

BD Climato: JSON, base de données et traitements

1. Historique

- V1.00 : 26/01/2022 : Nouvelle version totalement revue, et réorganisée
- V1.01 : 22/02/2022 : Corrections d'Hubert
- V1.02 : 24/02/2022 : Ajout contraintes « current » / « aggregations » page 5
- V1.03 : 24/02/2022 : Révision page 7, et page 5 une importation de valeurs pré-agrégées bloque toute importation d'autres mesures pré-agrégées à ce niveau.
- V1.20 : 27/02/2022 : Nouveau format Json, et précisions sur recalcul des extrêmes en cas de reprise à partir de l'historique WeeWX
- V1.21 : 27/02/2022 : Changement de place de la clé « duration »
- V1.22 : 07/03/2022 : Précision sur les dates locales/globales. Correction du Json exemple

2. TOC

1. Historique	1
2. TOC	1
3. Introduction	1
4. Système d'acquisition de données	2
5. WeeWX	2
6. Dates locales et globales	3
7. Spécification du fichier JSON d'importation	3
a. Format des données dans le JSON	3
b. Clé au niveau racine	3
c. Clé « info »	4
d. Clé « data »	4
e. Clé « extremes »	6
ANNEXE A : Définition des mesures	8
ANNEXE B : Nom des clés JSON/Observation/Agrégations	9
ANNEXE C : Unités utilisées	16
ANNEXE D : Service d'importation de fichiers JSON	17
ANNEXE E : Service de calcul d'agréations	18
ANNEXE F : Commentaires sur calcul OMM	20
ANNEXE G : Service de recalcul des extrêmes	21
ANNEXE H : changement du format Json	22
ANNEXE I : Exemple JSON données observation	23

3. Introduction

Toutes les agrégations sont automatiquement calculées à partir des valeurs élémentaires, qui sont stockées dans la table « observation ». La seule façon de modifier une agrégation est de

BD Climato: JSON, base de données et traitements

passer par une modification de l'observation associée (en fait une suppression, et insertion de nouvelles mesures). D'où l'importance de bien spécifier le contenu du fichier Json.

4. Système d'acquisition de données

La VP2 envoie des mesures instantanées toutes les deux secondes. Chaque envoi contient des mesures partielles, une fois la température, une autre fois la pression, etc... Le système d'acquisition doit regrouper ces informations partielles pour avoir une vision de tous les capteurs. Cela se faisant sur une certaine période, certaines valeurs doivent être agrégées à partir des mesures instantanées.

Chaque système d'acquisition peut avoir une stratégie différente dans les détails, mais ont globalement un fonctionnement assez semblable. Les mesures instantanées sont rarement archivées. Seule la version pré-agrégée sur cette courte période est stockée et est utilisée par les interfaces graphiques du système d'acquisition. Ce sont ces données agrégées que le système d'acquisition va stocker, ou distribuer. Nous appellerons ces données les **données élémentaires**, pour ne pas les confondre avec les agrégations faites dans la BD Climato.

A partir des mesures instantanées le système d'acquisition détecte les Min/Max, et garde en mémoire ces valeurs extrêmes, ainsi que l'heure ou cet extrême a été détecté. D'autres informations, suivant les instruments, comme la direction du vent au moment de l'extrême sont aussi gardées en mémoire.

5. WeeWX

Le réseau MeteOR-OI utilise principalement WeeWX comme « système d'acquisition ». Donc ces deux termes seront utilisés de façon interchangeable dans ce document, la BD Climato est conçue pour utiliser d'autres « système d'acquisition »

WeeWX reçoit les mesures instantanées, appelé « LOOP » (correspond aux mesures instantanées). Un « LOOP » va contenir les informations de température, le suivant sur la pluie, le suivant la température, puis le vent... « LOOP » ne contient que des informations partielles sur un domaine donné. WeeWX analyse ces données et les sauve de deux façons :

- **Données élémentaires**, regroupement sur une période de 5mn, stockées dans la table « archive ». Le dateTime de WeeWX contient l'heure de fin de la mesure (stop_dat)
- **Valeurs Extrêmes** ainsi que des informations complémentaires (heure de l'extrême, direction du vent au moment de l'extrême relevé pour le vent...) dans des tables « records » spécifiques suivant différents niveaux de regroupement (daily, month,...). *Les extrêmes WeeWX sont archivés par jour, mois, année seulement. En cas de reprise à partir de l'historique WeeWX, cette limitation va avoir un certain impact sur les données récupérées (voir chapitre 6.e : Clé « extremes »)*

Plus d'info (anglais) sur : <http://www.weewx.com/docs/customizing.htm>

Le module de génération des fichiers à partir de WeeWx part du principe que les blocks de données élémentaires stockés dans la base de données ne se recouvrent pas, et sont contigus.

BD Climato: JSON, base de données et traitements

Un « template » est utilisé pour générer un fichier JSON tous les 5mn, qui est envoyé vers le serveur Climato. Le « template » utilisé par WeeWX est défini dans (A vérifier...):

/srv/weewx/root/skins/Bootstrap/json/obs.YYYY-MM-DD-HH-mm.json

Le fichier JSON généré est stocké dans le raspberry dans /srv/weewx/root/public_html/json (A vérifier), avant envoi automatique vers le serveur BD Climato.

6. Dates locales et globales

Toutes les dates sont stockées en heure locale, sans time zone. Les champs stop_dat, start_dat sont des datetime sql. Le JSON ne connaissant pas le type date, celles-ci sont stockées comme une chaîne de caractères sous le format : YYYY-MM-DDTHH-mm-ss

Les agrégations sont calculées à partir des date-heures locale. La première heure de la journée sera donc 00 :05 :00.

