

In [1]:

## Data

- Les colonnes représentent des bâtiments différents
- Le fichier "sample\_data.csv" contient les consommations électriques au pas de temps 10 minutes de chacun des bâtiments sur la période 01/04/2017 - 20/12/2017

	Bat1	Bat2	Bat3	Bat4	Bat5	Bat6	Bat7	Bat8	Bat9	Bat10
2017-04-01 00:00:00	387.0	542.0	484.0	443.0	367.0	681.0	632.0	508.0	277.0	468.0
2017-04-01 00:10:00	384.0	544.0	518.0	459.0	362.0	717.0	652.0	518.0	330.0	481.0
2017-04-01 00:20:00	373.0	540.0	532.0	467.0	361.0	732.0	639.0	504.0	334.0	472.0
2017-04-01 00:30:00	347.0	542.0	521.0	448.0	358.0	723.0	619.0	427.0	329.0	451.0
2017-04-01 00:40:00	342.0	549.0	504.0	455.0	372.0	661.0	639.0	408.0	326.0	406.0
2017-04-01 00:50:00	350.0	563.0	535.0	452.0	392.0	685.0	630.0	405.0	317.0	414.0
2017-04-01 01:00:00	353.0	547.0	536.0	454.0	397.0	717.0	672.0	409.0	313.0	389.0
2017-04-01 01:10:00	337.0	551.0	513.0	428.0	365.0	731.0	672.0	407.0	294.0	417.0
2017-04-01 01:20:00	341.0	550.0	476.0	433.0	387.0	642.0	671.0	403.0	293.0	400.0
2017-04-01 01:30:00	349.0	566.0	551.0	426.0	378.0	699.0	663.0	405.0	299.0	393.0

- Le fichier "label\_data.csv" contient l'état d'ouverture des bâtiments (1 pour ouvert, 0 pour fermé) au pas de temps horaire sur la période 01/04/2017 - 20/12/2017

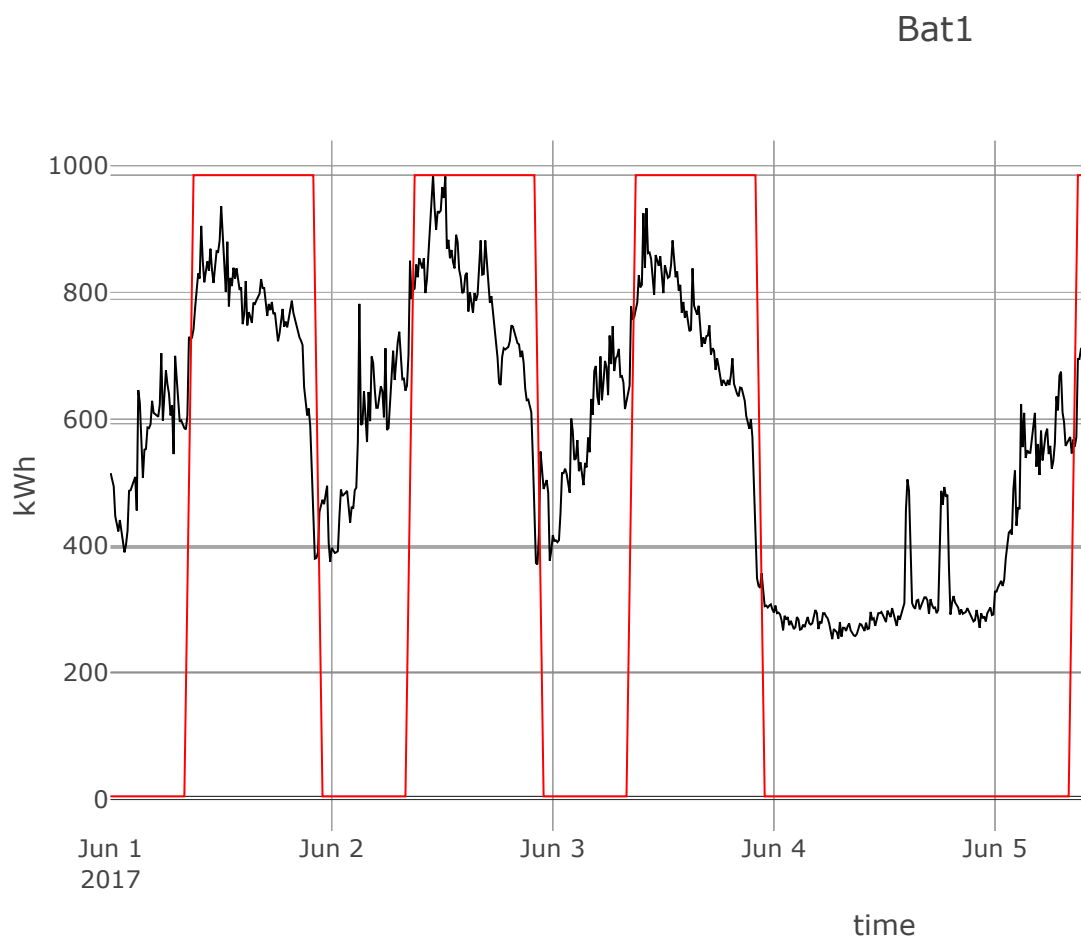
	Bat1	Bat2	Bat3	Bat4	Bat5	Bat6	Bat7	Bat8	Bat9	Bat10
2017-04-01 00:00:00	0	0	0	0	0	0	0	0	0	0
2017-04-01 01:00:00	0	0	0	0	0	0	0	0	0	0
2017-04-01 02:00:00	0	0	0	0	0	0	0	0	0	0
2017-04-01 03:00:00	0	0	0	0	0	0	0	0	0	0
2017-04-01 04:00:00	0	0	0	0	0	0	0	0	0	0
2017-04-01 05:00:00	0	0	0	0	0	0	0	0	0	0
2017-04-01 06:00:00	0	0	0	0	0	0	0	0	0	0
2017-04-01 07:00:00	0	0	0	0	0	0	0	0	0	0
2017-04-01 08:00:00	0	0	0	0	0	0	0	0	0	0
2017-04-01 09:00:00	1	1	1	1	1	1	1	1	1	1

