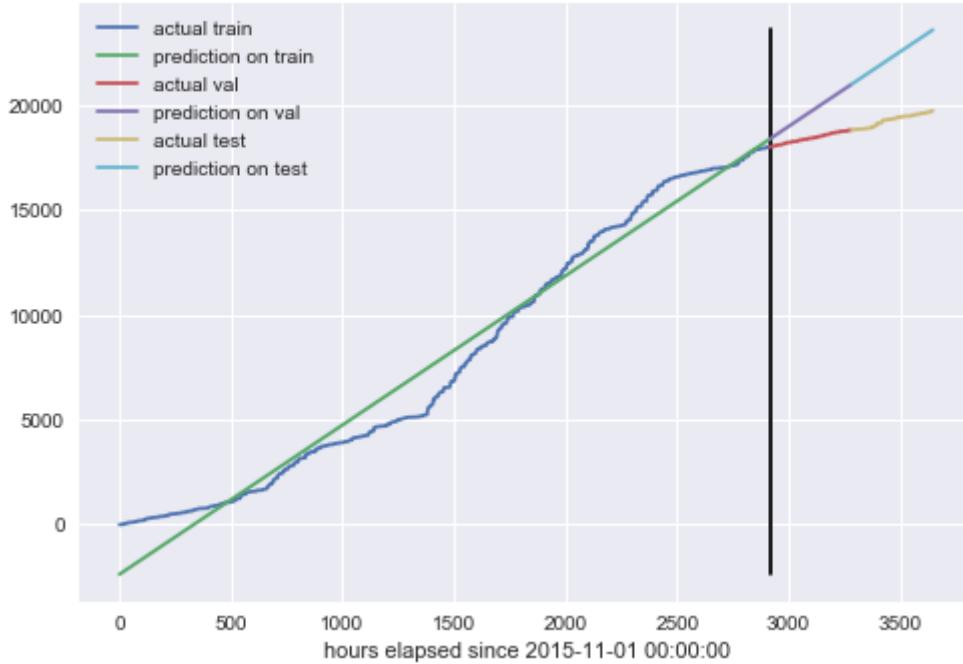




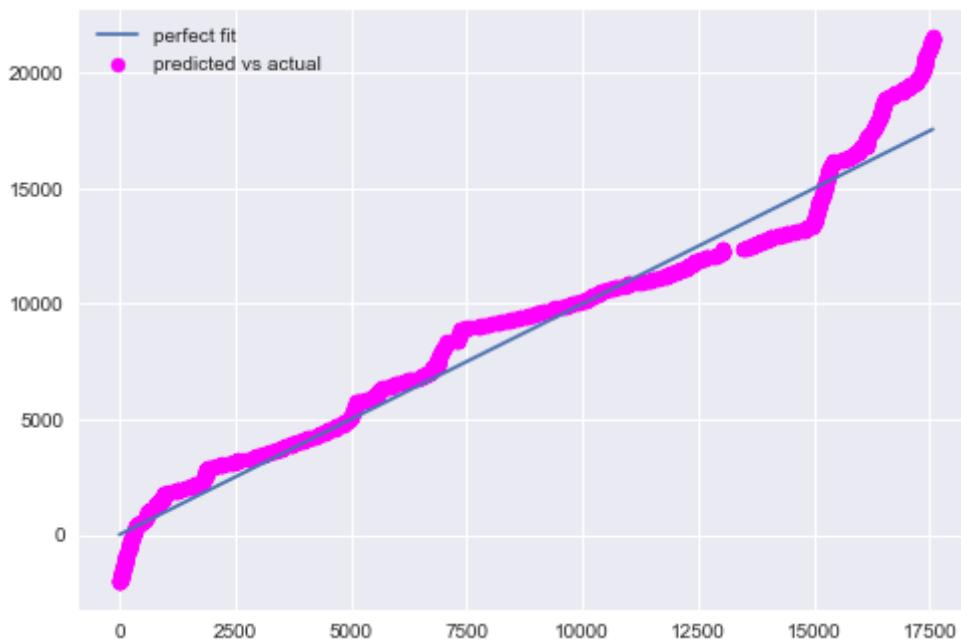
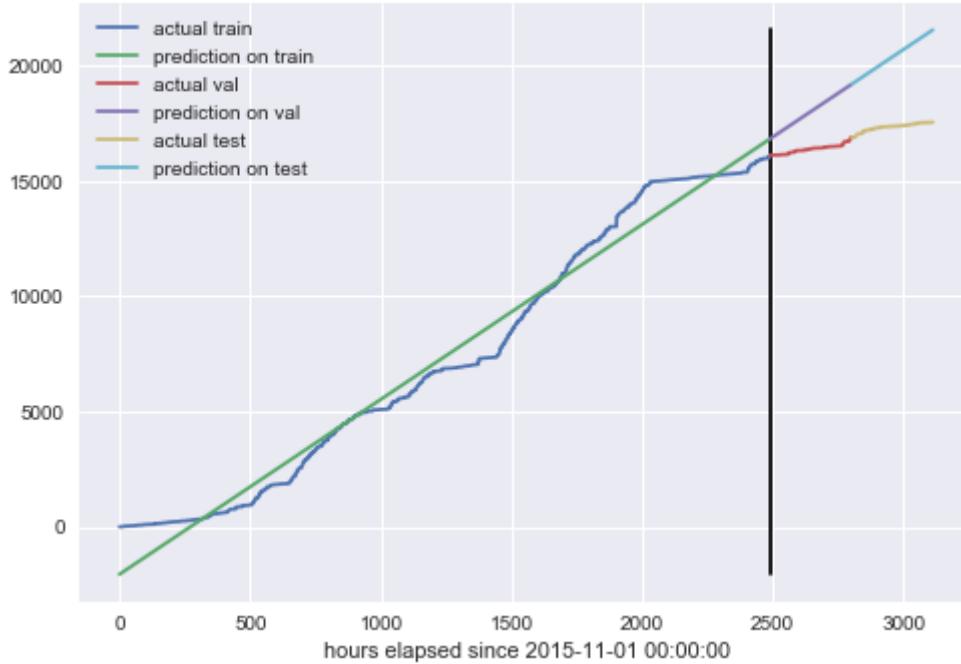
meterid 8084; r2_train: 0.968; r2_val: -37.507; r2_test: -184.915



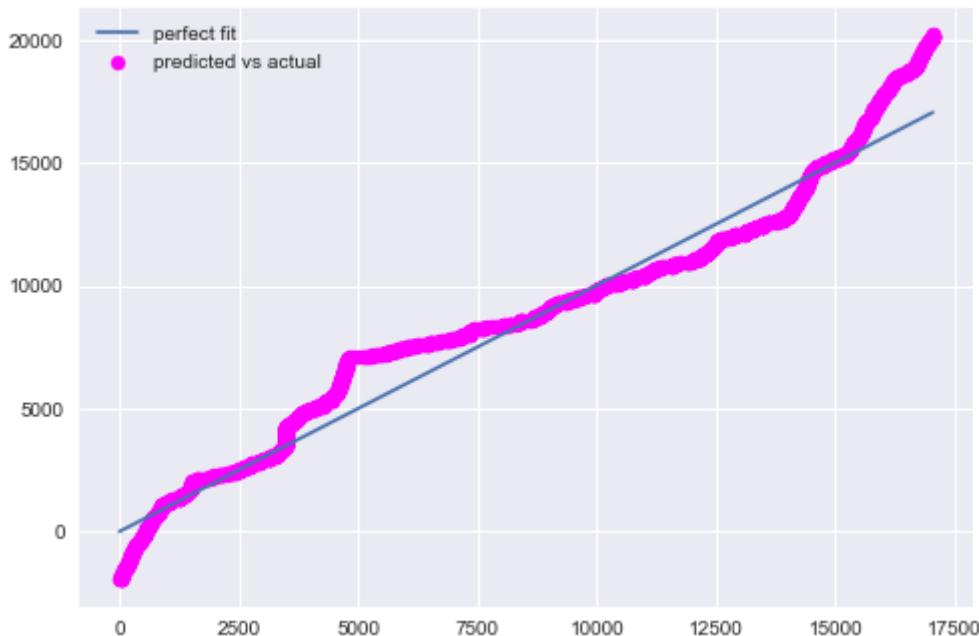
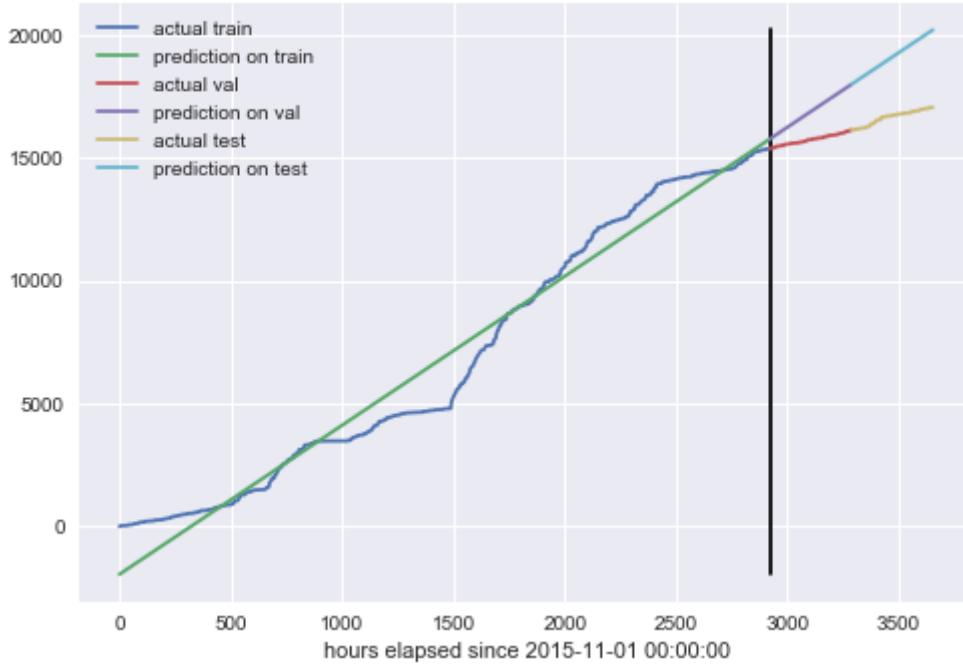
meterid 8086; r2_train: 0.974; r2_val: -31.824; r2_test: -108.244