Il est possible de corriger l'heure locale en ramenant tout à l'heure de la Réunion (peut être heure TU si désiré), en additionnant le nombre d'heures mis dans le champ fuseau_delta de la table poste, pour la station donnée. Cela permet par exemple de pouvoir comparer les pressions sur une même heure, même sur des stations se trouvant sur des fuseaux horaires différents, donc ayant une heure locale différente. Cette conversion doit être faite au niveau du programme d'affichage (Grafana, ou autre)

7. Spécification du fichier JSON d'importation

La structure définie ci-après peut être répétée (contenu dans un tableau) ou pas (simple parenthèses) dans le fichier JSON. Si elle est répétée, chaque entrée doit avoir le même code *meteor* pour chacune des répétitions (= doit venir d'une même station du réseau)

a. Format des données dans le JSON

Les valeurs numériques sont stockées sans les « » (comme des valeurs numériques JSON), avec un point décimal, et en respectant l'unité spécifiée en appendice

Les dates sont en heure locale, dans une chaîne de caractères, sous la forme :

« yyyy-mm-dd hh:mm:ss »

b. Clé au niveau racine

Les clés suivantes sont stockées à ce niveau :

- **meteor** : Code meteor de la station
- **info** : Sous JSON contenant les informations sur le contenu du fichier, et information de la station.
- **data** : Tableau de mesures pour une même station.
- **extremes** : Sous JSON facultatif contenant les extrêmes journaliers, présent seulement quand le JSON contient des données régénérées au niveau observation, ou agrégation horaire.

meteor	Nom meteor -> permet de retrouver le poste_id dans la base de données
---------------	---

BD Climato: JSON, base de données et traitements

c. Clé « info »

Information pour les pages html qui exploiteraient ce fichier, non utilisé par Climato (liste exacte à préciser plus tard par le groupe en charge de la génération des pages HTML). Les deux premières clefs sont obligatoires.

json_type	Type de fichier Json : « O » : données observation venant de la station au fil de l'eau « C » : régénération de données élémentaires « H » : régénération à partir des agrégations horaires de WeeWX « D » : régénération à partir des agrégations journalières de WeeWX « M » : régénération à partir des agrégations mensuelles de WeeWX « Y » : régénération à partir des agrégations annuelles de WeeWX « A » : régénération à partir des agrégations globales de WeeWX
version	1 pour cette spécification
Les informations suivantes ne sont pas utilisées par le serveur	
title	Météo en direct
location	\$station.location
lat	\$station.latitude[0]° \$station.latitude[1]' \$station.latitude[2]
lon	\$station.longitude[0]° \$station.longitude[1]' \$station.longitude[2]
alt	\$station.altitude
alt_unit	«meter»
hardware	\$station.hardware
uptime	\$station.uptime
serverUptime	\$station.os_uptime
softVersion	\$station.version
softName	« weewx », ou autre

d. Clé « data »

Tableau de données élémentaires, ou pré-agrégées.

Le fichier Json correspond à plusieurs cas différents d'utilisation :

- Chargement de données élémentaires venant au fil de l'eau d'une station (« json_type » = « O »). Pour chaque « data » la « stop_dat » de la période d'observation doit être fournie, ainsi que les valeurs observées, et extrêmes de la « duration » dans la sous clé « valeurs »
- Chargement de données élémentaires régénérées à partir de WeeWX (« json_type » = « C »). Pour chaque « data » la « stop_dat » de la période d'observation doit être fournie, ainsi que les valeurs observées dans la sous clé « valeurs ». Les valeurs extrêmes du jour correspondant seront aussi fournies dans la clé « extremes », ce qui limite la plage de dates des observations au jour concerné, sans obligatoirement couvrir l'intégralité du jour. Seules les « extremes » tombant dans une période d'observation fourni seront pris en compte
- Chargement de données agrégées par WeeWX pour un niveau donné (autres « json_type »). Pour chaque « data » la date de début de la période d'agrégation doit être fournie, ainsi que les valeurs agrégées dans la sous clé « valeurs ». Pour le niveau

BD Climato: JSON, base de données et traitements

horaire (« json_type » = « H »), les valeurs extrêmes du jour concerné seront aussi fournies dans la clé « extremes », et donc seules les données agrégée d'un jour pourront se trouver dans une clause « data »

Une seule des dates « stop_dat » ou « start_dat » doit être fournie suivant les cas d'utilisation.

stop_dat	Datetime de la fin de la durée de l'agrégation des données élémentaires. Doit être fourni que pour les « type_json » « O » ou « C »
duration	Duration en minutes, seulement pour les « type_json » « O » ou « C »
start_dat	DateTime du début de la période agrégée. Doit être fourni que pour les « type_json » « H », « D », « M », « Y » ou « A »
force_replace	Par défaut, si l'observation ou les valeurs pré-agrégées existent déjà pour cette date, les données sont ignorées, et un incident est créé coté serveur. Le flag « force_replace » avec la valeur « true » permet de forcer la suppression de l'observation existante et le remplacement par les nouvelles valeurs fournies. Les agrégations seront tenues à jour.

Chaque valeur a une clé bien définie, tel que décrit en appendice B, en fonction du « type_json »

i. Clé des mesures, fichier « type_json » « O » ou « C »

Les données de « valeurs » correspondent aux données élémentaires de la station. Chaque valeur a une clé bien définie, tel que décrit en appendice B, colonne « JSON type O ou C »

La « duration » de la période d'observation doit être fournie dans « valeurs »

La clé avec le nom de la mesure contient la valeur instantanée de la mesure.

