dataworth

July 1, 2019

1 PESTPP-GLM

In this notebook, we will run PESTPP-GLM in standard parameter estimation mode and regularization mode. In both cases, we will use the baked-in bayes-linear posterior monte carlo analysis to get posterior forecast PDFs. We will use the prior monte carlo outputs as the prior forecast PDF.

In [1]:
```
%matplotlib inline
import os
import shutil
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import matplotlib as mpl
plt.rcParams['font.size']=12
import flopy
import pyemu
```

```
flopy is installed in /Users/jeremyw/Dev/gw1876/activities_csiro/notebooks/flopy
```

In [2]:
```
m_d = "master_glm"
```

In [3]:
```
pst = pyemu.Pst(os.path.join(m_d,"freyberg_pp.pst"))
pst.write_par_summary_table(filename="none")
```

Out[3]:

	type	transform	count	initial value	upper bound	\
cn_strt8	cn_strt8	log	1	0	0.0211893	
cn_rech5	cn_rech5	log	1	0	0.0413927	
cn_vka6	cn_vka6	log	1	0	1	
gr_vka4	gr_vka4	fixed	705	1	10	
cn_hk6	cn_hk6	log	1	0	1	
welflux_k02	welflux_k02	log	6	0	1	
gr_ss3	gr_ss3	fixed	705	1	10	
cn_ss8	cn_ss8	log	1	0	1	
pp_strt1	pp_strt1	fixed	32	1	1.05	
gr_sy4	gr_sy4	fixed	705	1	1.75	
pp_rech0	pp_rech0	log	32	0	0.0413927	
gr_prsity4	gr_prsity4	fixed	705	1	1.5	

gr_rech2	gr_rech2	fixed	705	1	1.1
flow	flow	log	1	0	0.09691
gr_prsity3	gr_prsity3	fixed	705	1	1.5
cn_hk8	cn_hk8	log	1	0	1
pp_strt2	pp_strt2	fixed	32	1	1.05
pp_sy0	pp_sy0	log	32	0	0.243038
gr_rech3	gr_rech3	fixed	705	1	1.1
cn_ss7	cn_ss7	log	1	0	1
gr_strt5	gr_strt5	fixed	705	1	1.05
pp_prsity0	pp_prsity0	log	32	0	0.176091
gr_ss5	gr_ss5	fixed	705	1	10
pp_hk0	pp_hk0	log	32	0	1
gr_strt4	gr_strt4	fixed	705	1	1.05
gr_sy3	gr_sy3	fixed	705	1	1.75
gr_ss4	gr_ss4	fixed	705	1	10
pp_prsity1	pp_prsity1	log	32	0	0.176091
cn_sy7	cn_sy7	log	1	0	0.243038
drncond_k00	drncond_k00	log	10	0	1
...	...	...	...	...	...
pp_hk1	pp_hk1	log	32	0	1
cn_prsity8	cn_prsity8	log	1	0	0.176091
pp_ss1	pp_ss1	log	32	0	1
cn_rech4	cn_rech4	log	1	0	0.0413927
pp_vka0	pp_vka0	fixed	32	1	10
cn_ss6	cn_ss6	log	1	0	1
cn_vka8	cn_vka8	log	1	0	1
gr_vka3	gr_vka3	fixed	705	1	10
cn_sy6	cn_sy6	log	1	0	0.243038
gr_prsity5	gr_prsity5	fixed	705	1	1.5
strk	strk	log	40	0	2
pp_sy1	pp_sy1	log	32	0	0.243038
pp_strt0	pp_strt0	fixed	32	1	1.05
pp_prsity2	pp_prsity2	log	32	0	0.176091
gr_hk4	gr_hk4	fixed	705	1	10
cn_strt6	cn_strt6	log	1	0	0.0211893
cn_vka7	cn_vka7	log	1	0	1
pp_vka1	pp_vka1	log	32	0	1
cn_hk7	cn_hk7	log	1	0	1
cn_strt7	cn_strt7	log	1	0	0.0211893
gr_sy5	gr_sy5	fixed	705	1	1.75
pp_ss2	pp_ss2	log	32	0	1
cn_prsity6	cn_prsity6	log	1	0	0.176091
welflux	welflux	log	2	0	1
gr_strt3	gr_strt3	fixed	705	1	1.05
pp_sy2	pp_sy2	log	32	0	0.243038
pp_vka2	pp_vka2	fixed	32	1	10
gr_hk3	gr_hk3	fixed	705	1	10
cn_sy8	cn_sy8	log	1	0	0.243038

pp_hk2	pp_hk2	log	32	0	1

	lower bound	standard deviation
cn_strt8	-0.0222764	0.0108664
cn_rech5	-0.0457575	0.0217875
cn_vka6	-1	0.5
gr_vka4	0.1	2.475
cn_hk6	-1	0.5
welflux_k02	-1	0.5
gr_ss3	0.1	2.475
cn_ss8	-1	0.5
pp_strt1	0.95	0.025
gr_sy4	0.25	0.375
pp_rech0	-0.0457575	0.0217875
gr_prsity4	0.5	0.25
gr_rech2	0.9	0.05
flow	-0.124939	0.0554622
gr_prsity3	0.5	0.25
cn_hk8	-1	0.5
pp_strt2	0.95	0.025
pp_sy0	-0.60206	0.211275
gr_rech3	0.9	0.05
cn_ss7	-1	0.5
gr_strt5	0.95	0.025
pp_prsity0	-0.30103	0.11928
gr_ss5	0.1	2.475
pp_hk0	-1	0.5
gr_strt4	0.95	0.025
gr_sy3	0.25	0.375
gr_ss4	0.1	2.475
pp_prsity1	-0.30103	0.11928
cn_sy7	-0.60206	0.211275
drncond_k00	-1	0.5
...	...	...
pp_hk1	-1	0.5
cn_prsity8	-0.30103	0.11928
pp_ss1	-1	0.5
cn_rech4	-0.0457575	0.0217875
pp_vka0	0.1	2.475
cn_ss6	-1	0.5
cn_vka8	-1	0.5
gr_vka3	0.1	2.475
cn_sy6	-0.60206	0.211275
gr_prsity5	0.5	0.25
strk	-2	1
pp_sy1	-0.60206	0.211275
pp_strt0	0.95	0.025
pp_prsity2	-0.30103	0.11928

```
gr_hk4              0.1             2.475
cn_strt6     -0.0222764          0.0108664
cn_vka7              -1               0.5
pp_vka1              -1               0.5
cn_hk7               -1               0.5
cn_strt7     -0.0222764          0.0108664
gr_sy5             0.25             0.375
pp_ss2               -1               0.5
cn_prsity6     -0.30103           0.11928
welflux              -1               0.5
gr_strt3           0.95             0.025
pp_sy2          -0.60206          0.211275
pp_vka2             0.1             2.475
gr_hk3              0.1             2.475
cn_sy8          -0.60206          0.211275
pp_hk2               -1               0.5

[65 rows x 7 columns]
```

```
In [4]: cov = pyemu.Cov.from_binary(os.path.join(m_d,"prior_cov.jcb")).to_dataframe()
        cov = cov.loc[pst.adj_par_names,pst.adj_par_names]
        cov = pyemu.Cov.from_dataframe(cov)
```

```
new binary format detected...
```

```
In [5]: sc = pyemu.Schur(jco=os.path.join(m_d,"freyberg_pp.jcb"),parcov=cov)
```