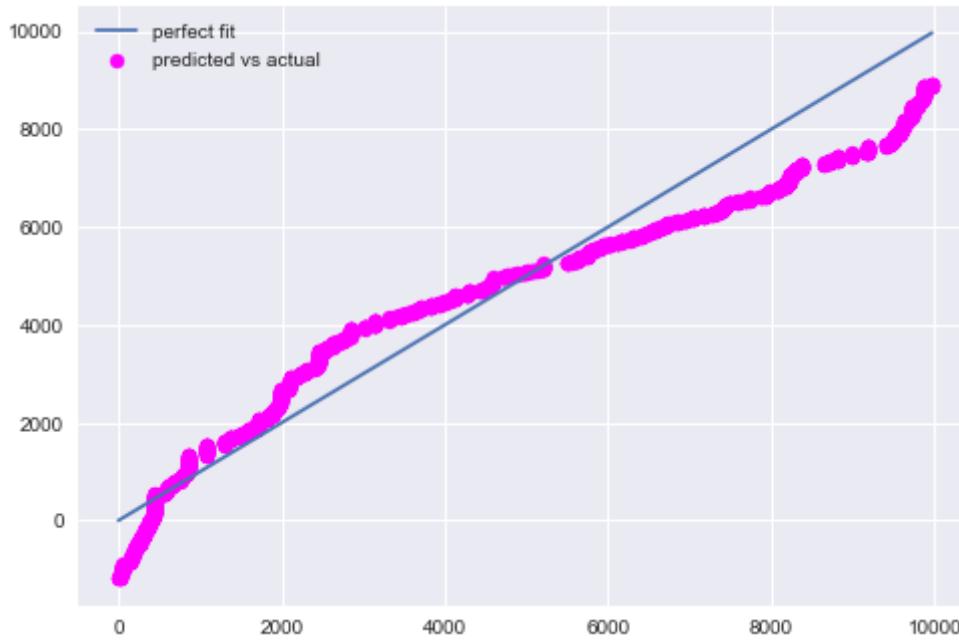
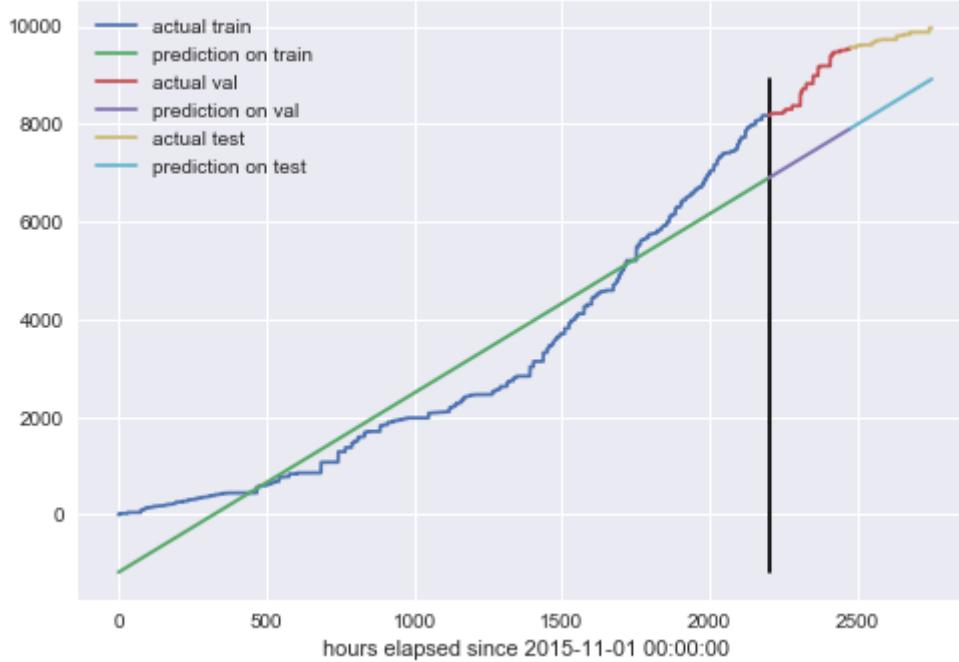
meterid 8155; r2_train: 0.979; r2_val: -87.925; r2_test: -309.409



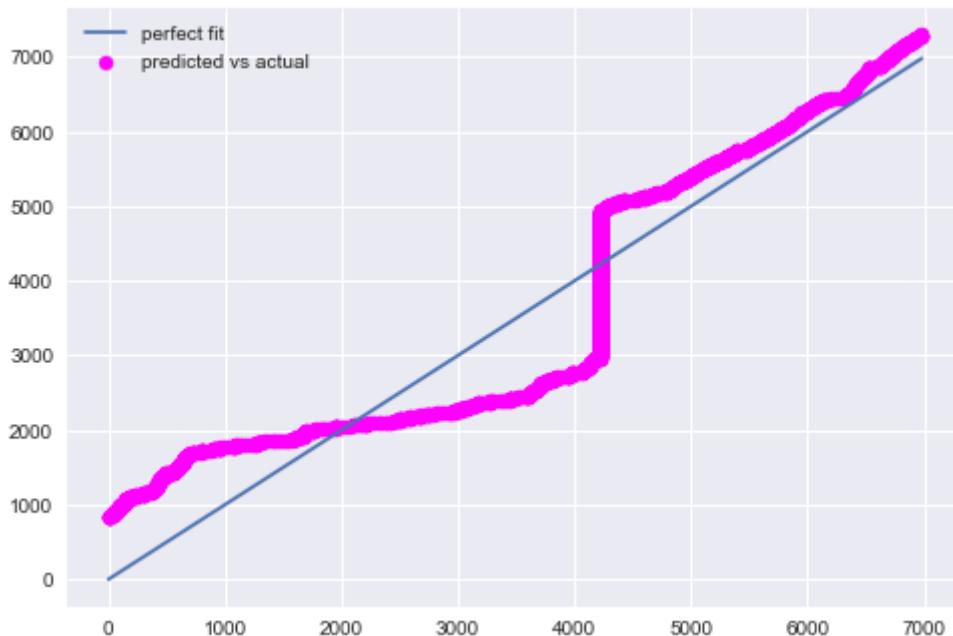
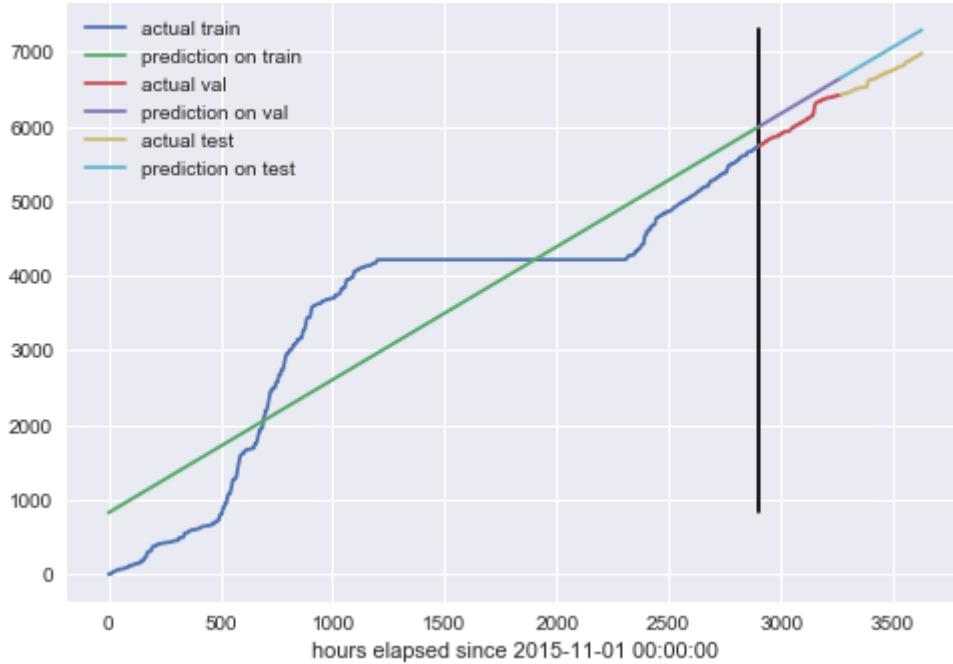
meterid 8156; r2_train: 0.971; r2_val: -35.158; r2_test: -71.424



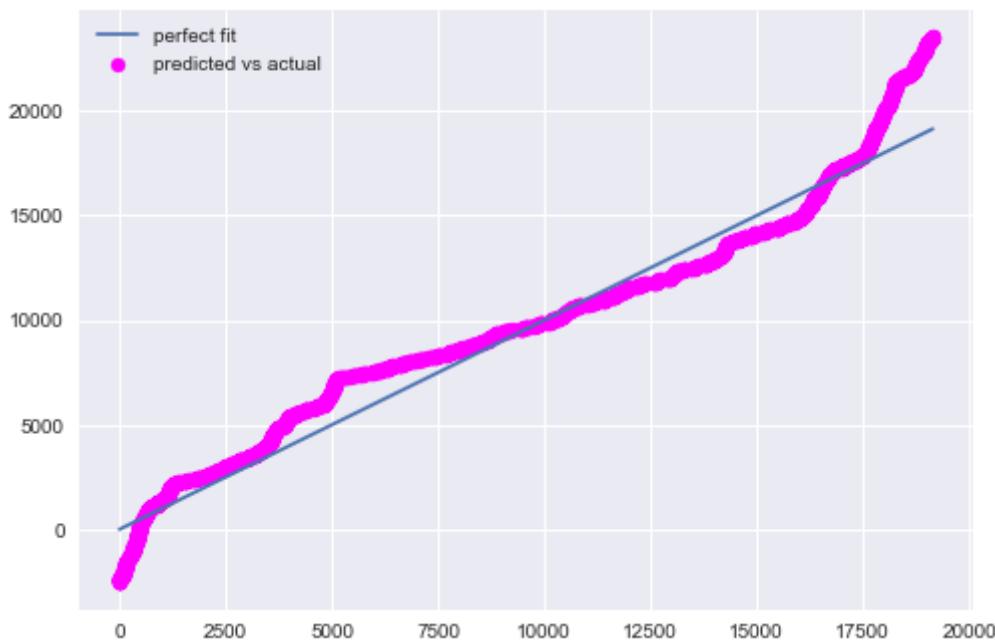
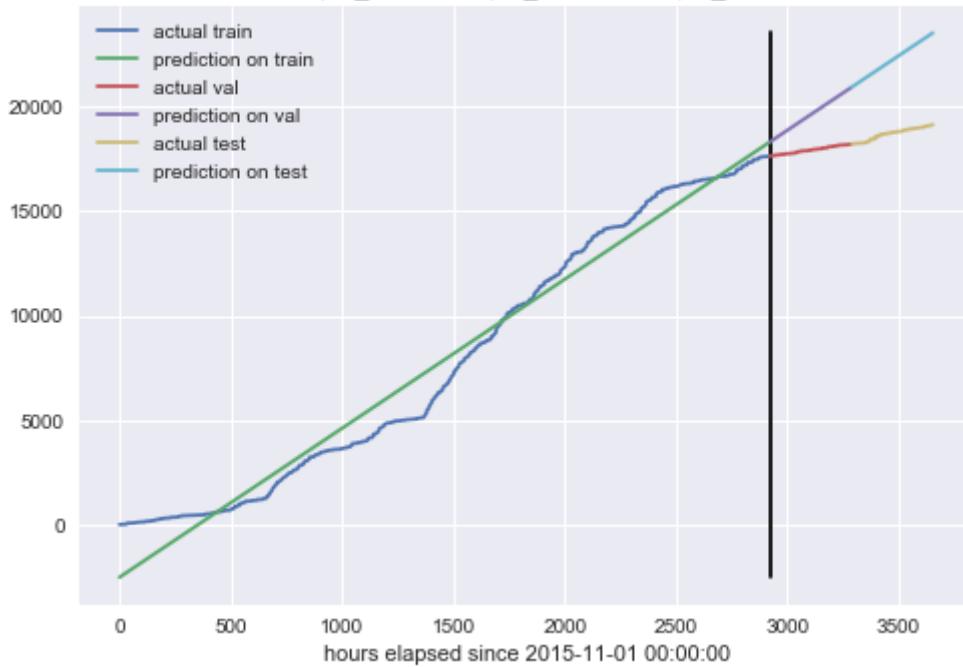
meterid 8386; r2_train: 0.926; r2_val: -7.416; r2_test: -139.653