La clé de la mesure, avec un suffixe « _avg » ou « _sum » contient une valeur agrégée sur la période d'observation (LOOP).

Quelques exemples :

- « out_temp » représente une valeur instantanée
- « out_temp_avg » correspond à une valeur agrégée sur la période de l'observation.
- « rain_sum » correspond à la somme des pluies de la période (duration).
- « voltage » correspond à une valeur instantanée

Certaines valeurs doivent être fournies sous forme instantanée (par exemple voltage). Ces contraintes sont spécifiées dans la définition des mesures ci-après.

En général quand une des formes n'est pas fournie, le programme essayera d'utiliser l'autre forme si elle est disponible, sauf indication contraire dans la définition des mesures.

Les valeurs extrêmes fournies, si elles le sont, doivent être des valeurs instantanées, et doivent utiliser la clé de la valeur avec un suffixe « _max » ou « _min » (par exemple

BD Climato: JSON, base de données et traitements

« *out_temp_max* »). Elles sont associées à une date/heure qui a un suffixe « *_time* » (par exemple « *out_temp_max_time* »).

Dans le cas de mesure de vent, un suffixe « *_dir* » peut être utilisé pour indiquer la direction du vent. Cela est utilisable sur la mesure élémentaire, ou sur les extrêmes. (Exemple : « *wind_dir* », « *wind_dir_max_dir* »)

Pour les mesures qui sont agrégées sous forme simple et sous forme OMM, il est recommandé de fournir l'information sous les deux formes, la forme agrégée étant utilisée pour les calculs d'agrégation de la valeur, alors que la valeur instantanée sera utilisée pour les calculs OMM, et pour simuler les valeurs extrêmes OMM ou non si non fournies.

ii. Clé des mesures, fichier « type_json » H-D-M-Y ou A

Les données présentes sous « valeurs » correspondent aux données pré-agrégées par WeeWX pour le niveau demandé. Chaque valeur a une clé bien définie, tel que décrit en appendice B, colonne « JSON type HDMYA »

Pour chaque mesure la somme pondérée par la somme des « *duration* » est stockée dans la clé « *mesure_s* », et la somme des « *duration* » des valeurs agrégée est stocké dans la clé « *mesure_d* ».

Quelques exemples :

- « *out_temp_s* » : Somme des « *out_temp* » * « *duration* » de la période concernée
- « *out_temp_d* » : Somme des durations agrégées

Au niveau horaire, les extrêmes d'agrégation ne doivent pas être fournis. Il seront simulés à partir du calcul des données agrégées fournies (« *mesure_s* »/« *mesure_d* » pour les moyennes). Ces extrêmes seront corrigés avec les extrêmes journaliers fournis dans la clé « *extremes* »

Pour les autres niveaux (journaliers et supérieurs), les extrêmes doivent être fournis à partir des extrêmes stockés par WeeWX.

e. Clé « extremes »

WeeWX ne sauvegarde dans sa base de données que les extrêmes journaliers et niveaux supérieurs. Cela va avoir un impact en cas de récupération des données élémentaires et agrégée au niveau horaire. Dans ce cas les valeurs extrêmes journalières de weeWX sont fournies dans le JSON sous la clé « *extremes* », permettant au programme d'importation de mettre les bonnes valeurs extrêmes dans les observations, ou agrégations horaires correspondantes. Voici comment cela est géré :

- *Cas de reprise d'historique au niveau observations : Le Json va contenir les valeurs observées, ainsi que les extrêmes journaliers du jour concerné (dans la clé « extremes »). Le programme d'importation utilisera les extrêmes journaliers pour mettre à jour l'observation correspondante à cet extrême, cela sera automatiquement propagé dans l'agrégation horaire correspondante, donc pas de perte d'information pour l'extrême journalier (et niveaux supérieurs). Cependant pour les autres observations, et autres agrégations horaires, les extrêmes seront*

BD Climato: JSON, base de données et traitements

recalculés à partir des valeurs instantanées fournies.

- *Reprise d'historique au niveau agrégation horaire : Le Json va contenir les valeurs agrégées au niveau horaire, ainsi que les extrêmes journaliers du jour concerné (dans la clé « extremes »). Les extrêmes du jour seront mis à jour dans l'agrégation horaire correspondante, sans de perte d'information pour l'extrême journalier (et niveaux supérieurs). Cependant pour les autres agrégations horaires, les extrêmes seront recalculés à partir des valeurs moyennées sur la période, donc ne seront pas exactement correcte au niveau horaire seulement.*

Reprise d'historique au niveau agrégation journalière ou au-dessus : Pas de perte d'information, et donc pas besoin de la clé « extremes »

« extremes » va stocker uniquement les max/min journalier du jour correspondant à la période concernée. Les clés sont les mêmes que celles utilisées dans les tables d'agrégation et sont spécifiées dans l'appendice C, colonne « JSON, type : HDMYA »

BD Climato: JSON, base de données et traitements

ANNEXE A : Définition des mesures

La mise à jour de cette définition est à vérifier dans les fichiers se trouvant dans le répertoire du projet : app/classe/typeInstruments.