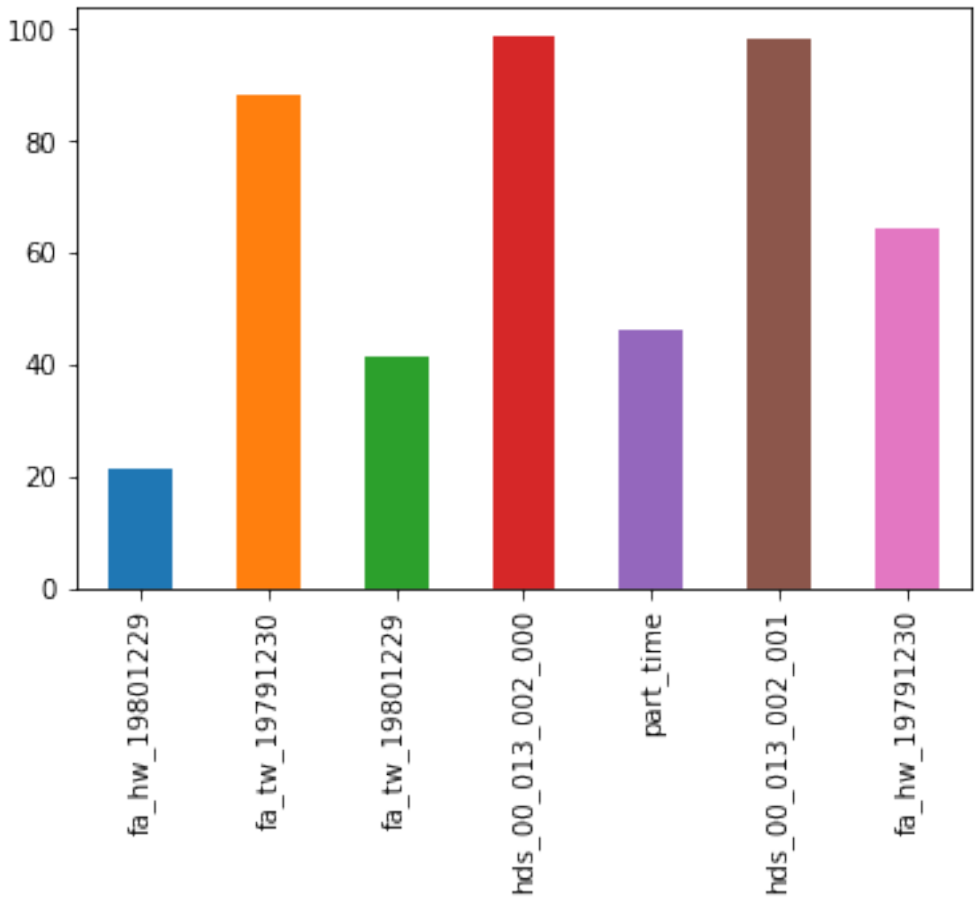
```
In [6]: df = sc.get_forecast_summary()
        df
```

Out[6]:

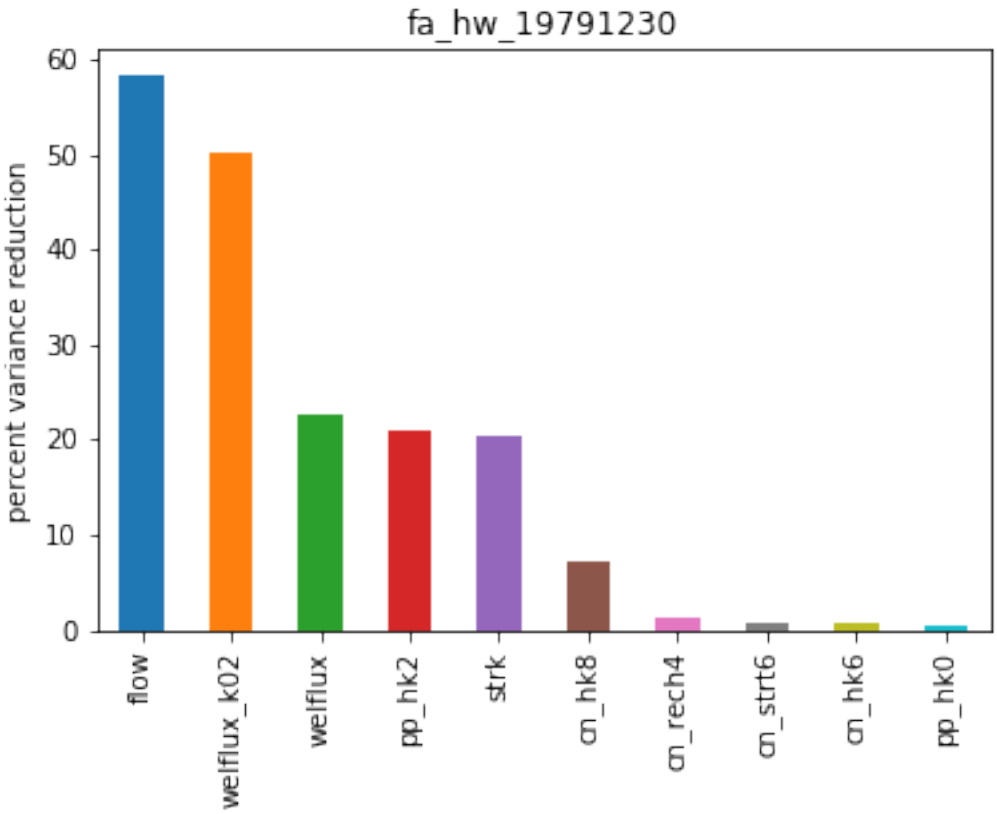
	percent_reduction	post_var	prior_var
part_status	NaN	0.000000	0.000000
fa_hw_19801229	21.344340	273159.074817	347284.702910
fa_tw_19791230	88.128060	20437.305568	172147.988428
fa_tw_19801229	41.327445	218490.578206	372389.747704
hds_00_013_002_000	98.824948	0.089684	7.632373
part_time	46.148749	113237.855892	210278.969728
hds_00_013_002_001	98.009634	0.168864	8.484073
fa_hw_19791230	64.350208	45999.044550	129030.329448

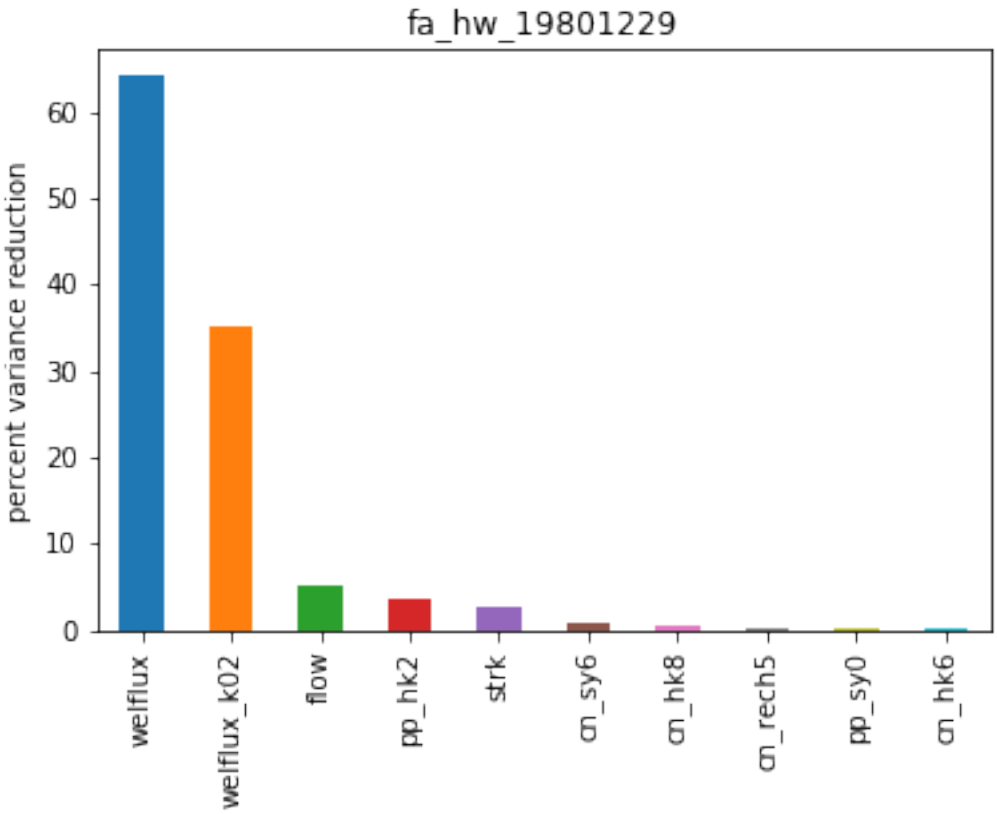
```
In [7]: df = df.loc[:,"percent_reduction"].dropna()
        df.plot(kind="bar")
```

