## Historique des versions

v0.1 : Keryl – premer draft

v0.2 : Nicolas – Proposition de modifications, à valider par Keryl. Ajout des actions à effectuer en cas de rejet de contrôle, et codes qualité.

Afin d’assurer la robustesse et pertinence de la BDD, il sera nécessaire de définir différents contrôles des données qui y sont insérées. Ces contrôles peuvent être de plusieurs types :

* **Filtres en entrée :** Un premier niveau de filtre s’assurera de la pertinence des données, notamment vis-à-vis des plages de mesure de chaque capteur. Ces données, qualifiées d’absurdes, s’avèreront donc fausses quoi qu’il arrive.
* **Trigger :** Le trigger est un niveau de cohérences inter-paramètres simples et aura une action de modification dans la base. Tout simplement, si une température minimale est manquante, le paramètre « heure de la température minimale » sera automatiquement mise à manquante.
* **Contrôles qualité :** Ce contrôle qualité sera le dernier niveau de contrôle et sera effectué quotidiennement. Celui-ci testera la cohérence temporelle ou inter-paramètres de manière plus élaborée que celui des filtres en entrée. À terme, un contrôle spatial pourra aussi être intégré lorsque le maillage de stations sur l’île sera plus fin.

Chacun de ces contrôles qui s’avéreront non conformes devront aboutir **sur une action ciblée, ainsi que des mécanismes d’agrégations et de calculs des extrêmes**. Les actions à effectuer sont décrits en fin de document, un fois l’ensemble des controles décrits.

## Filtres en entrée

### Temporalité du contrôle

Ce contrôle doit être fait quelque soit la façon dont les données sont modifiées (insertion/ suppression) dans la base de données.

### Champ d’action du contrôle

Le contrôle se fera préférentiellement **par paramètres** et non pas par type de paramètres. Les paramètres peuvent appartenir au même type mais ne pas être soumises à la même plage de variation.

Ex :

* La pression au niveau de la Mer et la pression de la station sont tous les 2 des paramètres de type « Pression » mais ces 2 paramètres ne vont pas être soumis au même filtre.
* Temp et Humidity sont tous les 2 de type « Temp » mais ne possèdent pas la même unité.

Un filtre pourra être différent suivant le niveau de l’agrégation (la largeur d’intervalle pour un cumul horaire ne sera pas le même qu’un cumul quotidien). Ex : Durée d’insolation sur 1h : Entre 0 et 60min ; Durée d’insolation sur 24h : Entre 0 et 960 min (18h). Il existe aussi des fitres qui sont les mêmes quelque soit le niveau d’agrégation (comme la température). Ces contraintes de filtrage existent surtout pour les agrégations hour et day.

Enfin ces fitres devront aussi être executés en cas d’insertion de valeurs pré-agrégées, sans données élémentaires, surtout au niveau heure et journalier.

On peut accepter que les données pre-agrégées au niveau mois/an doivent êtreverifiées manuellement avant insertion (cas pour l’historique d’une seule station du réseau)

### Contrôle d’un paramètre

**Table agg\_hour :**

Nomenclature :