Clé mesure	Clé dans la BD	Type	Agrégation	Max	Min	Divers
<i>out temp</i>	out temp	A	avg	Oui	Oui	
<i>out temp</i>	out_temp_omm	I	avgomm	-7	+5	
<i>Windchill</i>	windchill	A+I	Non	Non	Oui	
<i>Heatindex</i>	heatindex	I	Non	Oui	Non	
<i>Dewpoint</i>	dewpoint	I	Non	Oui	Oui	
<i>Soiltemp</i>	soiltemp	I	Non	Non	Oui	
<i>Humidity</i>	humidity	A	avg	Oui	Oui	
<i>Humidity</i>	humidity_omm	I	avgomm	Oui	Oui	
<i>Barometer</i>	barometer	A	avg	Oui	Oui	
<i>Barometer</i>	barometer_omm	I	avgomm	Oui	Oui	
<i>Pressure</i>	pressure	I	Non	Oui	Oui	
<i>Wind inst</i>	wind inst	I	avg – no calc	Non	Non	+ wind inst dir
<i>Wind</i>	wind	A	Non	Oui	Non	+ wind dir
<i>Gust</i>	gust	A	Non	Non	Non	+ gust_time, + gust_dir
<i>Wind10</i>	wind10	A	avg	Oui	Non	+wind10_dir
<i>Wind10</i>	wind10_omm	I	avgomm	Oui	Non	+wind10_dir
<i>Rain</i>	rain	A	sum	Non	Non	
<i>Rain</i>	rain_omm	I	sumomm	-7	-7	
<i>Rain rate</i>	rain rate	I	Non	Oui	Non	
<i>Uv indice</i>	uv indice	I	Non	Oui	Non	
<i>Radiation</i>	radiation	A	sum	Oui	Oui	
<i>Etp</i>	etp		sum	Non	Non	Fourni à l'heure ronde pour 60 mn. cause de weeWX
<i>In temp</i>	in temp	A	avg	Oui	Oui	
<i>In humidity</i>	in_humidity	A	avg	Oui	Oui	
<i>rx</i>	rx	I	avg	Oui	Oui	
<i>Voltage</i>	voltage	I	avg	Oui	Oui	

BD Climato: JSON, base de données et traitements

ANNEXE B : Nom des clés JSON/Observation/Agrégations

Légende des colonnes :

Clé Json : nom de la clé dans le JSON.

Deca : décalage en nombre d'heure du max/min. pas de valeur = décalage 0.

Clef table Obs : nom de la clé dans la colonne j (col 1), ou j_agg/j_xtreme (col 3) de la table observation

Clef table agg_XXX : nom de la clé dans la colonne j des tables d'agrégations

Les clés en **bleu** dans la colonne « Clef JSON » sont nécessaires si on désire le calcul de la mesure (donnée élémentaire).

La colonne « **Clef table Obs** » les clés en **rouge** sont les clés créées uniquement si la mesure est fournie. Les clés en **vert** correspondent aux valeurs OMM.

Les clés en **violet** sont présentes dans les tables agg_XXX seulement, et ne doit pas être fournie dans le JSON (ne passe pas l'étape de validation)

Clés JSON.valeurs (type : O ou C)	De ca	Clés obs.j	obs.j_agg & j_xtreme Clés agg_XXX JSON.valeurs/extremes (type : HDMYA)	Commentaire
Température				
out_temp_avg out_temp out_temp_max_time out_temp_max_time out_temp_min out_temp_min_time out_temp_omm		out_temp_s out_temp_d out_temp_max out_temp_max_time out_temp_min out_temp_min_time out_temp_omm_s out_temp_omm_d	out_temp_s out_temp_d out_temp_avg out_temp_max_time out_temp_max_time out_temp_min out_temp_min_time out_temp_omm_s out_temp_omm_d out_temp_omm_avg	Mesure agrégée de la température Mesure instantanée de la température Somme des valeurs agrégées* durées Somme des durées Calculé (seulement dans table) Max soit fourni, soit max des instantanées Date du max Min soit fourni, soit min des instantanées Date du min Somme des valeurs omm* durées omm Somme des durées omm Calculé (seulement dans table)

BD Climato: JSON, base de données et traitements

	-7 -7 +5 +5	out_temp_omm_max out_temp_omm_max_time out_temp_omm_min out_temp_omm_min_time	out_temp_omm_max out_temp_omm_max_time out_temp_omm_min out_temp_omm_min_time	Valeur omm max Date du max Valeur omm min Date du min
windchill windchill_min windchill_min_time		windchill_min windchill_min_time	windchill_min windchill_min_time	Température ressentie. Valeur instantanée préférée seul le min est agrégé. Min soit fourni, soit min des instantanées Date du min
heatindex		heatindex_max heatindex_max_time	heatindex_max heatindex_max_time	Index de chaleur, valeur instantanée préférée, seul le max est agrégé Max soit fourni, soit max des instantanées Date du max
dewpoint dewpoint_max dewpoint_max_time dewpoint_min dewpoint_min_time		dewpoint_max dewpoint_max_time dewpoint_min dewpoint_min_time	dewpoint_max dewpoint_max_time dewpoint_min dewpoint_min_time	Point de rosé, valeur instantanée préférée, seuls les extrêmes sont agrégés Max soit fourni, soit max des instantanées Date du max Min soit fourni, soit min des instantanées Date du min
soiltemp soiltemp_min soiltemp_min_time		soiltemp_min soiltemp_min_time	soiltemp_min soiltemp_min_time	Température du sol à 10 m, valeur instantanée préférée, seul le min est agrégé Min soit fourni, soit min des instantanées Date du min
Type Humidité				
humidity humidity_avg humidity_max		humidity_s humidity_d humidity_max	humidity_s humidity_d humidity_avg humidity_max	Humidité valeur instantanée Moyenne de l'humidité Somme des valeurs agrégées* durées Somme des durées Calculé (seulement dans table) Max soit fourni, soit max des instantanées