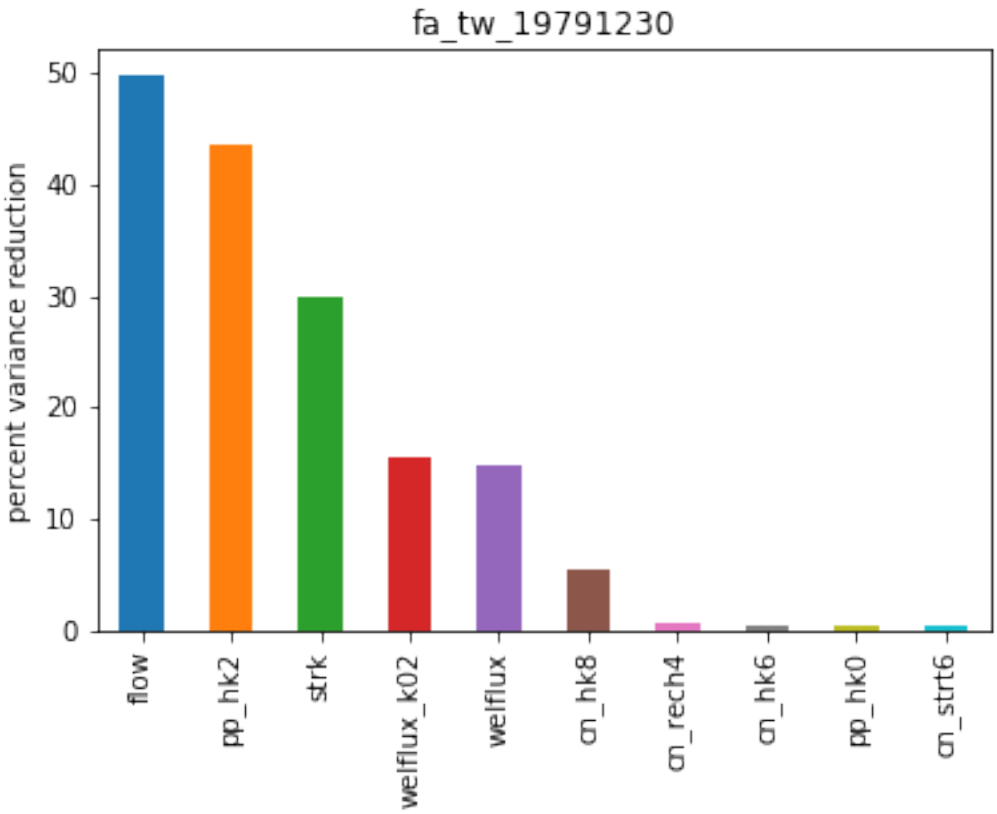
```
Out[7]: <matplotlib.axes._subplots.AxesSubplot at 0x10f4bdc18>
```

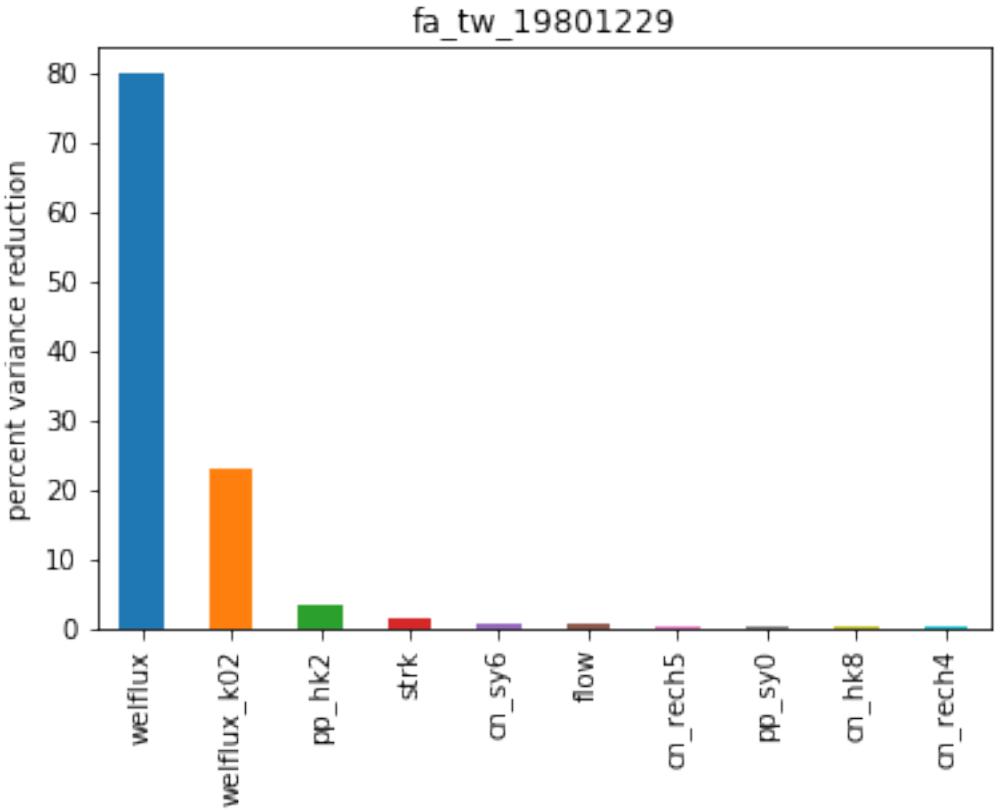


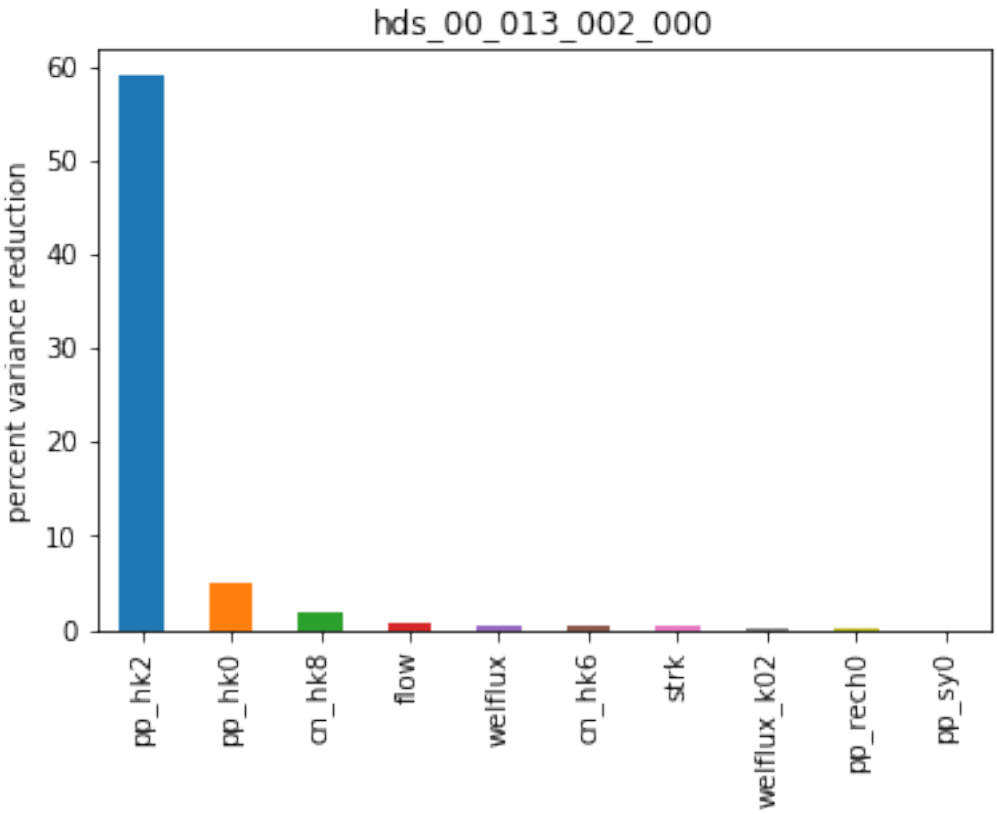

```
In [8]: df = sc.get_par_group_contribution()#.sort_values(by="reduction").iloc[:10].plot(kind=

In [9]: base = df.loc["base",:]
        df = 100.0 * (base - df) / base

In [10]: for forecast in df.columns:
             fore_df = df.loc[:,forecast].copy()
             fore_df.sort_values(inplace=True, ascending=False)
             ax = fore_df.iloc[:10].plot(kind="bar")
             ax.set_title(forecast)
             ax.set_ylabel("percent variance reduction")
             plt.show()
```