**FSAH\_X : Filtre Seuil Absolu dans agg\_hour pour le paramètre X**

|  |  |  |
| --- | --- | --- |
| **Paramètre** | **Nom du paramètre dans la table** | **Filtre** |
| **FSAH\_T** : Température |  | Entre -40/50°C |
| **FSAH\_TN** : Température minimale |  |
| **FSAH\_TX** : Température minimale |  |
| **FSAH\_TD** : Température point de rosée |  |
| **FSAH\_T10 :** Température à 10 cm du sol |  | Entre – 15°C et +50°C |
| **FSAH\_T20 :** Température à 20 cm du sol |  | Entre -10°C et +45°C |
| **FSAH\_T50 :** Température à 50 cm du sol |  | Entre -5°C et +40°C |
| **FSAH\_RR1 :** Cumul de précipitations sur 1h |  | Entre 0 et 400mm |
| **FSAH\_INS :** Durée horaire de l’insolation |  | Entre 0 et 60min |
| **FSAH\_GLO :** Rayonnement horaire global |  | Entre 0 et 500 J/cm2 |
| **FSAH\_INFRAR :** Rayonnement infra-rouge horaire |  | Entre 0 et 300 J/cm2 |
| **FSAH\_UVINDICE :** Indice UV |  | Entre 0 et 20 |
| **FSAH\_DD :** Direction du vent vent moyen |  | Entre 0 et 360° |
| **FSAH\_DXI :** Direction du vent de la rafale max |  |
| **FSAH\_FF :** Force du vent moyen |  | Entre 0 et 70 m/s |
| **FSAH\_FXI :** Force des rafales |  | Entre 0 et 100 m/s |
| **FASH\_U :** Humidité |  | Entre 0 et 110% |
| **FASH\_UN :** Humidité minimale |  |
| **FASH\_UX :** Humidité maximale |  |
| **FSAH\_PMER :** Pression au niveau de la mer |  | Entre 850 et 1060 hPa |
| **FSAH\_PSTAT :** Pression station |  | Entre 600 et 1060 hPa |
| **FSAH\_H{X}** : Heure des extrêmes, avec X le paramètre |  | Doit être compris dans [H-60mn,H] |

**Table agg\_day :**

Nomenclature :

**FSAQ\_X : Filtre Seuil Absolu dans agg\_day pour le paramètre X**

|  |  |  |
| --- | --- | --- |
| **Paramètre** | **Nom du paramètre dans la table** | **Filtre** |
| **FSAQ\_TN** : Température minimale |  | Entre -40°C et +50°C |
| **FSAQ\_TX** : Température minimale |  |
| **FSAQ\_RR24 :** Cumul de précipitations sur 24h |  | Entre 0 et 3000 mm |
| **FSAQ\_INS :** Durée quotidienne de l’insolation |  | Entre 0 et 960min (<18h) |
| **FSAQ\_FF :** Force du vent moyen |  | Entre 0 et 70 m/s |
| **FSAQ\_FXI :** Force des rafales |  | Entre 0 et 100 m/s |
| **FASQ\_UM:** Humidité moyenne |  | Entre 0 et 110% |
| **FASQ\_UN :** Humidité minimale |  |
| **FASQ\_UX :** Humidité maximale |  |
| **FSAQ\_PMERM :** Pression moyenne au niveau de la mer |  | Entre 850 et 1060 hPa |
| **FSAQ\_GLOT :** Rayonnement quotidien |  | Entre 0 et 5000 J/cm2 |
| **FSAQ\_H{X}** : Heure des extrêmes, avec X le paramètre |  | Doit être compris dans [00,23h59] |

### Contrôle inter-paramètres

* Peut on resumer ce paragraphe par :

Il faut vérifier que les valeurs minimumes soient bien inférieures aux valeurs maximum, et que la pression station est inférieure à la pression mer ?

En doublon de ces filtres viendront s’insérer des contraintes inter-paramètres.

Nomenclature :

* **FCAH\_X : Filtre Contrainte Absolue Horaire**
* **FCAQ\_X : Filtre Contrainte Absolue Quotidienne**

**Table agg\_hour :**

|  |  |  |
| --- | --- | --- |
| **Paramètre** | **Nom du paramètre dans la table** | **Filtre** |
| **FCAHT\_TN** | **T(H) >= TN(H)** | Température supérieure à la température minimale |
| **FCAHT\_TX** | **T(H) <= TX(H)** | Température inférieure à la température maximale |
| **FCAHU\_UN** | **U(H) >= UN(H)** | Humidité supérieure à l’humidité minimale |
| **FCAHU\_UX** | **U(H) <= UX(H)** | Humidité inférieure à l’humidité maximale |
| **FCAHPMER\_PSTAT** | **PMER(H) >= PSTAT(H)** | Pression mer supérieure à la pression station |

**Table agg\_day :**

|  |  |  |
| --- | --- | --- |
| **Paramètre** | **Nom du paramètre dans la table** | **Filtre** |
| **FCAQTX\_TN** | **TX(Q) >= TN(Q)** | Température maximale supérieure à la température minimale |
| **FCAQUX\_UN** | **UX(Q) >= UN(Q)** | Humidité maximale supérieure à l’humidité minimale |

## Triggers simples

Il s’agit de contrôle d’intégrité et de mise en cohérence automatique entre plusieurs paramètres.