BD Climato: JSON, base de données et traitements

humidity_max_time humidity_min humidity_min_time humidity_omm		humidity_max_time humidity_min humidity_min_time humidity_omm_s humidity_omm_d humidity_omm_max humidity_omm_max_time humidity_omm_min humidity_omm_min_time	humidity_max_time humidity_min humidity_min_time humidity_omm_s humidity_omm_d humidity_omm_avg humidity_omm_max humidity_omm_max_time humidity_omm_min humidity_omm_min_time	Date du max Min soit fourni, soit min des instantanées Date du min Somme des valeurs omm* durées omm Somme des durées omm Calculé (seulement dans table) Valeur omm max Date du max Valeur omm min Date du min
Type Pression				
barometer barometer_avg barometer_max barometer_max_time barometer_min barometer_min_time barometer_omm		barometer_s barometer_d barometer_max barometer_max_time barometer_min barometer_min_time barometer_omm_s barometer_omm_d barometer_omm_max barometer_omm_max_time barometer_omm_min barometer_omm_min_time	barometer_s barometer_d barometer_avg barometer_max barometer_max_time barometer_min barometer_min_time barometer_omm_s barometer_omm_d barometer_omm_avg barometer_omm_max barometer_omm_max_time barometer_omm_min barometer_omm_min_time	Pression instantanée Moyenne des pressions (niv. mer) Somme des valeurs agrégées* durées Somme des durées Calculé (seulement dans table) Max soit fourni, soit max des instantanées Date du max Min soit fourni, soit min des instantanées Date du min Somme des valeurs omm* durées omm Somme des durées omm Calculé (seulement dans table) Valeur omm max Date du max Valeur omm min Date du min

BD Climato: JSON, base de données et traitements

pressure pressure_max pressure_max_time pressure_min pressure_min_time		pressure_max pressure_max_time pressure_min pressure_min_time	pressure_max pressure_max_time pressure_min pressure_min_time	Pression instantanée au niveau de la station Max soit fourni, soit max des instantanées Date du max Min soit fourni, soit min des instantanées Date du min
Wind				
wind_inst		wind_inst_max wind_inst_max_time	wind_inst_max wind_inst_max_time	Vent instantané. (peu utilisée) La moyenne n'est jamais calculée, mais peut être fourni dans la clause « aggregations » du Json.
wind_avg wind_dir wind_max wind_max_time wind_max_dir		wind_s wind_d wind_dir_nb wind_dir_sin wind_dir_cos wind_max wind_max_time wind_max_dir	wind_s wind_d wind_avg wind_dir wind_dir_nb wind_dir_sin wind_dir_cos wind_max wind_max_time wind_max_dir	Mesure moyenne sur la duration. Pas de valeur instantanée (wind n'est pas permis !) Somme des sommes * durées Somme des durées Calculé (seulement dans table) Direction du vent moyen - calculé Nb mesure direction vent Somme des sinus des mesures Somme des cosinus des mesures Max soit fourni, soit max des instantanées Date du max Direction du max
wind10 (wind10_avg) wind10_dir		wind10_s wind10_d	wind10_s wind10_d wind10_avg wind10_dir	Vent moyen sur 10mn Direction moyenne des vent 10 mn (pas dispo sur VP2) Somme des wind10 * durée Somme des durées Moyenne des vents 10mn (seulement dans table) Direction moyenne du win10. Calculé

BD Climato: JSON, base de données et traitements

<p>wind10_omm</p>		<p>wind10_dir_nb wind10_dir_sin wind10_dir_cos</p> <p>wind10_omm_s wind10_omm_d</p> <p>wind10_omm_dir_nb wind10_omm_dir_cos wind10_omm_dir_sin</p> <p>wind10_omm_max wind10_omm_max_time wind10_omm_max_dir</p>	<p>wind10_dir_nb wind10_dir_sin wind10_dir_cos</p> <p>wind10_omm_s wind10_omm_d wind10_omm_avg</p> <p>wind10_omm_dir wind10_omm_dir_nb wind10_omm_dir_cos wind10_omm_dir_sin</p> <p>wind10_omm_max wind10_omm_max_time wind10_omm_max_dir</p>	<p>Nombre de mesure Somme des sinus Somme des cosinus</p> <p>Somme des valeurs omm* durées omm Somme des durées omm Moyenne des valeurs omm (seulement dans table)</p> <p>Direction moyenne omm (calculé) Nombre de mesure de direction Somme des cosinus des directions Somme des sinus des directions</p> <p>Valeur omm max Date du max Direction du max omm</p>
<p>gust</p> <p>gust_max gust_max_time gust_max_dir</p>		<p>gust_max gust_max_time gust_max_dir</p>	<p>gust_max gust_max_time gust_max_dir</p>	<p>gust/gust_max est une mesure remontée par la console VP2, qui peut être légèrement différente du wind_max. Cette valeur est stockée dans la base de données pour analyse ultérieure. A n'utiliser que si le wind_max est donné. (cad si on a que le gust/gust_max et pas de wind_max, il faut mettre le gust/gust_max de la station dans wind/wind_max du json).</p>
Type Rain				
<p>rain_sum rain (synonyme)</p> <p>rain_omm</p> <p>rain_omm_max rain_omm_max_time</p>	<p>7 7</p>	<p>rain_s rain_d</p> <p>rain_omm_s rain_omm_d rain_omm_max rain_omm_max_time</p>	<p>rain_s rain_d</p> <p>rain_omm_s rain_omm_d rain_omm_max rain_omm_max_time</p>	<p>Somme de la pluie sur la période.</p> <p>Somme des pluies Somme des durées</p> <p>Valeur OMM Somme des durées OMM max des valeurs OMM Date du max</p>

