

RAPPORT D'ETUDE

Pascal Doré

Limoges 29 juillet 2013

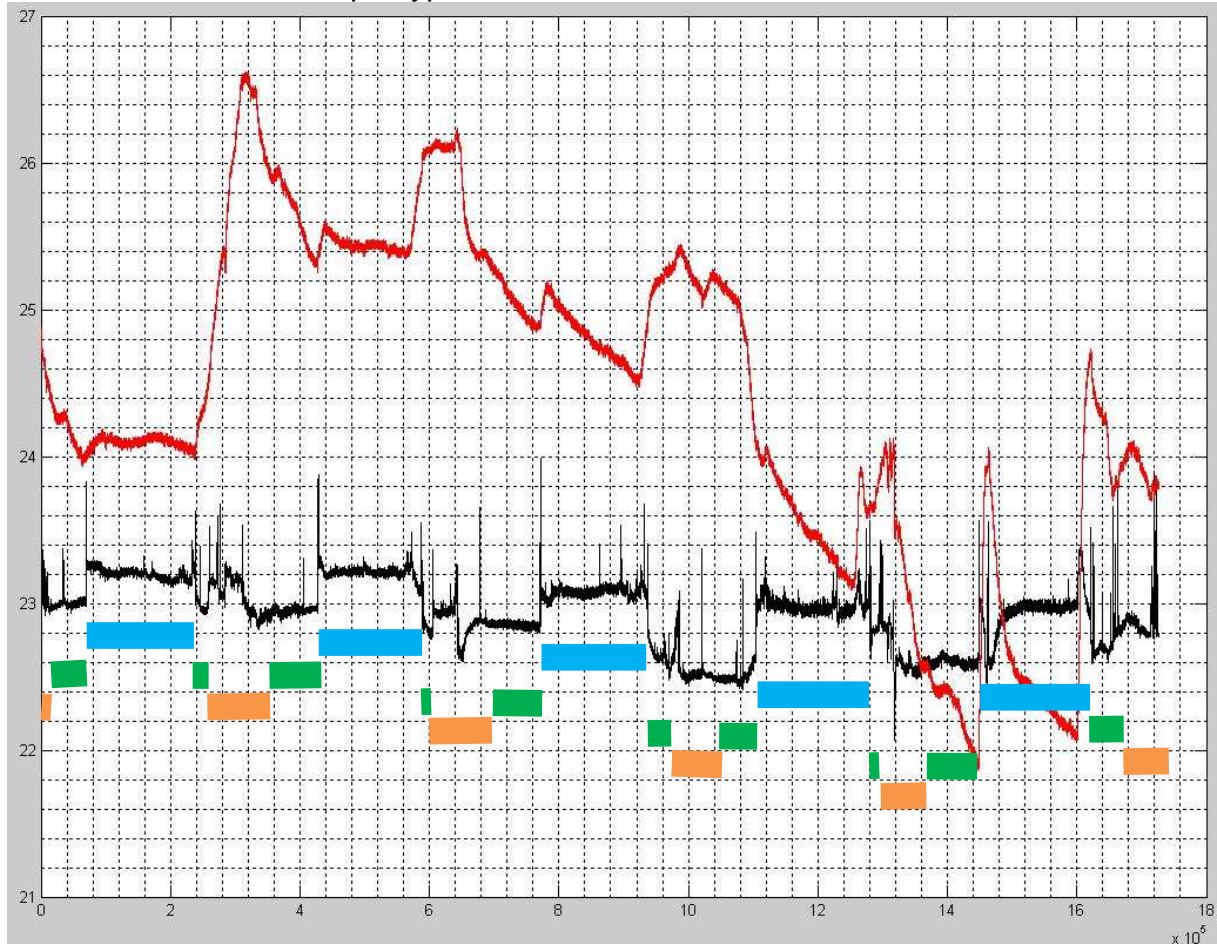
Objet : Sujet SEME, algorithme de détection de motif dans une série temporelle longue

Nous disposons de données échantillonnées à 4 Hz sous la forme d'un fichier CSV

jour	hh :mm :ss	Température envir.	Température mesurée	Filtre médian Temp. envir.	Filtre médian Temp. mes.
0	00:00:00	26,222	24,381	26,222	24,381
0	00:00:01	26,222	24,388	26,222	24,381
0	00:00:01	26,222	24,388	26,222	24,388
0	00:00:01	26,211	24,387	26,222	24,388
0	00:00:01	26,2	24,372	26,211	24,387
0	00:00:02	26,211	24,38	26,211	24,38
0	00:00:02	26,2	24,379	26,2	24,379
0	00:00:02	26,222	24,398	26,211	24,38
0	00:00:02	26,2	24,369	26,2	24,379
0	00:00:03	26,211	24,38	26,211	24,38
0	00:00:03	26,222	24,398	26,211	24,38
0	00:00:03	26,211	24,39	26,211	24,39
0	00:00:03	26,222	24,395	26,222	24,395
0	00:00:04	26,211	24,387	26,211	24,39

L'horodatage est relatif et démarre à 0 à chaque initialisation du capteur (il s'agit d'un prototype). La première colonne donne le n° du jour, la deuxième donne l'heure, la troisième la température de l'environnement (nous sommes dans un bâtiment, cette température est égale à la température ambiante + 1,1°C) la quatrième, la température que l'on mesure. Les 2 dernières sont le résultat de l'application d'un filtre médian effectué sur les 3 dernières mesures de la troisième et de la quatrième colonne. Cette opération a pour but de supprimer les valeurs aberrantes issues du prototype.

Sur ce graphe sont représentés en rouge la courbe qui correspond à la colonne 5 (temp. environnement filtrée) et en noir la différence colonne 6 – colonne 5 à laquelle on a ajouté une constante (juste pour l'affichage). C'est ce signal qui nous intéresse, malheureusement, il a une amplitude faible, il est soumis à des perturbations quand le $\Delta\theta/\Delta t$ de l'environnement est important et par ailleurs il dérive dans le temps ce qui exclue un traitement simple type détection à seuil.



L'objectif serait d'avoir la réponse niveau haut (en bleu) et niveau bas (en vert). Nous avons représenté ces informations qui nous intéressent, et ajouté un niveau 'intermédiaire' en orange pour lequel la fiabilité de la réponse nous importe peu. Sont représentés les 2 niveaux intéressants : en bleu le niveau haut et en vert le niveau bas. En orange le niveau intermédiaire moins intéressant. En clair la mesure de performance d'un algorithme ne tiendrait compte que des résultats sur les classes bleu et vert (niveau haut & niveau bas)

Nous disposons d'un fichier labellisé, c'est-à-dire que nous avons une colonne supplémentaire précisant la classe, qui représente donc le résultat attendu.