In [4]:

```
input_path = "/home/ubuntu/notebooks/exercice/"  
data = pd.read_csv(input_path + "sample_data.csv", index_col=0, parse_dates=[0])  
labels = pd.read_csv(input_path + "label_data.csv", index_col=0, parse_dates=[0])
```

In [8]:

```
start_date = dt.datetime(2017, 6, 1)  
end_date = dt.datetime(2017, 6, 8)  
plot_results_classif(data, labels, 'Bat1', start_date, end_date)
```



## Exercice 1: Preprocessing

Se ramener au pas de temps horaire en calculant la moyenne horaire des consommations électriques

## Exercice 2: Feature Engineering

- (a) Utiliser Scikit Learn pour normaliser les consommations électriques
- (b) Flagger les weekends et jours fériés

résultat attendu:

	Bat1	Bat2	Bat3	Bat4	Bat5	Bat6	Bat7	Bat8	Bat9	Bat10	day_off
<b>time</b>											
<b>2017-04-01 00:00:00</b>	v01	v02	v03	v04	v05	v06	v07	v08	v09	v010	d0
<b>2017-04-01 01:00:00</b>	v11	v12	v13	v14	v15	v16	v17	v18	v19	v110	d1
<b>2017-04-01 02:00:00</b>	v21	v22	v23	v24	v25	v26	v27	v28	v29	v210	d2
<b>2017-04-01 03:00:00</b>	v31	v32	v33	v34	v35	v36	v37	v38	v39	v310	d3
<b>2017-04-01 04:00:00</b>	v41	v42	v43	v44	v45	v46	v47	v48	v49	v410	d4
<b>2017-04-01 05:00:00</b>	v51	v52	v53	v54	v55	v56	v57	v58	v59	v510	d5
<b>2017-04-01 06:00:00</b>	v61	v62	v63	v64	v65	v66	v67	v68	v69	v610	d6
<b>2017-04-01 07:00:00</b>	v71	v72	v73	v74	v75	v76	v77	v78	v79	v710	d7
<b>2017-04-01 08:00:00</b>	v81	v82	v83	v84	v85	v86	v87	v88	v89	v810	d8
<b>2017-04-01 09:00:00</b>	v91	v92	v93	v94	v95	v96	v97	v98	v99	v910	d9

## Exercice 3: Jointure

- (a) Redimensionner la table des consommations afin d'avoir les bâtiments en ligne
- (b) Ajouter les labels à la table des consommations

La table attendue devra avoir la forme:

		day_off	value	label
time	building			
2017-04-01 00:00:00	Bat1	d1	v0	l0
2017-04-01 01:00:00	Bat1	d1	v1	l1
2017-04-01 02:00:00	Bat1	d1	v2	l2
2017-04-01 03:00:00	Bat1	d1	v3	l3
2017-04-01 04:00:00	Bat1	d1	v4	l4
2017-04-01 05:00:00	Bat1	d1	v5	l5
2017-04-01 06:00:00	Bat1	d1	v6	l6
2017-04-01 07:00:00	Bat1	d1	v7	l7
2017-04-01 08:00:00	Bat1	d1	v8	l8
2017-04-01 09:00:00	Bat1	d1	v9	l9

## Exercice 4: Classification

- (a) Utiliser un algorithme de classification pour prédire les heures d'ouvertures
- (b) Evaluer votre/vos modèle(s)

## Exercice 5: Interprétation

- (a) Proposer une visualisation des résultats par bâtiment
- (b) Quelle(s) autre(s) variable(s) explicative(s) peut/peuvent être utilisée(s) pour améliorer les résultats du modèle