BD Climato: JSON, base de données et traitements

rain_rate_avg (rain_rate)				Moyenne du rain_rate sur la période synonyme
rain_rate_max rain_rate_max_time		rain_rate_max rain_rate_max_time	rain_rate_max rain_rate_max_time	Rate max Date du max
Type Solar				
uv_indice		uv_indice_max uv_indice_max_time	uv_indice_max uv_indice_max_time	Max des indices
radiation_inst		uv_indice_s uv_indice_d	uv_indice_s uv_indice_d	Somme des radiations
etp_sum (etp_s) etp_d		etp_s etp_d	etp_s etp_d	Fourni dans les « valeurs » d'observation aux heures rondes, avec un etp_d=60
Type Intérieur				
in_temp in_temp_avg		in_temp_d in_temp_s	in_temp_d in_temp_s in_temp_avg	Température intérieure instantanées Moyenne des températures intérieure Somme des durées de temp. Int Somme des temp * durée Moyenne des temp (seulement dans table)
in_temp_max in_temp_max_time in_temp_min in_temp_min_time		in_temp_max in_temp_max_time in_temp_min in_temp_min_time	in_temp_max in_temp_max_time in_temp_min in_temp_min_time	Température intérieure max Date du max Température intérieure min Date du min
in_humidity in_humidity_avg		in_humidity_d in_humidity_s	in_humidity_d in_humidity_s in_humidity_avg	Humidité intérieure instantanées Moyenne des Humidité intérieure Somme des durées Somme des durées * Humidité Moyenne de l'humidité (calculé) (seulement dans table)
in_humidity_max in_humidity_max_time in_humidity_min		in_humidity_max in_humidity_max_time in_humidity_min	in_humidity_max in_humidity_max_time in_humidity_min	Humidité intérieure max Date du max Humidité intérieure min

BD Climato: JSON, base de données et traitements

in_humidity_min_time		in_humidity_min_time	in_humidity_min_time	Date du min
Type Divers				
rx rx_avg		rx_d rx_s	rx_d rx_s rx_avg	Taux réception VP2 instantanées Moyenne des Taux réception Somme des durées Somme des durées * Taux réception Moyenne des taux de réception (calculé) (seulement dans table)
rx_max rx_max_time rx_min rx_min_time		rx_max rx_max_time rx_min rx_min_time	rx_max rx_max_time rx_min rx_min_time	Taux réception VP2 max Date du max Taux réception VP2 min Date du min
voltage voltage_avg		voltage_d voltage_s	voltage_d voltage_s voltage_avg	Voltage VP2 instantané Moyenne du voltage Somme des durées Somme des durées * Voltage Moyenne des voltages (calculé) (seulement dans table)
voltage_max voltage_max_time voltage_min voltage_min_time		voltage_max voltage_max_time voltage_min voltage_min_time	voltage_max voltage_max_time voltage_min voltage_min_time	Voltage max Date du max Voltage min Date du min

BD Climato: JSON, base de données et traitements

ANNEXE C : Unités utilisées

Mesure	Unité	Commentaire
out_temp in_temp heatindex dewpoint soil_temp	°C	degrés Celsius
windchill		sans unité (assimilable degrés celsius)
humidity in_humidity	%	pourcents
barometer pressure	hPa	hectopascals
wind_avg wind10_avg gust_max wind_max	m/s	mètres par seconde
wind_dir gust_dir wind_max_dir	dg	degrés (rose de 360)
rain	mm	millimètres
rain_rate	mm/h	millimètres par heure
radiation radiation_sum	J/s/cm ² J/cm ²	Joules par seconde par cm ² (valeur instantanée) Joules par cm ² (valeur cumulée)
rx	%	pourcentage
Voltage	V	volts

ANNEXE D : Service d'importation de fichiers JSON

Les observations ont une date de fin de période d'observation (stop_dat), ainsi qu'une durée d'observation (duration), le plus souvent de 5 mn, mais pas toujours. Il ne faut pas importer des dates qui se recouvrent.

Toutes les dates/heures sont en heure locale

Le service d'importation de fichier JSON lit chaque mesure, calcule des extrêmes à partir des valeurs si non fournies, calcule les valeurs omm, vérifie s'il existe un extrême plus fort dans la clause « extremes », et stocke dans la colonne j de la table obs les données normalisées, telles que définies dans l'appendice B

Aussi un enregistrement est ajouté dans la table agg_todo, avec un status = 0, qui sera utilisé pour lancer le service de calcul des agrégations. Cet enregistrement contient le delta des valeurs qui sera propagé dans les agrégations.

Ce service est activé toutes les minutes. Il est possible de le contrôler grâce aux commandes :

```
python3 manage.py json --stop : stop le service
python3 manage.py json --start : démarre le service
python3 manage.py json --run : active le service
python3 manage.py json --trace : trace les activations du service
python3 manage.py json --notrace : stop le tracing du service
```

ANNEXE E : Service de calcul d'agrégations

Le service monitor la table `agg_todo`, s'il y a des enregistrements avec un `status = 0` (données à reporter dans les agrégations). Dès qu'un enregistrement est en cours de traitement, il prend un `status = 1`. Quand l'ajout dans les agrégations est terminé, cet enregistrement est supprimé (ou `status = 999` en phase de mise au point). En cas d'erreur, le `status` est mis à 99, et l'ajout de ses données dans les tables d'agrégation est annulé.

Les lignes de la table `agg_todo` avec un `status 1` ou `99` correspondent aux observations non intégrées dans les agrégations. Il est nécessaire alors de corriger la source du problème, puis de remettre le `status = 0`, pour relancer une nouvelle tentative. En cas d'erreur la raison du problème est signalée dans la table `incident`, et dans la colonne `j_error`

`Agg_todo` contient le delta de valeurs à appliquer aux différents niveaux d'agrégation, ainsi que les mesures extrêmes relevés sur la période d'observation. Extrêmes fournis, calculés à partir des données instantanées, ou obtenues à partir d'information de la clause « `extremes` »

Ces delta-valeurs sont passés à chaque niveau des agrégations, au niveau heure, puis jour, puis mois, puis année, puis global. Tant qu'un extrême reste extrême à un niveau d'agrégation, il est passé au niveau supérieur, sinon il est oublié.