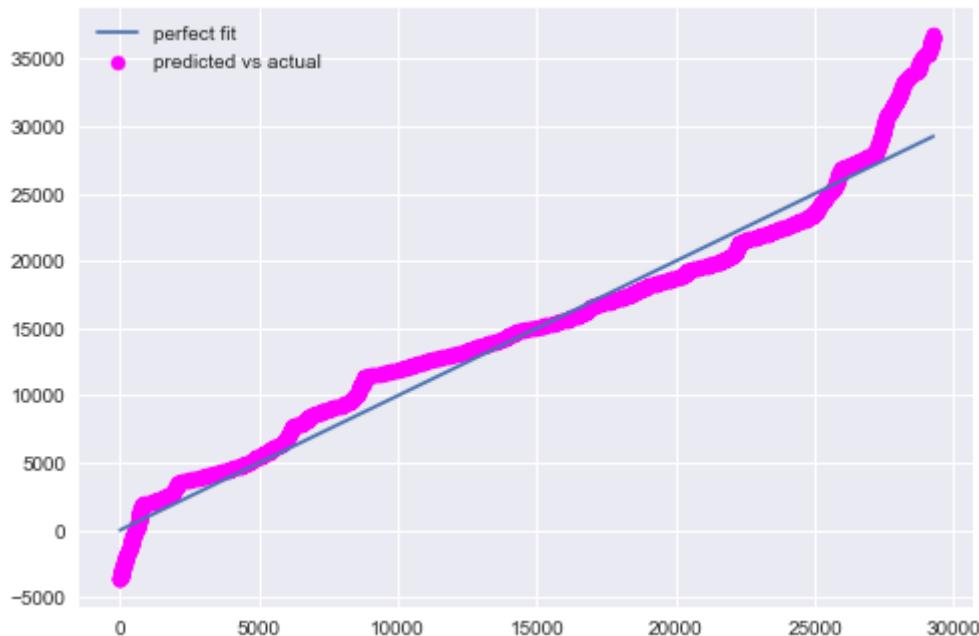
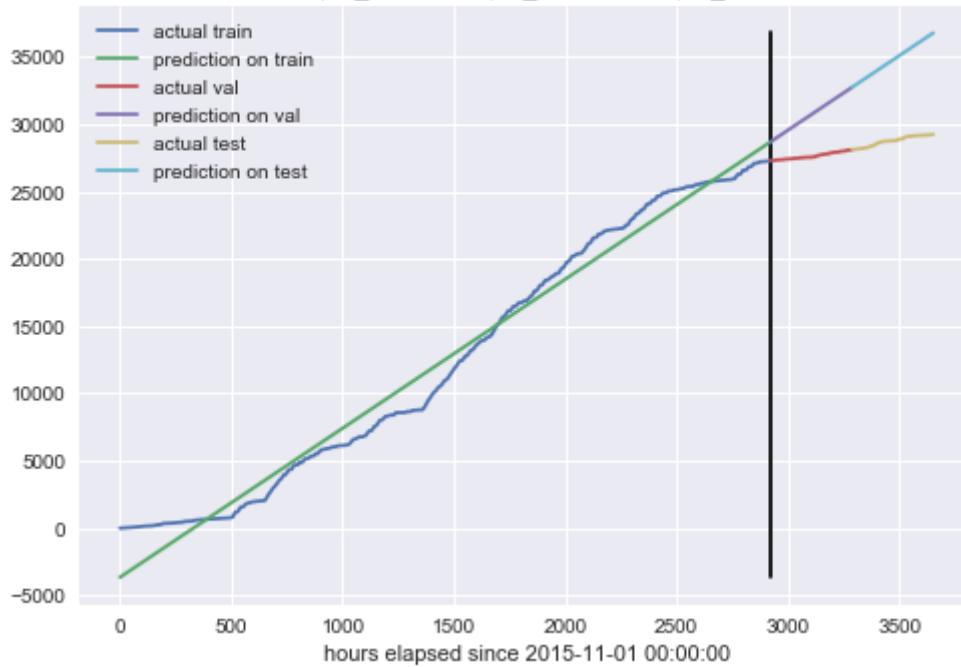
meterid 8467; r2_train: 0.812; r2_val: -0.211; r2_test: -2.097



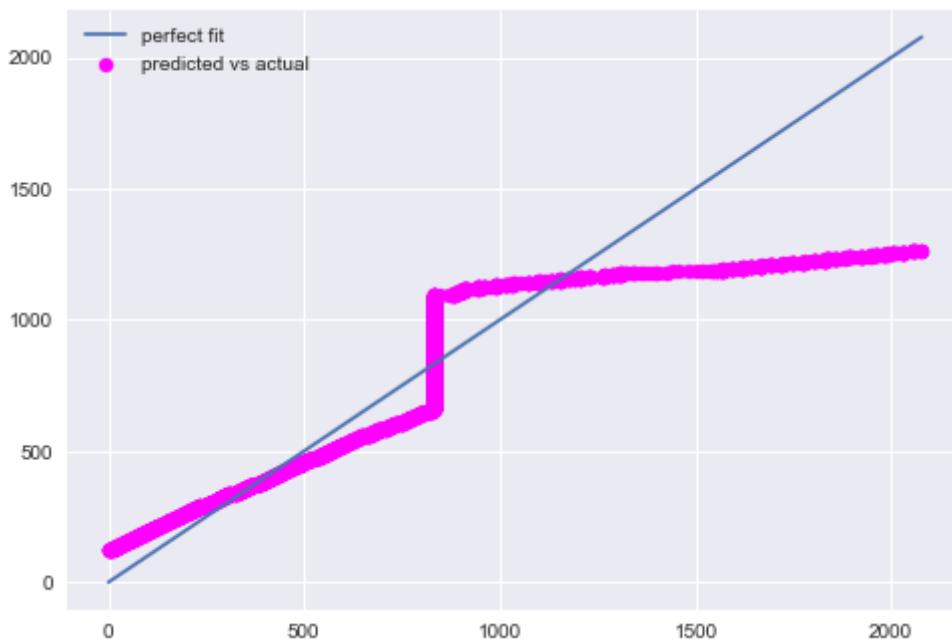
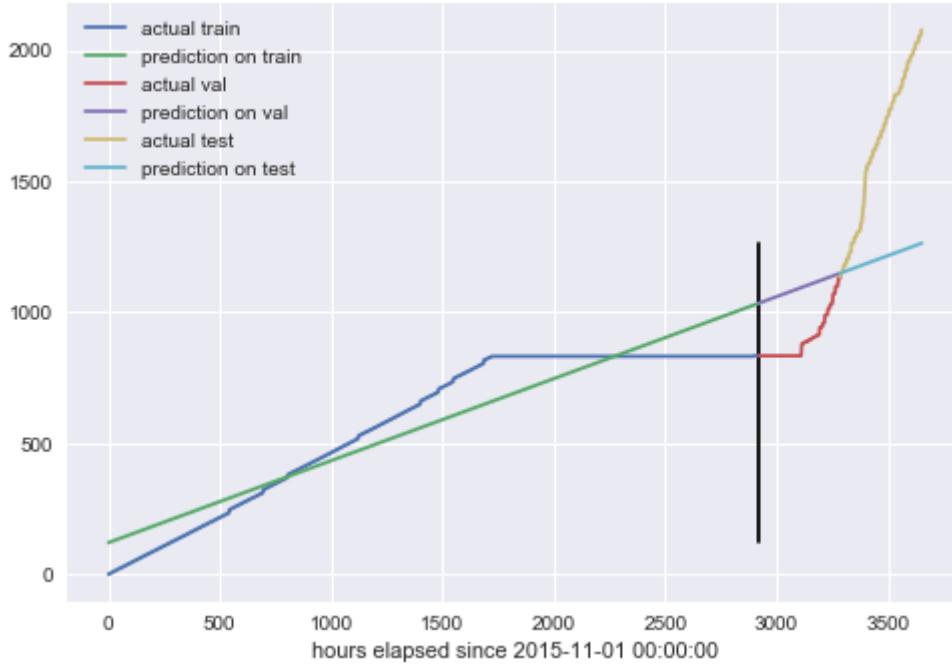
meterid 8829; r2_train: 0.974; r2_val: -105.614; r2_test: -151.027