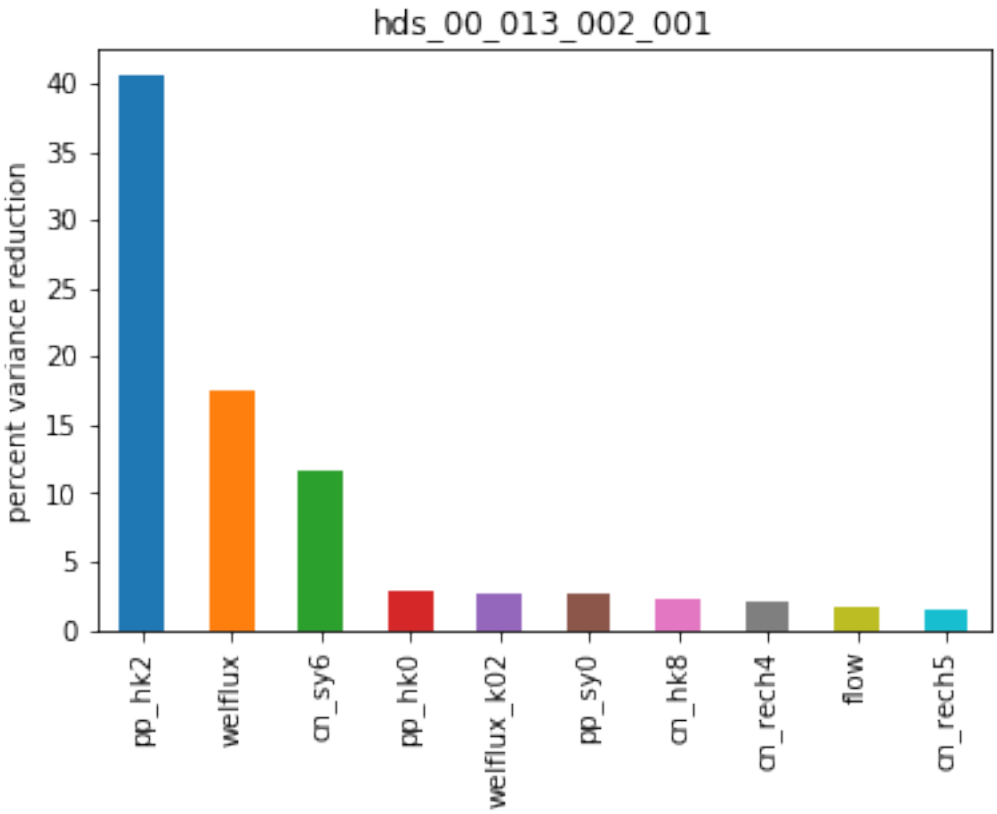

fa_hw_19791230
percent variance reduction
60
50
40
30
20
10
0
flow
welflux_k02
welflux
pp_hk2
strk
cn_hk8
cn_rech4
cn_strt6
cn_hk6
pp_hk0

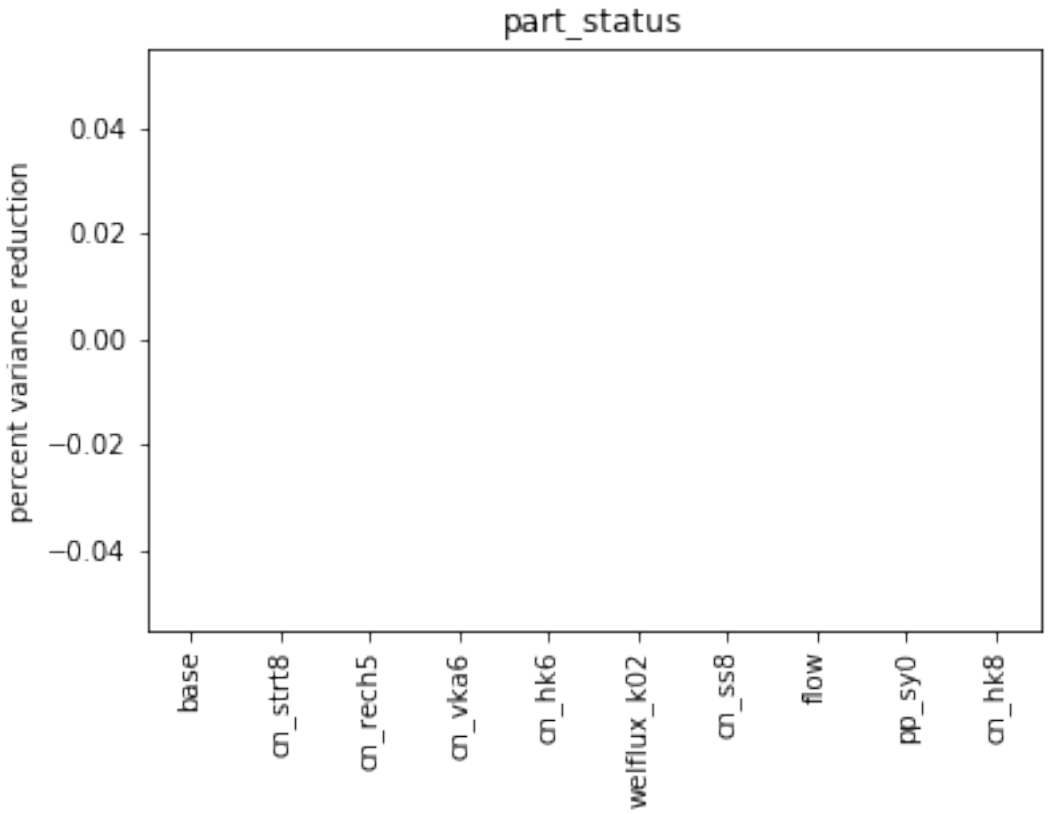

fa_hw_19801229
percent variance reduction
0
10
20
30
40
50
60
welflux
welflux_k02
flow
pp_hk2
strk
cn_sy6
cn_hk8
cn_rech5
pp_sy0
cn_hk6