En même temps, la durée de la donnée élémentaire est ajoutée dans le champ `duration_sum` de l'agrégation, permettant ainsi de quantifier le nombre de mesures ayant servi pour le calcul de cette agrégation. « `duration_sum` » ne peut pas être supérieur au champ « `duration_max` » qui correspond à la durée de l'agrégation (exemple 60mn pour une agrégation horaire)

Calcul fait suivant le type d'agrégation :

- Avg: il est stocké la somme des moyennes fois la durée, ainsi que la somme des durées. La moyenne est calculée
- Sum : La sommes des valeurs somme est stocké, ainsi que la somme des durées

Pour calculer la moyenne des directions des vents : on utilise la méthode des moyennes circulaires/Moyenne angulaire, tel que définie dans :

https://en.wikipedia.org/wiki/Circular_mean

Dans les agrégations on stocke la somme des sinus, des cosinus, et le nombre de mesures de direction de vent. La moyenne de la direction est ainsi calculée et stockée.

En cas de suppression de données un extrême peut être invalidé, et nécessiter une analyse des extrêmes de niveau inférieur, ou au niveau observation pour une invalidation d'un extrême au niveau horaire. Ces demandes sont stockées dans la table `extreme_todo`.

Ce service est activé toutes les deux minutes. Il est possible de le contrôler grâce aux commandes :

BD Climato: JSON, base de données et traitements

```
python3 manage.py agg --stop : stop le service  
python3 manage.py agg --start : démarre le service  
python3 manage.py agg --run : active le service  
python3 manage.py agg --trace : trace les activations du service  
python3 manage.py agg --notrace : stop le tracing du service
```

ANNEXE F : Commentaires sur calcul OMM

Les agrégations de valeurs OMM sont calculées suivant le type d'agrégation utilisé :

- **avgomm** : On utilise la donnée instantanée à l'heure ronde, qu'on considère pour une durée de 60mn. Heure ronde veut dire que l'heure ronde doit se trouver entre l'heure de début (non incluse), et l'heure de fin (incluse) de la période d'observation. Les valeurs extrêmes de la période sont prises en compte pour calculer les valeurs extrêmes OMM.
- **sumomm** : Toutes les données de la période sont sommées, pour calculer cette somme. Les valeurs extrêmes de la période sont prises en compte pour calculer les valeurs extrême OMM.

Coté calcul c'est surtout le type *avgomm* qui diffère.

Cependant une valeur OMM peut avoir une plage horaire différente :

- Pour le calcul des agrégations journalière et au-dessus (exemple la pluie OMM journalière est calculée à partir des données de 7 :00 à 6 :55 le lendemain)
- Pour le calcul des extrêmes min ou max (par exemple le out_temp_omm min journalier est calculé sur une période de 19 :00 la veille à 18 :55, tandis que le out_temp_omm_max est calculé sur une période allant de 7 :00 à 6 :55 le lendemain).

Pour contrôler cela, il y a trois paramètres dans la définition des valeurs OMM :

- **deca_hour** : gère le décalage de la période de la mesure, et des valeurs extrêmes, sauf notification contraire (min_deca, max_deca)
- **min_deca** : décalage de la période du min
- **max_deca** : décalage de la période du max

Afin d'homogénéiser le calcul des agrégations pour les valeurs OMM, l'agrégation horaire utilisée tient compte de ce décalage. Par exemple la valeur rain_omm du 1^{er} janvier 2022 @ 5h00 sera stockée dans l'agrégation horaire du 31 décembre 2021 @ 22h00. De cette façon cette valeur sera correctement agrégée dans l'agrégation du 21 décembre 2021, puis dans l'agrégation annuelle 2021.

C'est la raison pour laquelle il est recommandé d'utiliser les données OMM qu'à partir des agrégations journalières, et au-dessus, et d'éviter d'utiliser les agrégations horaires pour les données OMM (seulement pour les mesures OMM ayant un décalage)

BD Climato: JSON, base de données et traitements

ANNEXE G : Service de recalcul des extrêmes

- *Les extrêmes WeeWX sont archivés par jour, mois, année seulement. En cas de reprise à partir de l'historique WeeWX, cette limitation va avoir l'impact suivant pour les extrêmes stockés dans la BD Climato :*
 - *Cas de reprise d'historique au niveau observations : Le Json va contenir les valeurs observées, ainsi que les extrêmes journaliers du jour concerné. Les extrêmes du jour seront mis à jour dans l'observation correspondante, et dans l'agrégation horaire correspondante, donc pas de perte d'information pour l'extrême journalier (et niveaux supérieurs). Cependant pour les autres observations, et agrégations horaires, les extrêmes seront recalculés à partir des valeurs instantanées.*
 - *Reprise d'historique au niveau agrégation horaire : Le Json va contenir les valeurs agrégées au niveau horaire, ainsi que les extrêmes journaliers du jour concerné. Les extrêmes du jour seront mis à jour dans l'agrégation horaire correspondante, sans de perte d'information pour l'extrême journalier (et niveaux supérieurs). Cependant pour les autres agrégations horaires, les extrêmes seront recalculés à partir des valeurs instantanées.*

Reprise d'historique au niveau agrégation journalière ou au-dessus : Pas de perte d'information