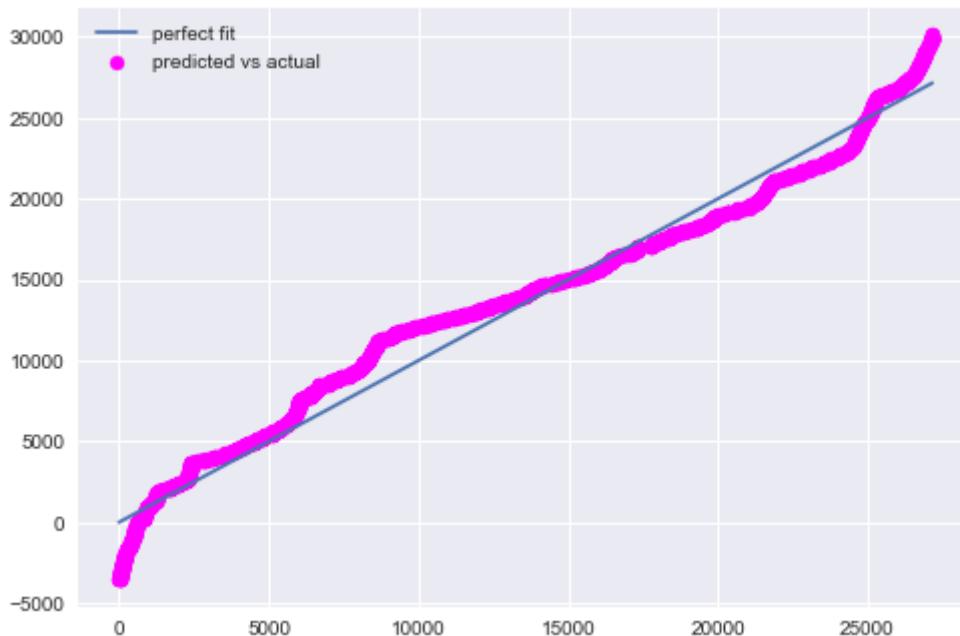
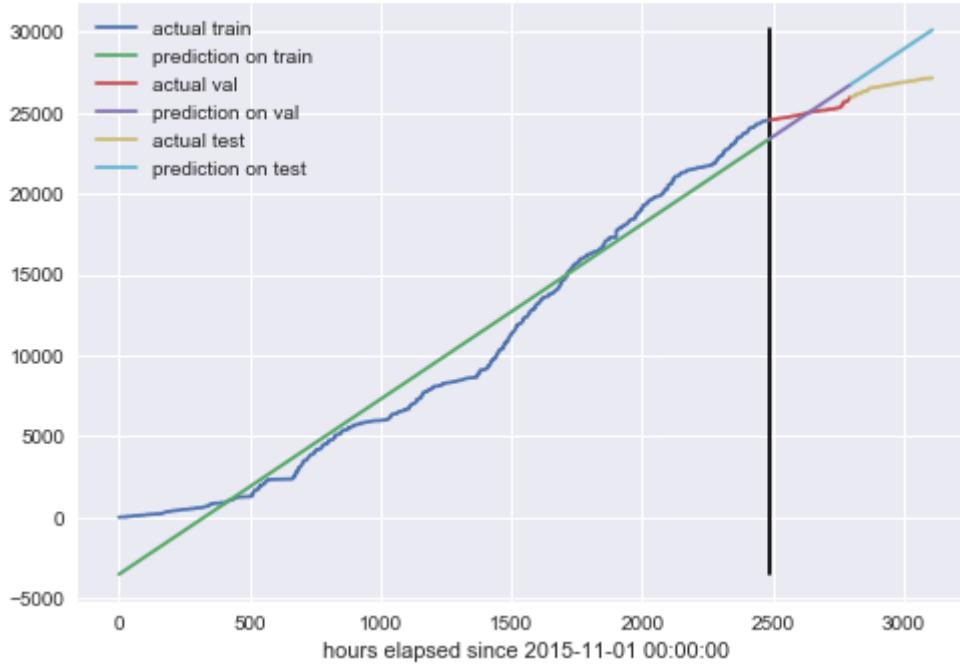
meterid 8890; r2_train: 0.981; r2_val: -169.271; r2_test: -260.258



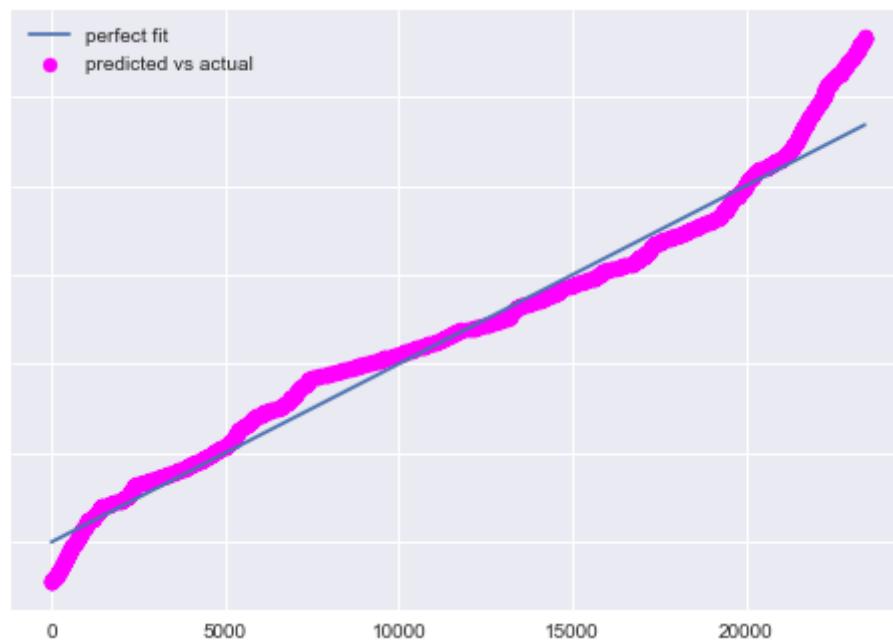
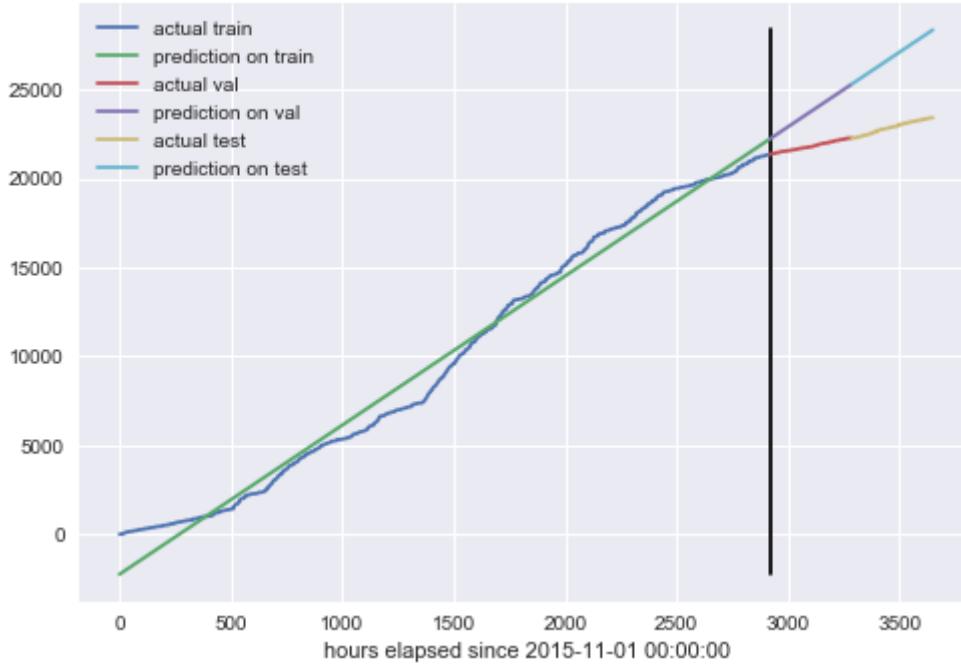
meterid 8967; r2_train: 0.884; r2_val: -4.518; r2_test: -2.104



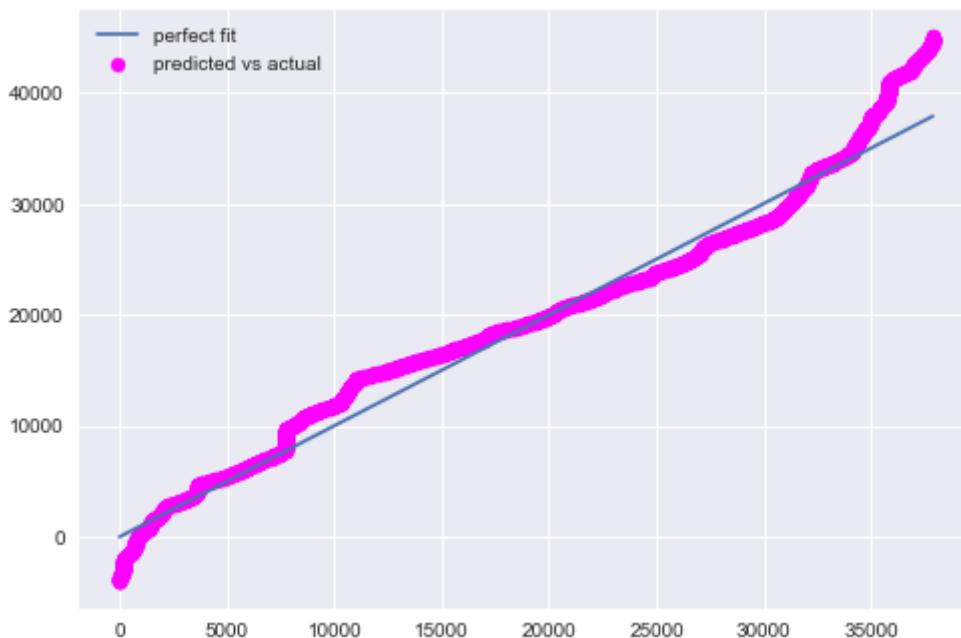
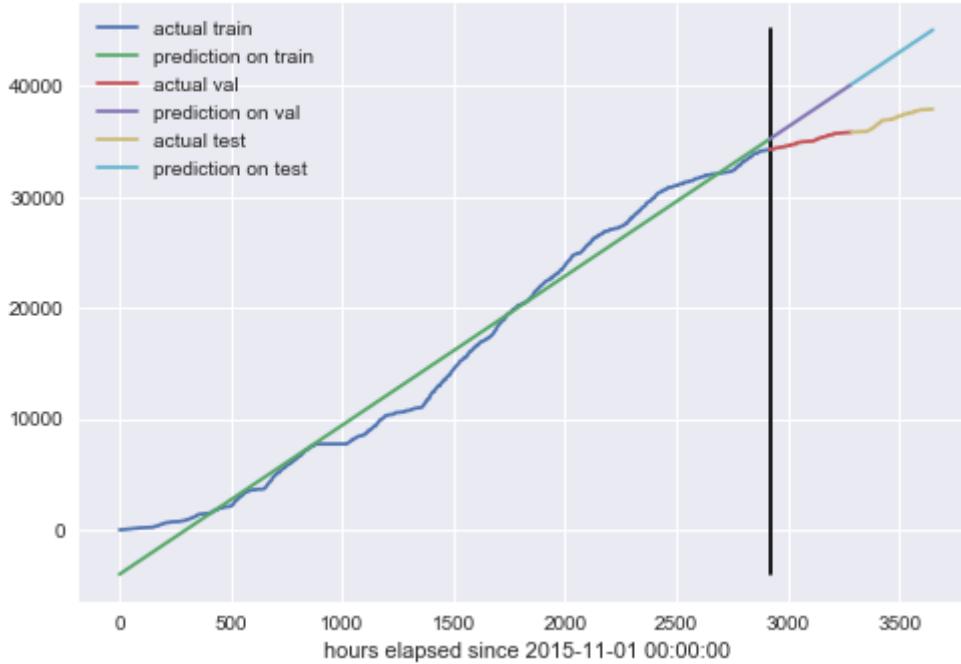
meterid 9052; r2_train: 0.971; r2_val: -2.801; r2_test: -31.686