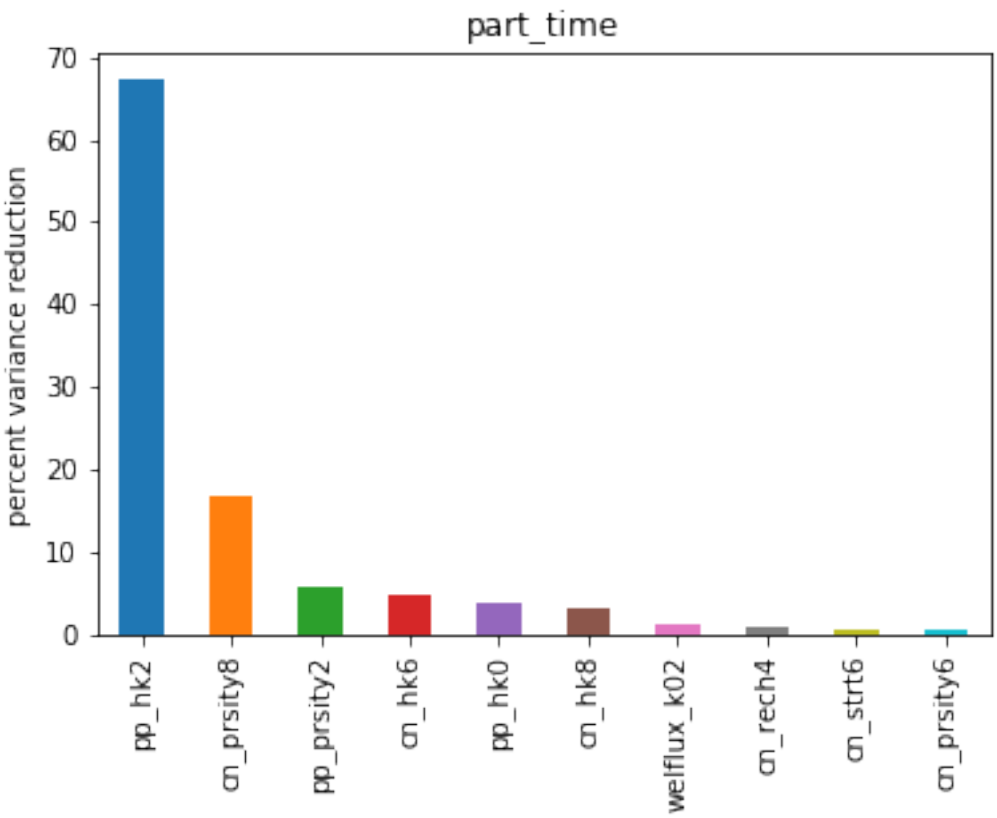

fa_tw_19791230
percent variance reduction
0
10
20
30
40
50
flow
pp_hk2
strk
welflux_k02
welflux
cn_hk8
cn_rech4
cn_hk6
pp_hk0
cn_strt6


fa_tw_19801229
percent variance reduction
80
70
60
50
40
30
20
10
0
welflux
welflux_k02
pp_hk2
strk
cn_sy6
flow
cn_rech5
pp_sy0
cn_hk8
cn_rech4


hds_00_013_002_000
percent variance reduction
60
50
40
30
20
10
0
pp_hk2
pp_hk0
cn_hk8
flow
welflux
cn_hk6
strk
welflux_k02
pp_rech0
pp_sy0


hds_00_013_002_001
percent variance reduction
0
5
10
15
20
25
30
35
40
pp_hk2
welflux
cn_sy6
pp_hk0
welflux_k02
pp_sy0
cn_hk8
cn_rech4
flow
cn_rech5


part_status
percent variance reduction
0.04
0.02
0.00
−0.02
−0.04
base
cn_strt8
cn_rech5
cn_vka6
cn_hk6
welflux_k02
cn_ss8
flow
pp_sy0
cn_hk8



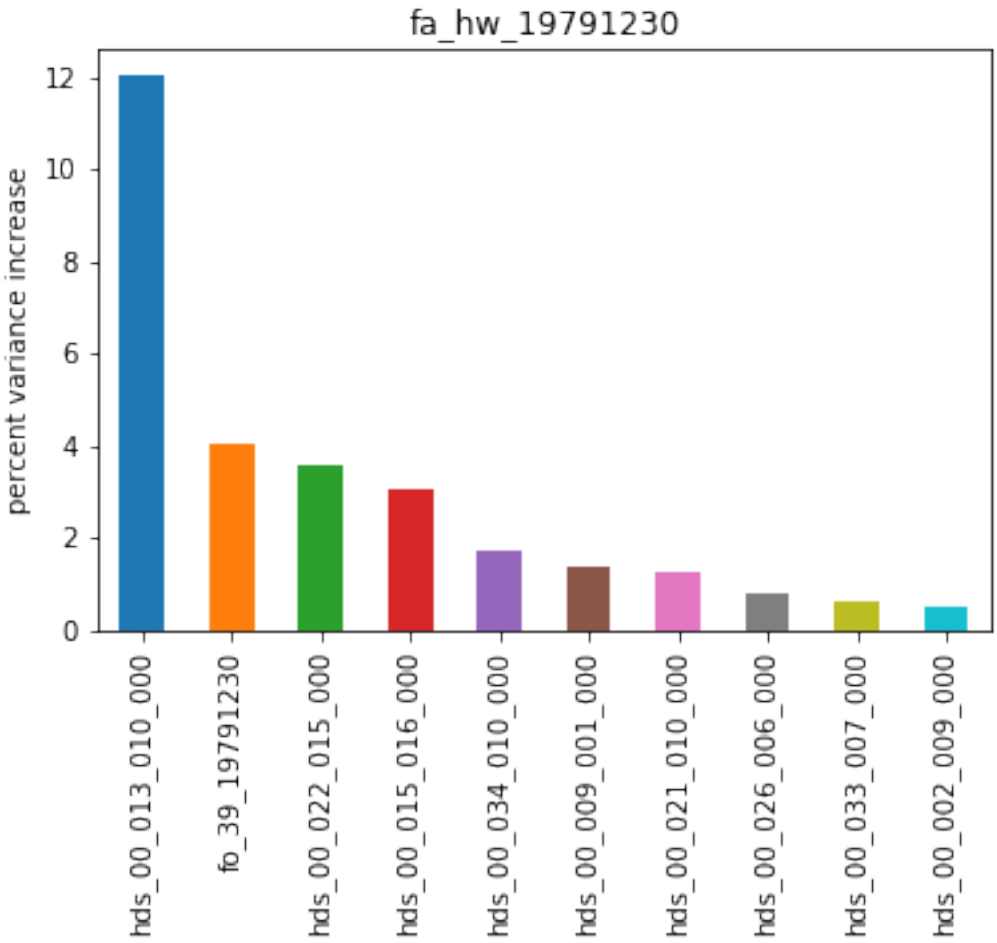

```
In [11]: df = sc.get_removed_obs_importance()
         base = df.loc["base",:]
         df = 100 * (df - base) / base
         df
```

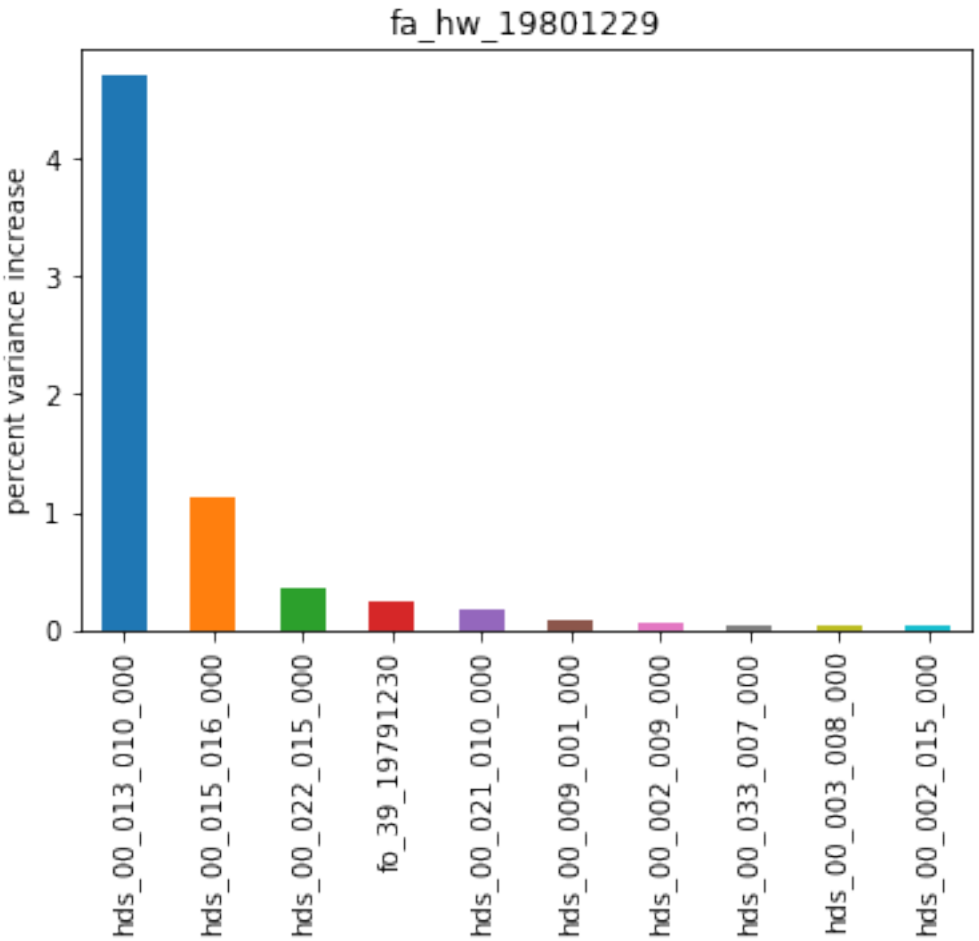
Out[11]:

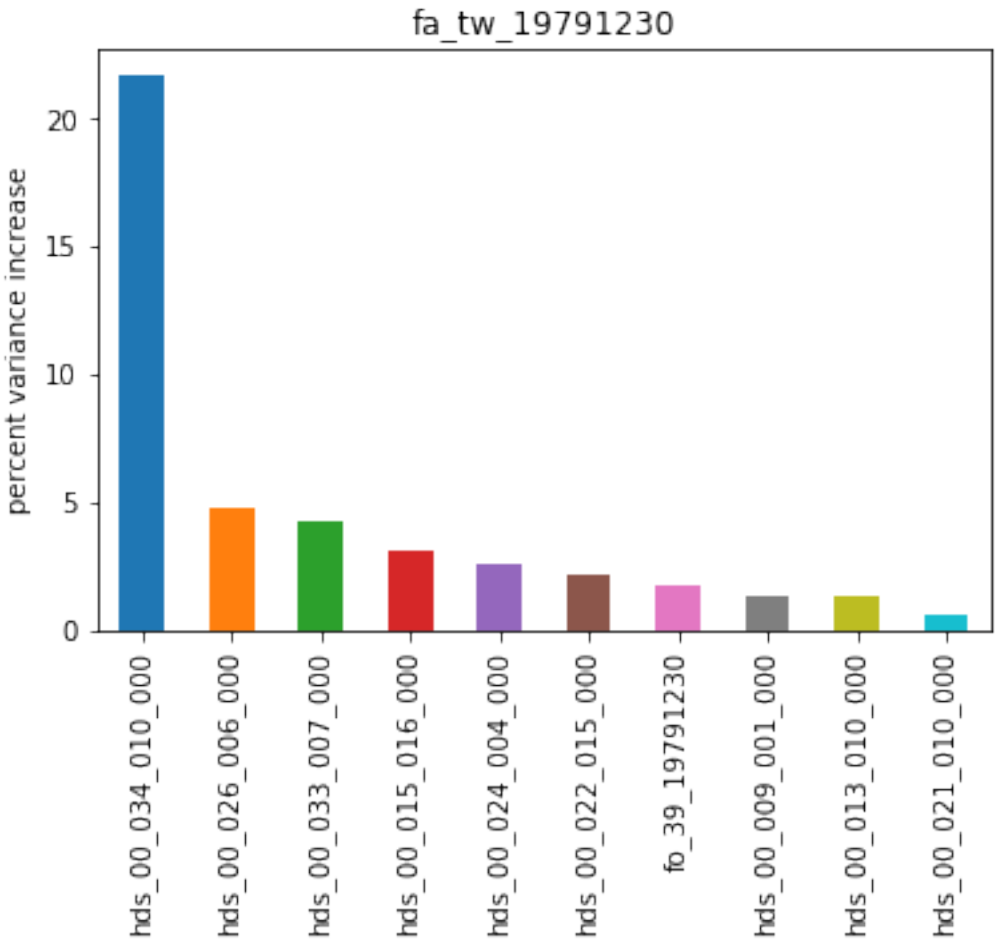
	fa_hw_19791230	fa_hw_19801229	fa_tw_19791230	\
base	0.000000	0.000000	0.000000	
fo_39_19791230	4.012877	0.241313	1.690854	
hds_00_029_015_000	0.247682	0.015491	0.231082	
hds_00_002_015_000	0.048175	0.030032	0.025861	
hds_00_034_010_000	1.742413	0.023564	21.674707	
hds_00_002_009_000	0.476543	0.056929	0.031141	
hds_00_026_006_000	0.799357	0.002194	4.780637	
hds_00_021_010_000	1.273090	0.182038	0.622944	
hds_00_024_004_000	0.335069	0.009109	2.535588	
hds_00_022_015_000	3.574861	0.353209	2.104635	
hds_00_009_001_000	1.380803	0.079675	1.358300	
hds_00_033_007_000	0.586518	0.044459	4.212460	
hds_00_003_008_000	0.289715	0.030337	0.000117	
hds_00_015_016_000	3.019865	1.114461	3.096199	
hds_00_013_010_000	12.012003	4.702808	1.276416	

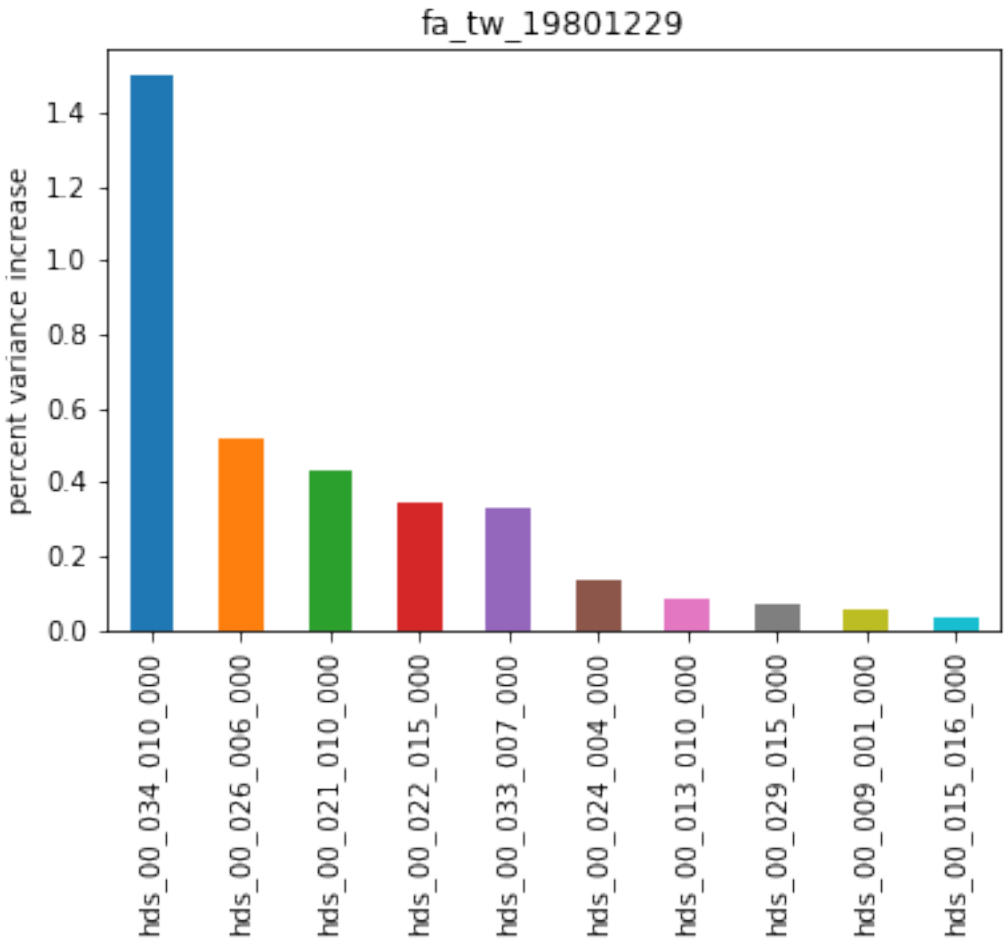
```
                      fa_tw_19801229  hds_00_013_002_000  hds_00_013_002_001  \
base                        0.000000            0.000000            0.000000
fo_39_19791230              0.013423            0.073945            0.176483
hds_00_029_015_000          0.066738            0.035723            0.029632
hds_00_002_015_000          0.000261            0.018346            0.001031
hds_00_034_010_000          1.497769            0.000073            0.000256
hds_00_002_009_000          0.000093            0.097950            0.037836
hds_00_026_006_000          0.514749            1.037770            0.534080
hds_00_021_010_000          0.428192            0.238214            0.209184
hds_00_024_004_000          0.133140           20.309296           13.241454
hds_00_022_015_000          0.343265            0.201569            0.167909
hds_00_009_001_000          0.055881          462.619008          262.711993
hds_00_033_007_000          0.331273            0.022534            0.014822
hds_00_003_008_000          0.000024            0.598035            0.289061
hds_00_015_016_000          0.030087            0.027176            0.093879
hds_00_013_010_000          0.081671            0.371386            0.101404

                      part_status  part_time
base                          NaN   0.000000
fo_39_19791230                NaN   0.032043
hds_00_029_015_000            NaN   0.090228
hds_00_002_015_000            NaN   0.018081
hds_00_034_010_000            NaN   0.016246
hds_00_002_009_000            NaN   0.041058
hds_00_026_006_000            NaN   0.748410
hds_00_021_010_000            NaN   0.005721
hds_00_024_004_000            NaN  13.423321
hds_00_022_015_000            NaN   0.254313
hds_00_009_001_000            NaN   0.144889
hds_00_033_007_000            NaN   0.086371
hds_00_003_008_000            NaN   1.177053
hds_00_015_016_000            NaN   0.041669
hds_00_013_010_000            NaN   0.725848
```