⇒ *Nouveau service à écrire pour recalculer les extrêmes suite à une suppression d'une observation ayant un extrême à un niveau d'agrégation quelconque.*

Ce service est activé toutes les minutes. Il est possible de le contrôler grâce aux commandes :

```
python3 manage.py maxmin --stop : stop le service
python3 manage.py maxmin --start : démarre le service
python3 manage.py maxmin --run : active le service
python3 manage.py maxmin --trace : trace les activations du service
python3 manage.py maxmin --notrace : stop le tracing du service
```

ANNEXE H : changement du format Json

- Suppression de la clause « aggregations »
- Nouvelle clause « extremes » au niveau data, simple Json, pas un tableau
- « version » dans « info »
- Nouveau champs : « json_type » dans « info » :
 - O -> Fichier contenant des observations
 - C -> Régénération d'observations, avec extrêmes journaliers
 - H -> Fichier contenant des aggregations horaires, avec des extrêmes journaliers
 - D -> Fichier contenant des agrégations et extrêmes journaliers
 - M -> Fichier contenant des agrégations et extrêmes mensuels
 - Y -> Fichier contenant des agrégations et extrêmes annuels
 - A -> Fichier contenant des agrégations et extrêmes globaux
- La clause « current » est remplacée par « valeurs »
- Dans « valeurs » il est possible d'ajouter une « clef_d » si la duration de la mesure clef est différente de la duration du « data » (comme pour etp)
- Simplification des dates dans le JSON :
 - Observation -> Seulement stop_dat et duration
 - Données agrégées par WeeWX : -> Seulement start_dat

ANNEXE I : Exemple JSON données observation

Un exemple des JSON générés par l'utilitaire *wee_json* se trouve dans le document *wee_json.doc*

```
[
  {
    "meteor": "BBF015",
    "info": {
      "version": 2,
      "json_type": "0"
    },
    "data": [
      {
        "stop_dat": "2022-02-18T00:05:00",
        "duration": 5,

        "valeurs": {
          "barometer_avg": 1008.3047501,
          "barometer_d": 5,
          "barometer_max_time": "2022-02-18T00:03:02",
          "barometer_max": 1008.36480205,
          "barometer_min_time": "2022-02-18T00:03:02",
          "barometer_min": 1008.22934652,
          "barometer": 1008.22934652,

          "dewpoint_d": 5,
          "dewpoint_max_time": "2022-02-18T00:03:12",
          "dewpoint_max": 23.8217176455,
          "dewpoint_min_time": "2022-02-18T00:03:02",
          "dewpoint_min": 23.7669838316,
          "dewpoint": 23.8217176455,

          "etp_s": 0.254,
          "etp_d": 60,

          "heatindex_d": 5,
          "heatindex_max_time": "2022-02-18T00:03:12",
          "heatindex_max": 25.7777777778,
          "heatindex": 25.7777777778,

          "humidity_avg": 89,
          "humidity_d": 5,
          "humidity_max_time": "2022-02-18T00:03:02",
          "humidity_max": 89,
          "humidity_min_time": "2022-02-18T00:03:02",
          "humidity_min": 89,
          "humidity": 89,

          "in_humidity_avg": 47.4,
```

```
"in_humidity_d": d,
"in_humidity_max_time": "2022-02-18T00:03:00",
"in_humidity_max": 49,
"in_humidity_min_time": "2022-02-18T00:03:02",
"in_humidity_min": 46,
"in_humidity": 49,

"in_temp_avg": 25.0222222222,
"in_temp_d": 5,
"in_temp_max_time": "2022-02-18T00:03:02",
"in_temp_max": 25.0555555556,
"in_temp_min_time": "2022-02-18T00:03:00",
"in_temp_min": 25,
"in_temp": 25.0555555556,

"out_temp_avg": 25.7644444444,
"out_temp_d": 5,
"out_temp_max_time": "2022-02-18T00:03:12",
"out_temp_max": 25.7777777778,
"out_temp_min_time": "2022-02-18T00:03:02",
"out_temp_min": 25.7222222222,
"out_temp": 25.7777777778,

"pressure_max_time": "2022-02-18T00:03:12",
"pressure_max": 1003.86606438,
"pressure_min_time": "2022-02-18T00:03:02",
"pressure_min": 1003.73121318,

"radiation_d": 5,
"radiation_max_time": "2022-02-18T00:03:02",
"radiation_max": 0,
"radiation_min_time": "2022-02-18T00:03:02",
"radiation_min": 0,
"radiation": 0,

"rain_rate_max_time": "2022-02-18T00:03:02",
"rain_rate_max": 0,

"rain": 0,
"rain_d": 5,

"rx": 97.375,
"rx_d": 5,

"uv_indice_avg": 0,
"uv_indice_d": 5,
"uv_indice_max_time": "2022-02-18T00:03:02",
"uv_indice_max": 0,
"uv_indice": 0,

"voltage_avg": 4.77,
```



```
"voltage_d": 5,  
"voltage_min_time": "2022-02-18T00:03:02",  
"voltage_min": 4.77,  
"voltage": 4.77,  
  
"wind_avg": 2.68224666883,  
"wind_d": 5,  
"wind_dir": 135,  
"wind_max_dir": 180,  
"wind_max_time": "2022-02-18T00:03:02",  
"wind_max": 5.36449333766,  
  
"wind10_avg": 2.68224666883,  
"wind10_dir": 135,  
"wind10_d": 5,  
  
"windchill_avg": 25.76444444444,  
"windchill_d": 5,  
"windchill_min_time": "2022-02-18T00:03:02",  
"windchill_min": 25.72222222222,  
"windchill": 25.7777777778  
}  
}  
]  
}  
]
```