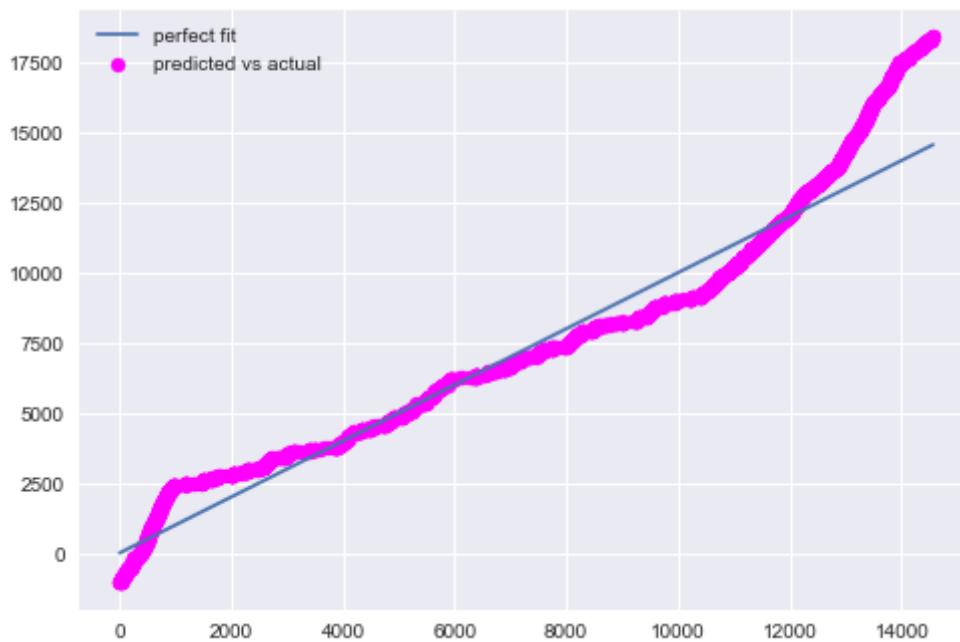
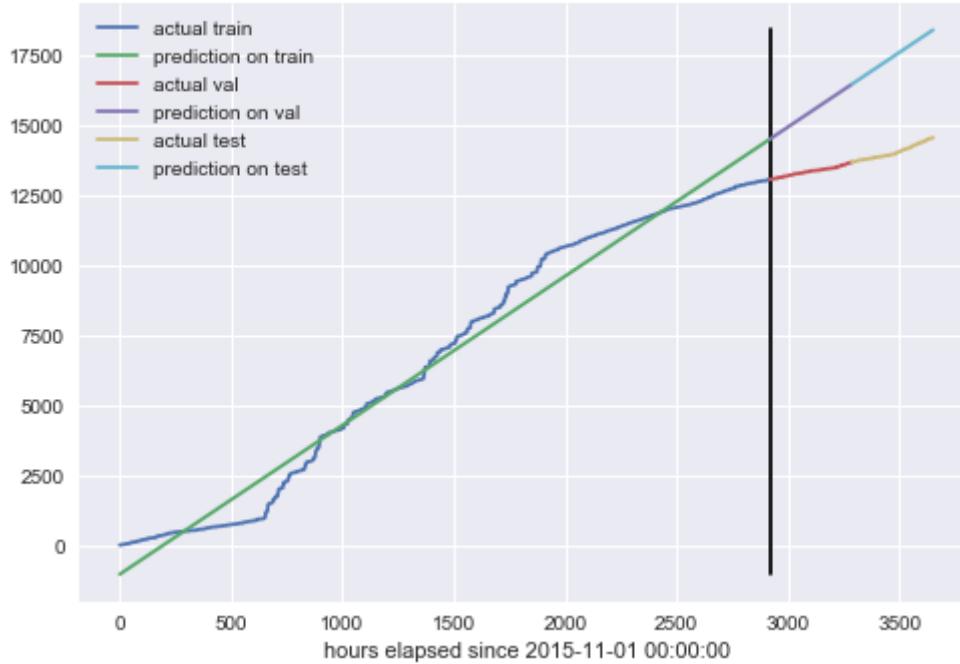
meterid 9121; r2_train: 0.986; r2_val: -55.813; r2_test: -123.987



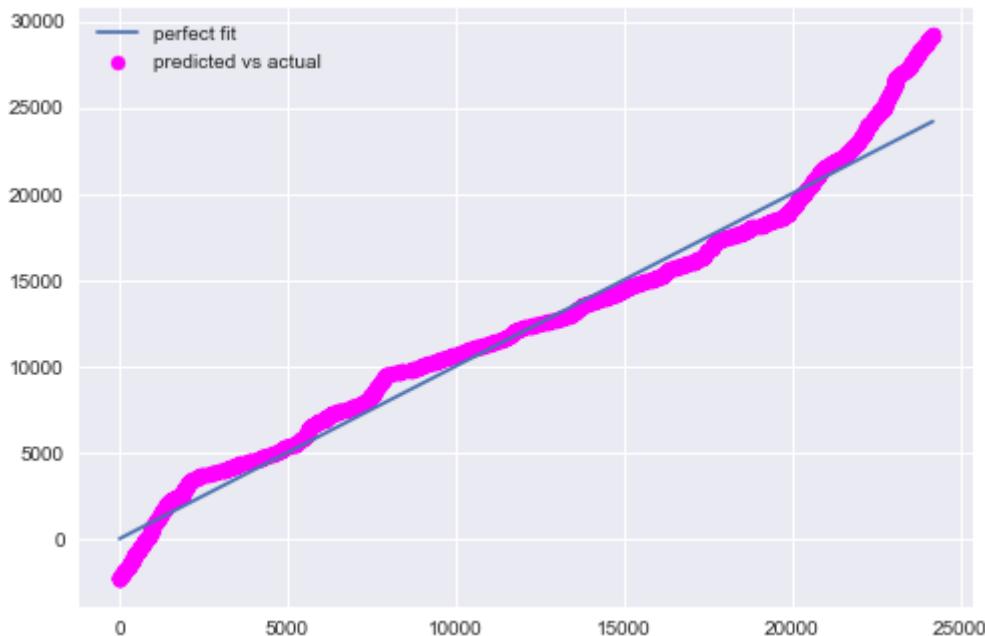
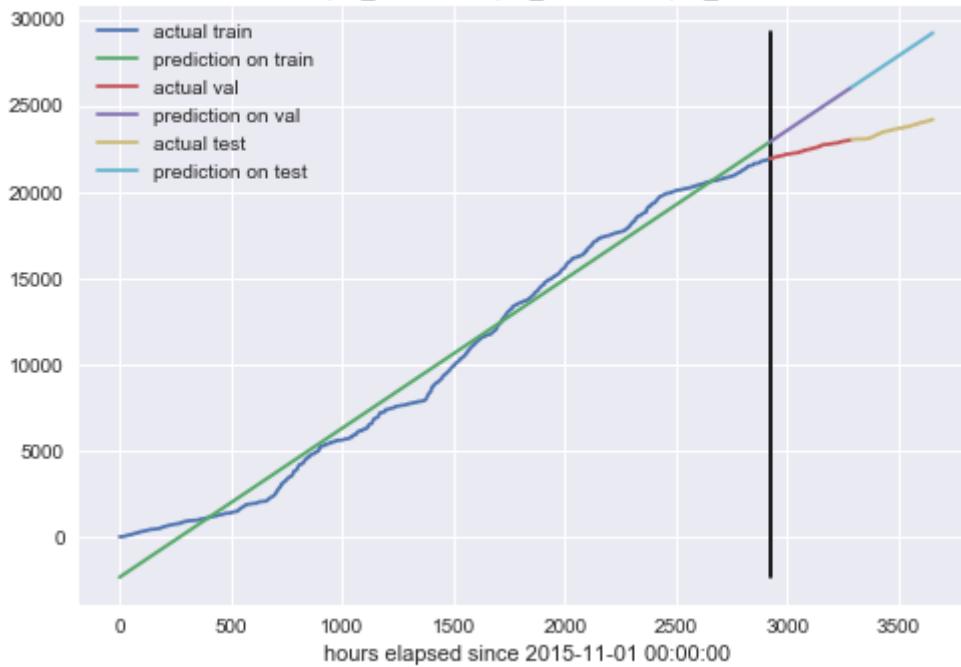
meterid 9134; r2_train: 0.983; r2_val: -29.507; r2_test: -58.263

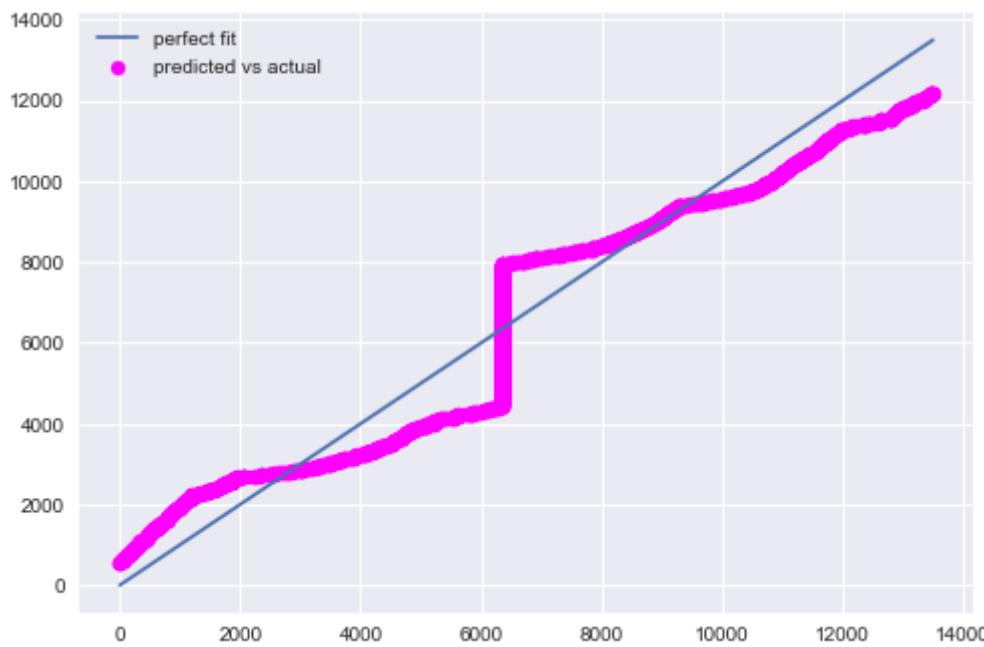
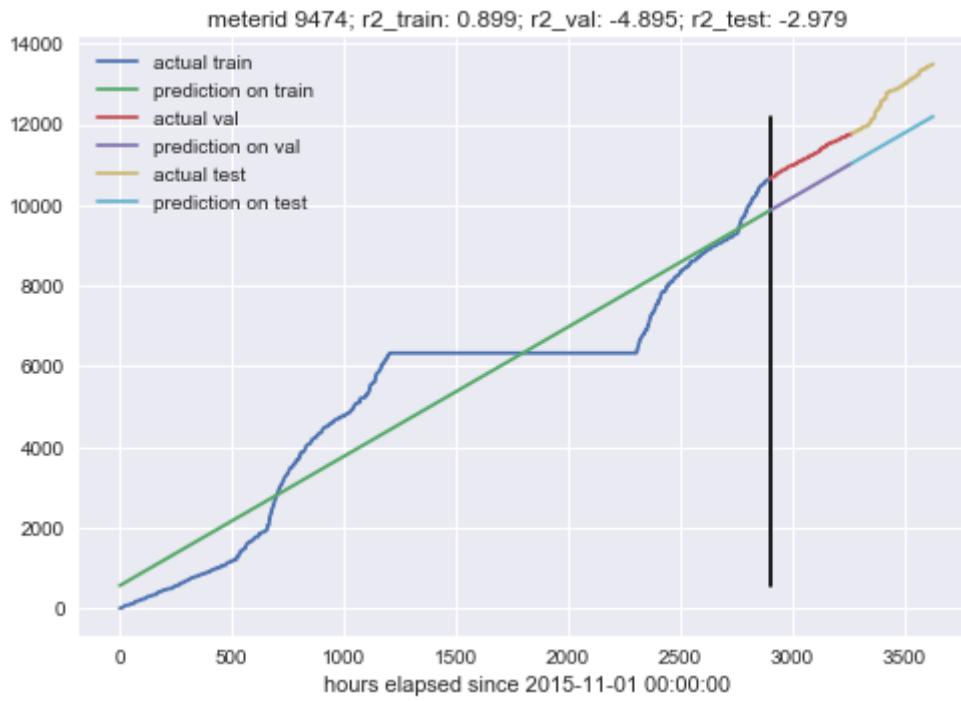