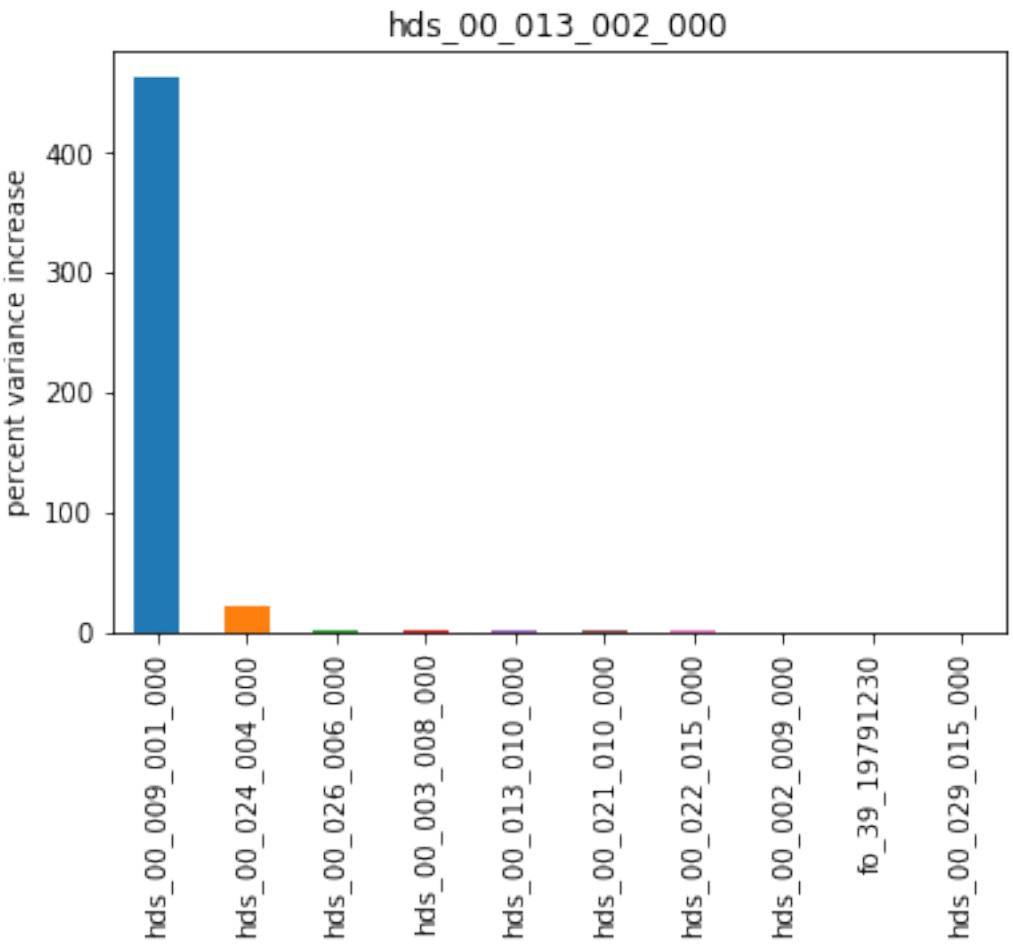
In [12]:
```
for forecast in df.columns:
    fore_df = df.loc[:,forecast].copy()
    fore_df.sort_values(inplace=True, ascending=False)
    ax = fore_df.iloc[:10].plot(kind="bar")
    ax.set_title(forecast)
    ax.set_ylabel("percent variance increase")
    plt.show()
```

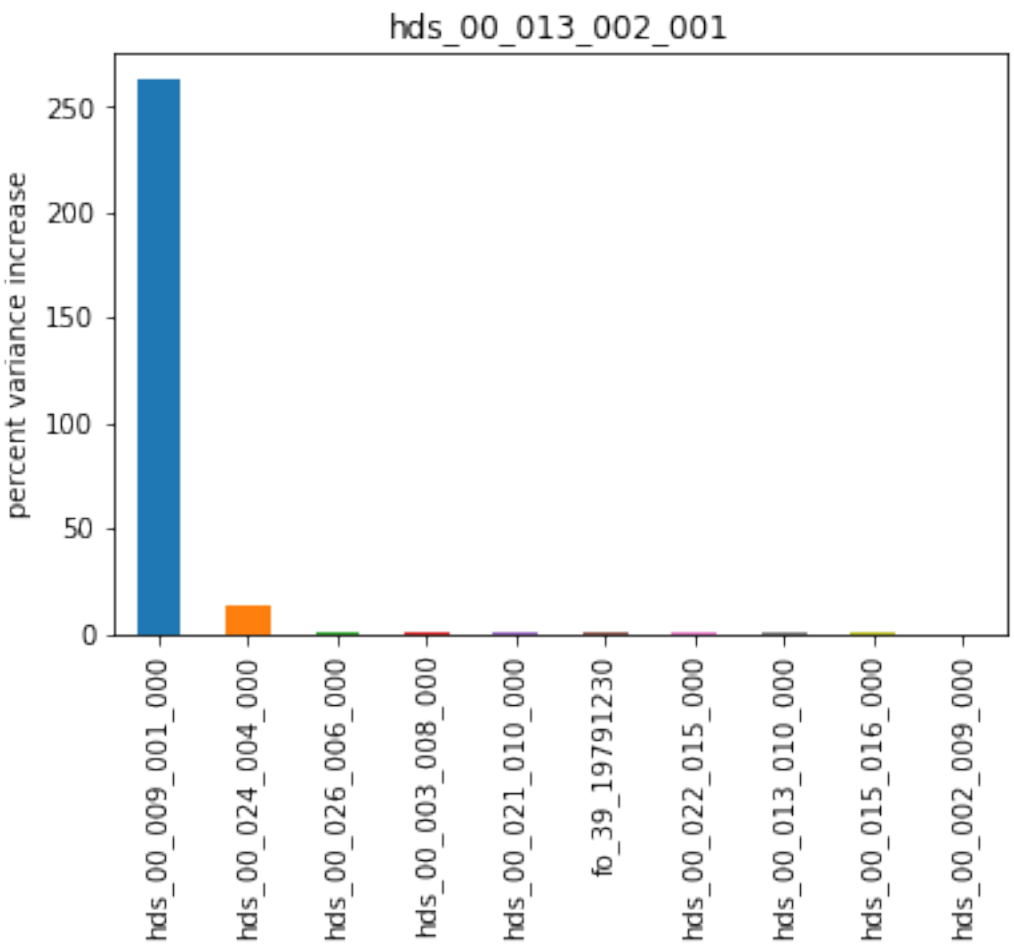

fa_hw_19791230
percent variance increase
0
2
4
6
8
10
12
hds_00_013_010_000
fo_39_19791230
hds_00_022_015_000
hds_00_015_016_000
hds_00_034_010_000
hds_00_009_001_000
hds_00_021_010_000
hds_00_026_006_000
hds_00_033_007_000
hds_00_002_009_000