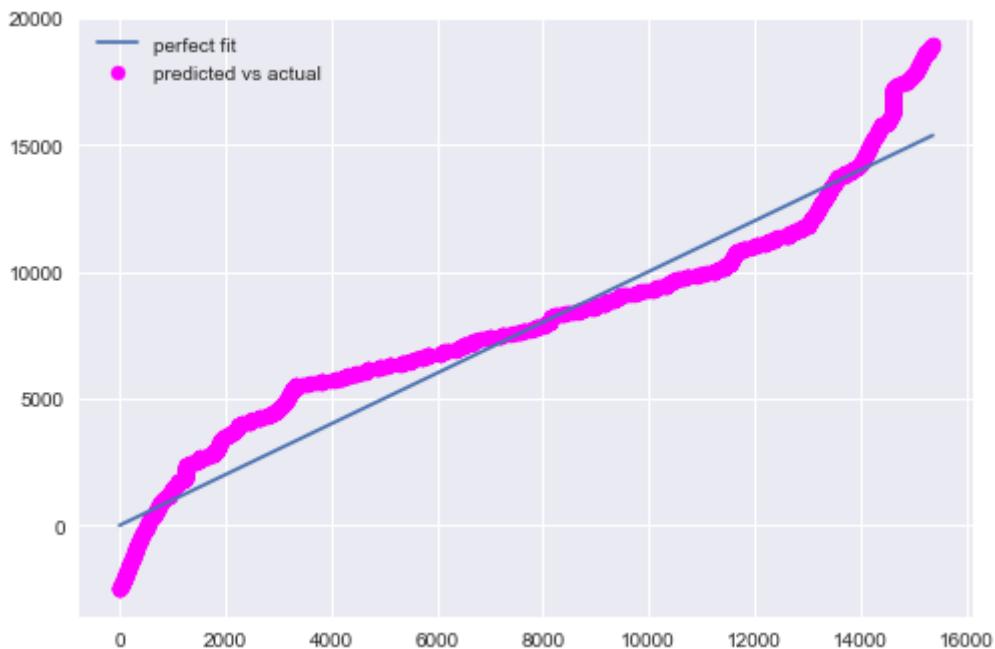
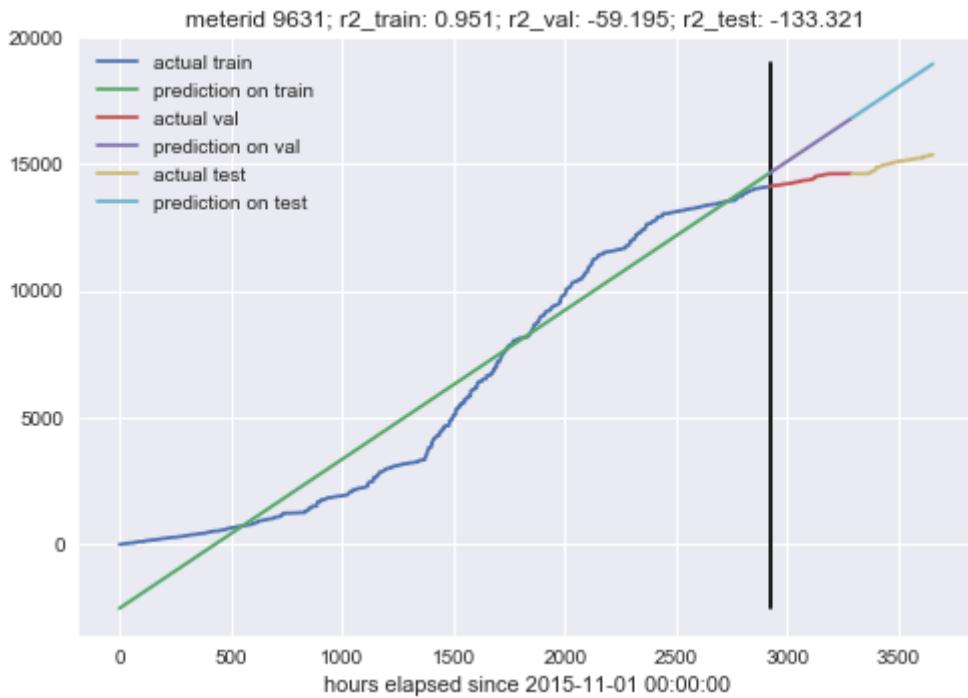
meterid 9278; r2_train: 0.978; r2_val: -178.071; r2_test: -175.017



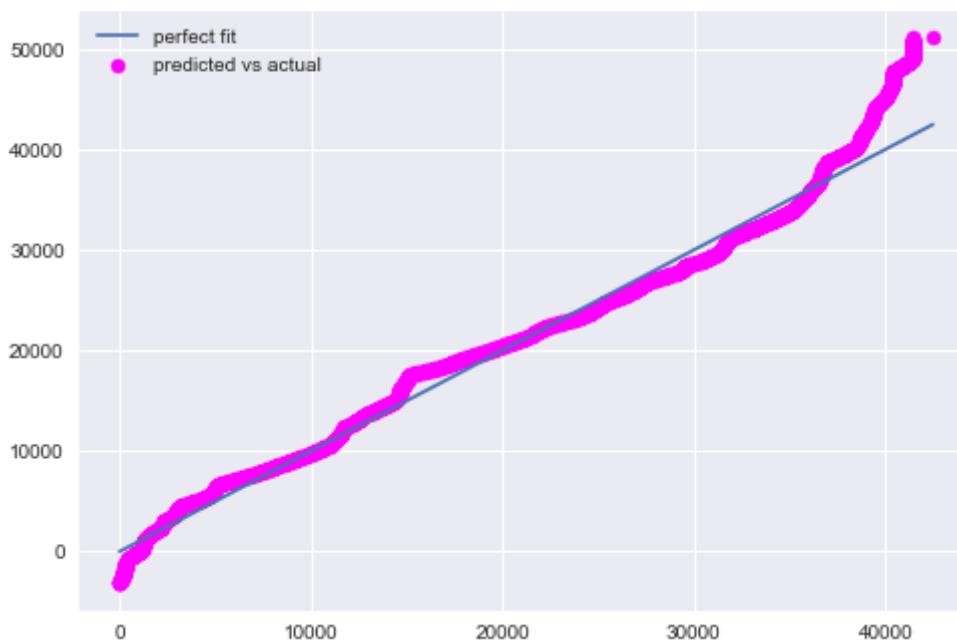
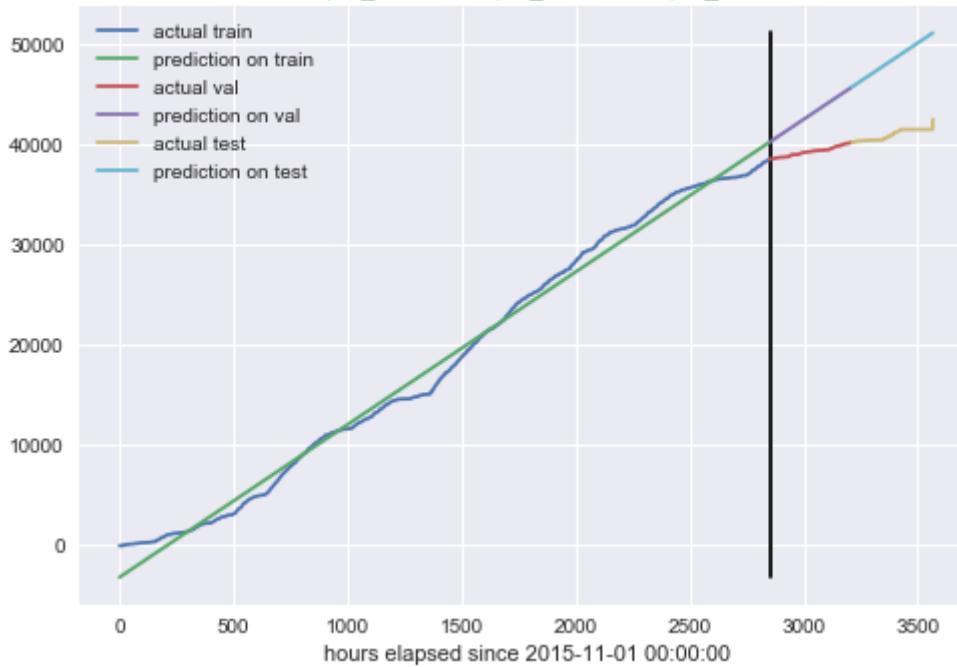
meterid 9295; r2_train: 0.987; r2_val: -40.034; r2_test: -124.598



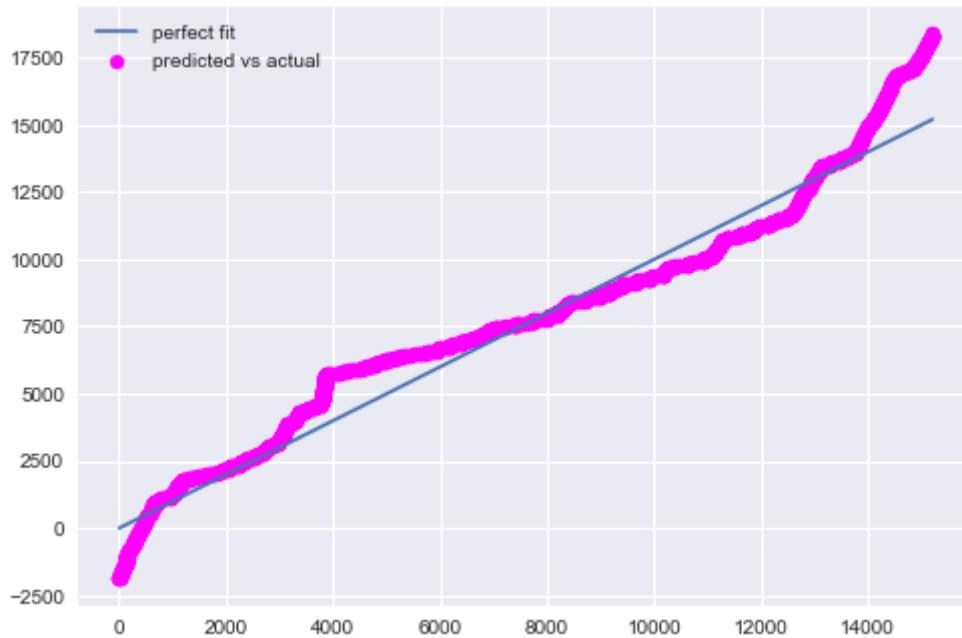
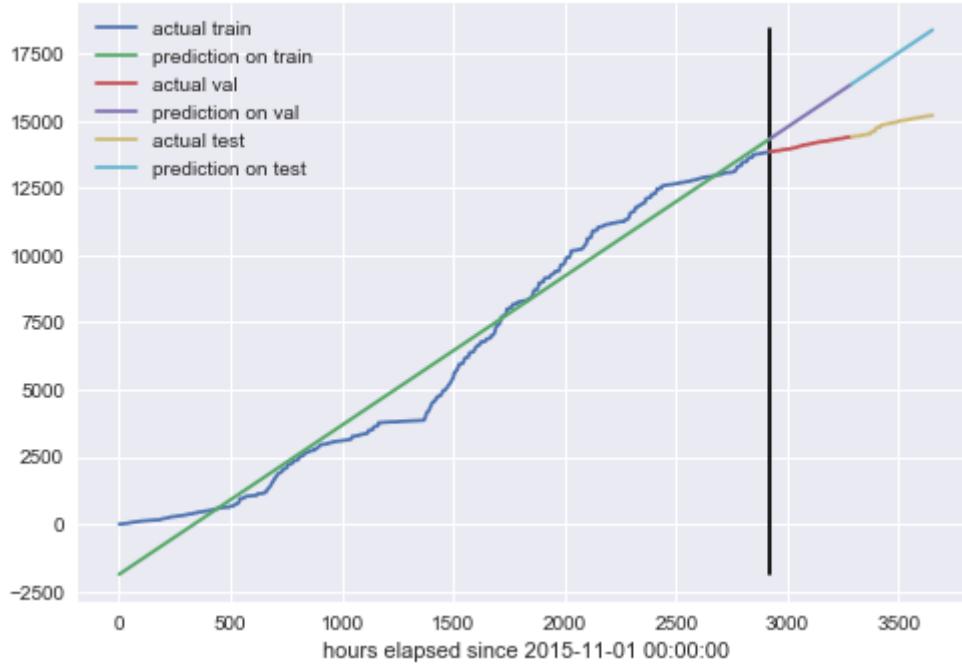


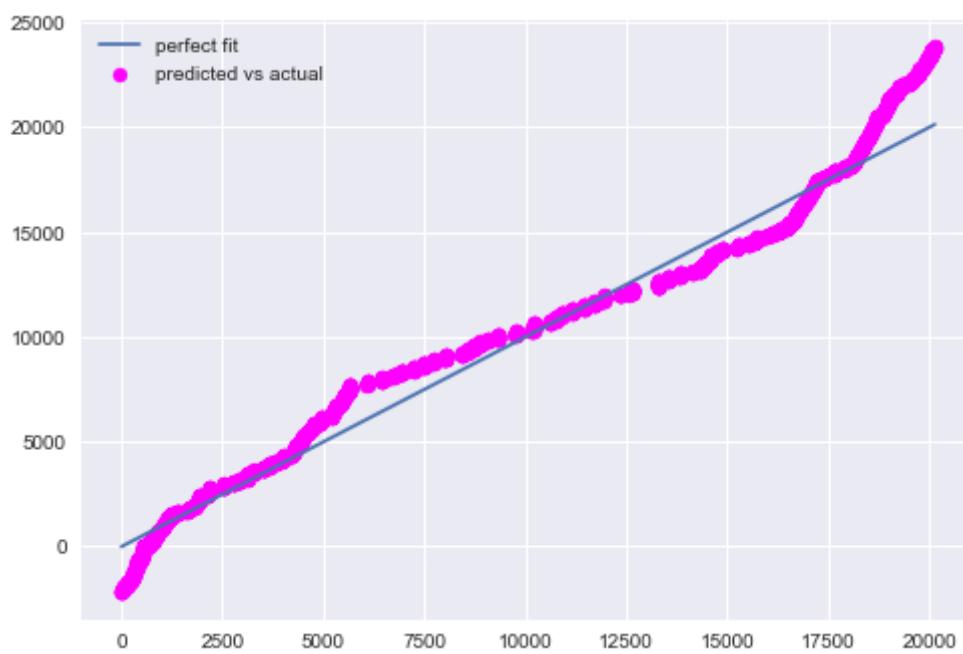
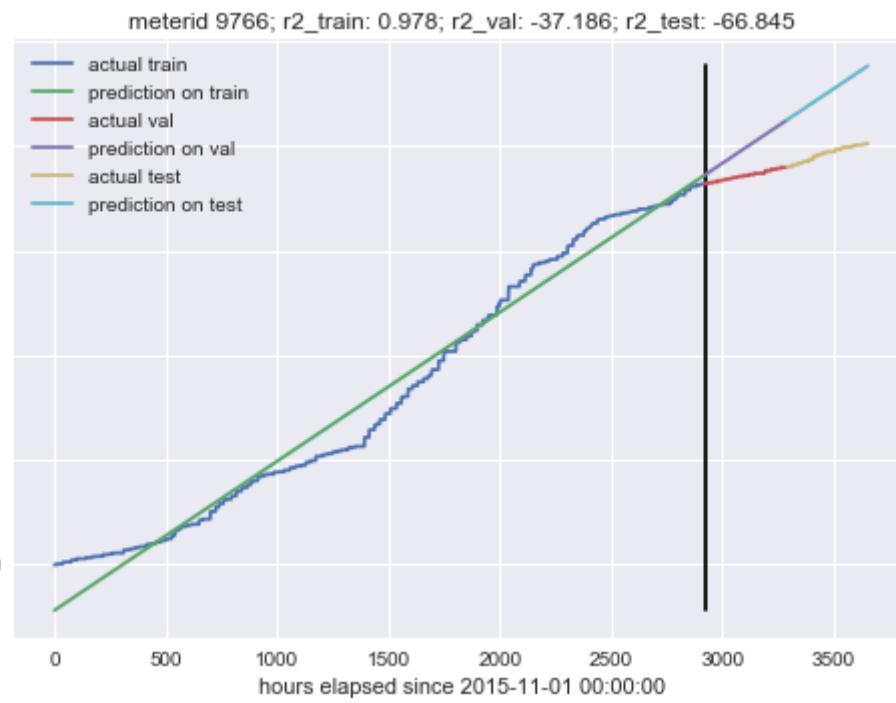


meterid 9639; r2_train: 0.992; r2_val: -66.815; r2_test: -211.364

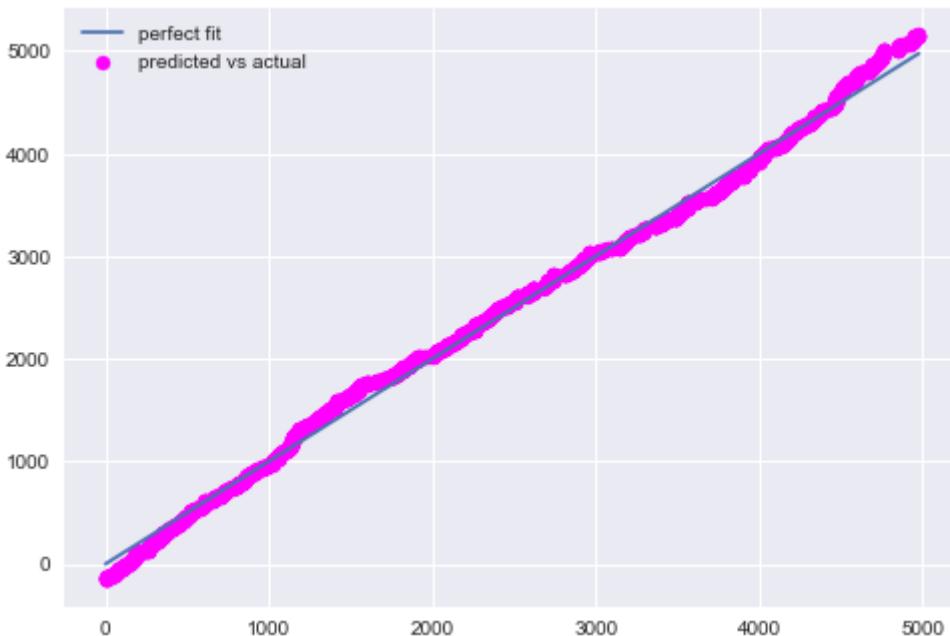
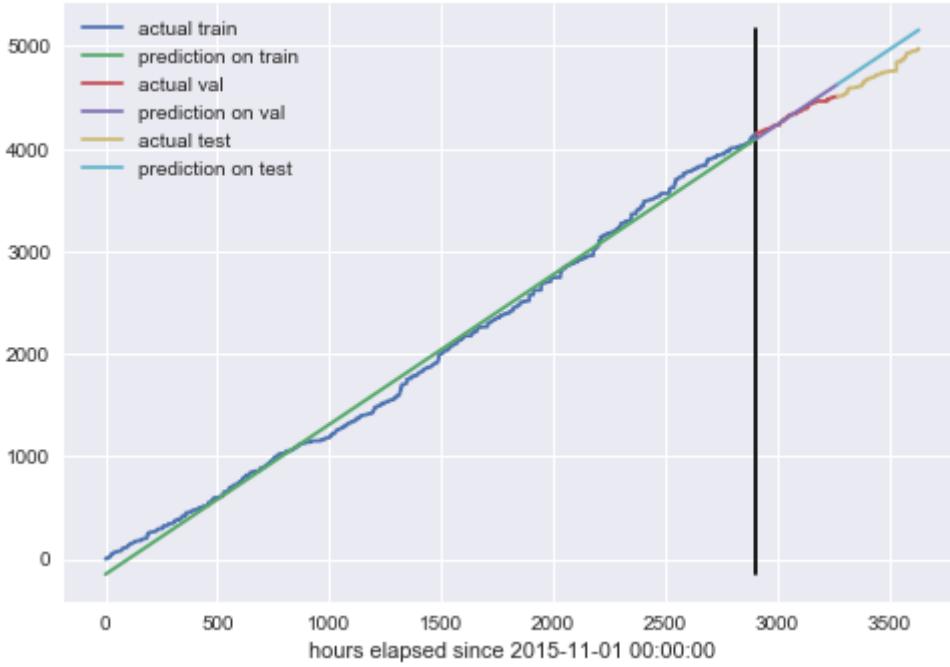