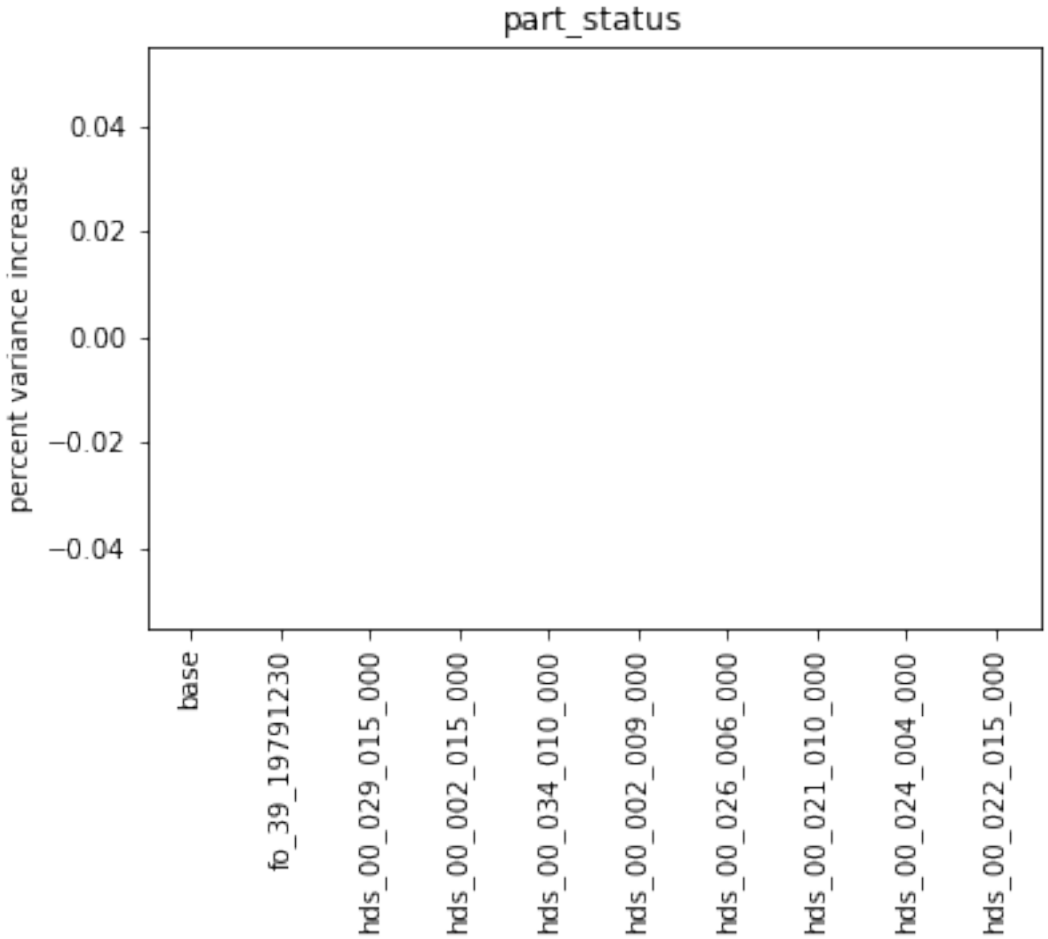

fa_hw_19801229
percent variance increase
0
1
2
3
4
hds_00_013_010_000
hds_00_015_016_000
hds_00_022_015_000
fo_39_19791230
hds_00_021_010_000
hds_00_009_001_000
hds_00_002_009_000
hds_00_033_007_000
hds_00_003_008_000
hds_00_002_015_000

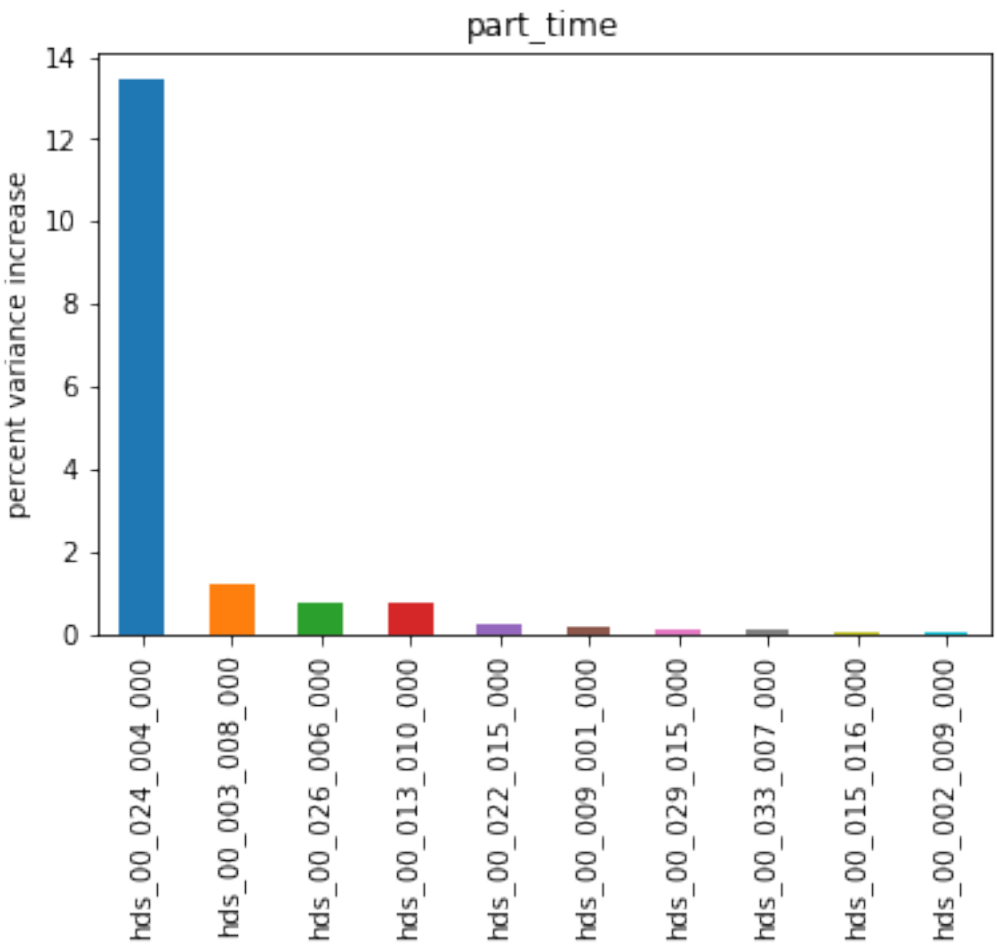

fa_tw_19791230
percent variance increase
0
5
10
15
20
hds_00_034_010_000
hds_00_026_006_000
hds_00_033_007_000
hds_00_015_016_000
hds_00_024_004_000
hds_00_022_015_000
fo_39_19791230
hds_00_009_001_000
hds_00_013_010_000
hds_00_021_010_000


fa_tw_19801229
percent variance increase
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
hds_00_034_010_000
hds_00_026_006_000
hds_00_021_010_000
hds_00_022_015_000
hds_00_033_007_000
hds_00_024_004_000
hds_00_013_010_000
hds_00_029_015_000
hds_00_009_001_000
hds_00_015_016_000


hds_00_013_002_000
percent variance increase
400
300
200
100
0
hds_00_009_001_000
hds_00_024_004_000
hds_00_026_006_000
hds_00_003_008_000
hds_00_013_010_000
hds_00_021_010_000
hds_00_022_015_000
hds_00_002_009_000
fo_39_19791230
hds_00_029_015_000


hds_00_013_002_001
percent variance increase
250
200
150
100
50
0
hds_00_009_001_000
hds_00_024_004_000
hds_00_026_006_000
hds_00_003_008_000
hds_00_021_010_000
fo_39_19791230
hds_00_022_015_000
hds_00_013_010_000
hds_00_015_016_000
hds_00_002_009_000


part_status
percent variance increase
0.04
0.02
0.00
−0.02
−0.04
base
fo_39_19791230
hds_00_029_015_000
hds_00_002_015_000
hds_00_034_010_000
hds_00_002_009_000
hds_00_026_006_000
hds_00_021_010_000
hds_00_024_004_000
hds_00_022_015_000




```
In [13]: df = sc.get_added_obs_importance()
```