meterid 9729; r2_train: 0.974; r2_val: -55.196; r2_test: -91.587

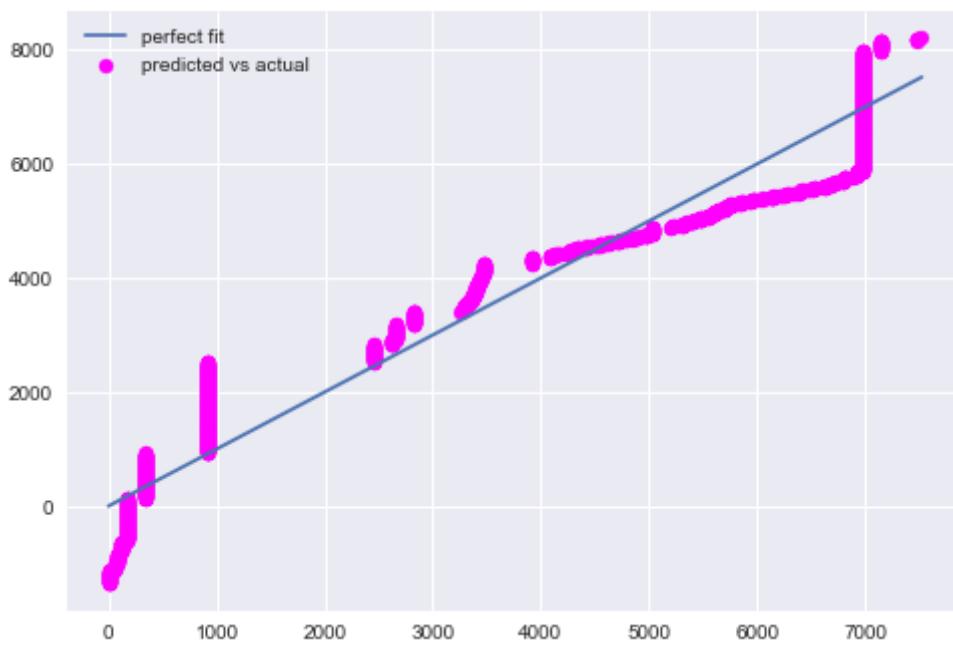
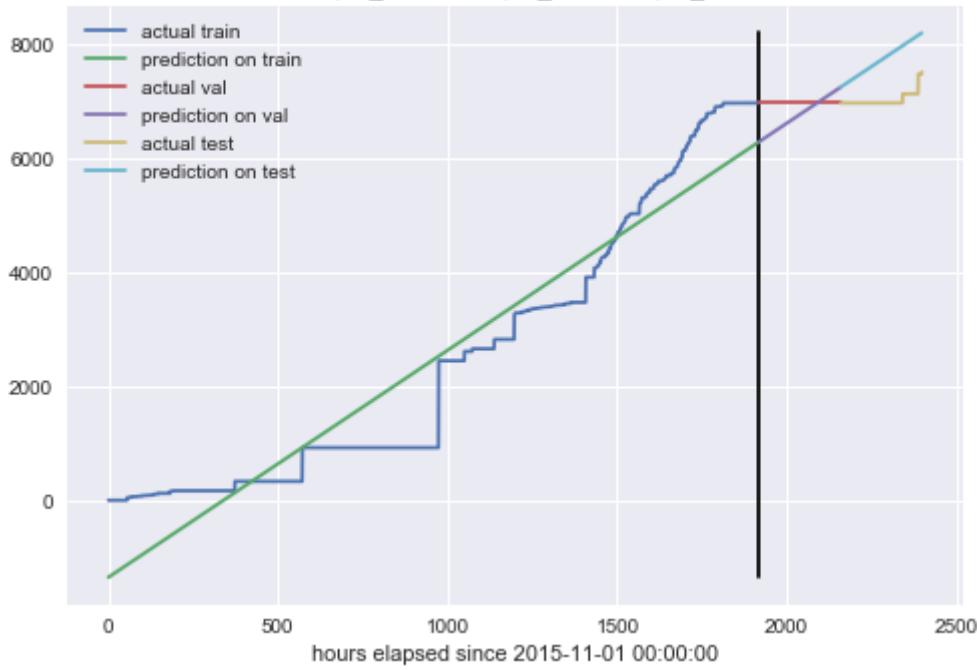




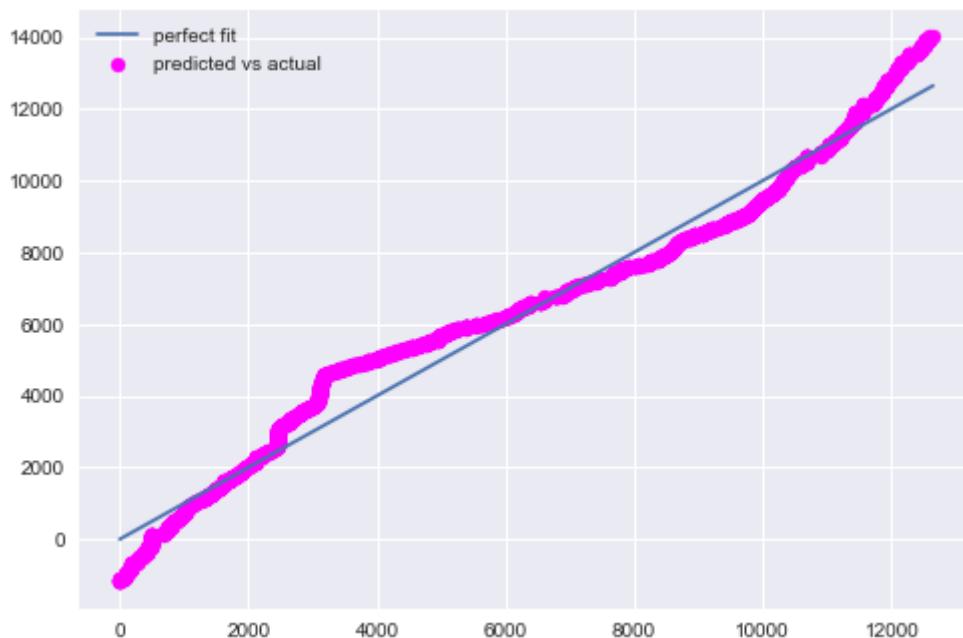
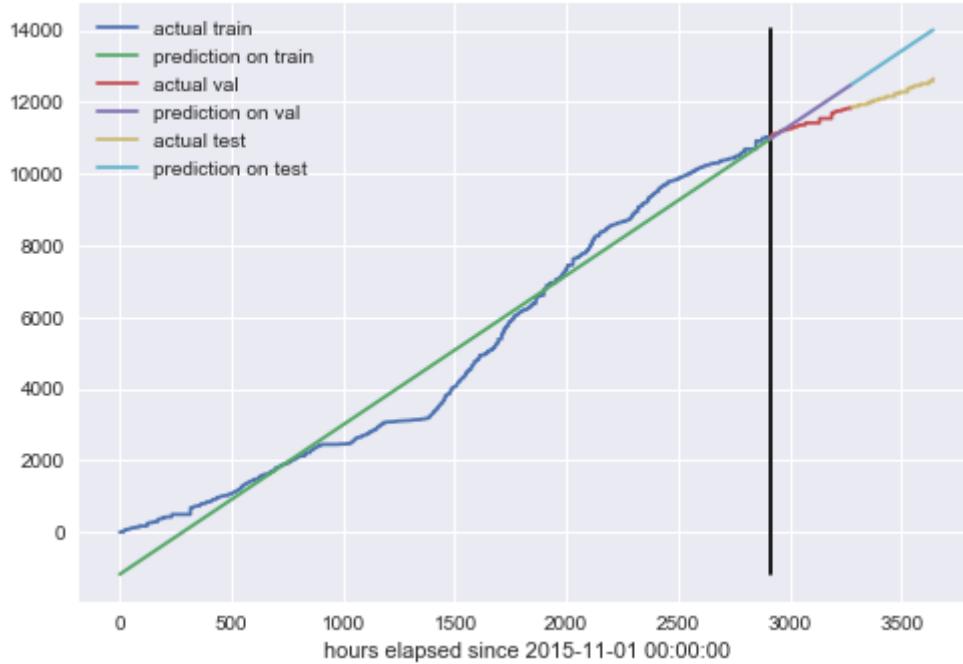
meterid 9849; r2_train: 0.996; r2_val: 0.844; r2_test: -0.540

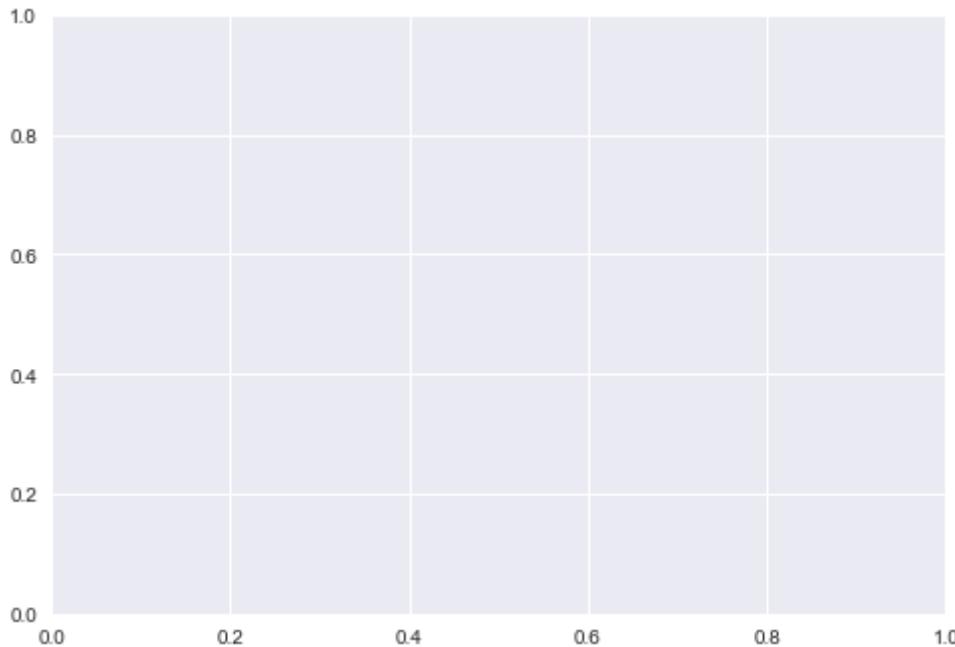


meterid 9956; r2_train: 0.910; r2_val: 0.000; r2_test: -37.961



meterid 9982; r2_train: 0.971; r2_val: -1.548; r2_test: -22.304





In [41]:

```
# plot line indicating train/test demarcation
test_mark = X[valid_start][0]
print(test_mark)
```

2922