### Temporalité du contrôle

Ces controles devront être effectués à chaque modification (au sens large) des données

### Champ d’action du contrôle

Ce contrôle se fera sur **l’ensemble des tables de la BDD**. En effet, toute table de la BDD peut être modifiée et la modification de la valeur d’un paramètre en « valeur manquante » doit pouvoir être répercutée sur les paramètres connexes. Par ailleurs, la modification d’une valeur doit pouvoir modifier la valeur des paramètres calculés/agrégés (ex : modification de rain\_sum1h entrainera la modification de rain\_sum3h).

### Triggers de mise à manquant

**Table agg\_hour :**

|  |  |
| --- | --- |
| **Si le paramètre est manquant** | **Les paramètres suivants sont mis à manquant :** |
| Vitesse du vent | Direction du vent + heures associées |
| Rafales | Direction de la rafale + heures associées |
| Température minimale | Heure de la température minimale |
| Température maximale | Heure de la température maximale |
| Température | Température minimale, Heure de la TN, Température maximale, Heure de la TX |
| Humidité minimale | Heure de l’humidité minimale |
| Humidité maximale | Heure de l’humidité maximale |
| Humidité | Humidité minimale, Heure de l’humidité minimale, Humidité maximale, Heure de l’humidité maximale |
| Pression au niveau de la mer | Pression au niveau de la mer minimale et heures associées |

**Table agg\_day :**

|  |  |
| --- | --- |
| **Si le paramètre est manquant** | **Les paramètres suivants sont mis à manquant :** |
| Rafales max | Direction de la rafale max + heures associées |
| Température minimale | Heure de la température minimale |
| Température maximale | Heure de la température maximale |
| Humidité minimale | Heure de l’humidité minimale |
| Humidité maximale | Heure de l’humidité maximale |
| Pression au niveau de la mer | Pression au niveau de la mer minimale et heures associées |

## Contrôle qualité

## Temporalité du contrôle et bases concernées

Ces contrôles peuvent être lancés :

* Quotidiennement sur les données temps réel inférieures à 60j pour les tables ‘données élémentaires’, H et Q.
* À la demande sur des données plus anciennes (ex : intégration d’archives) sur une ou plusieurs tables ‘données élémentaires’, H et Q.
* Lors d’une modification (ce n’est pas pareil qu’à la demande ?)

### Codes qualités

Question : quand une donnée a été modifiee par un filtre/trigger, faut il garder cette information dans les agregations superieure ? Si oui, est-ce qu’un simple compteur de modifications faites par les controles serait un meilleur critere, tout en gardant un historique dans une autre table des modifications faite automatiquement par les filtres ?

Aussi si une donnee a été reperee mais non modifiée, est ce qu’un cumul par agregation d’un compteur des incidents serait un bon critere, avec aussi un historique de ces incidents dans une table ?

On pourrait ajouter un champ contrôle\_qualite, boolean qui dirait si la donnee a été controlee.

Plusieurs niveaux de codes qualité peuvent être considérés :

* QC=1 Non contrôlé (cas contrôle\_qualite = false)
* QC=2 Contrôlé et non douteux (cas contrôle\_qualite = true & cpt\_incident = 0)
* QC=3 Contrôlé et douteux (cas contrôle\_qualite = true & cpt\_incident > 0)
* ( ??) QC = 4 Donnée filtrée (cas contrôle\_qualite = true & cpt\_modif > 0)
* ( ??) QC = 5 Trigger de mise à manquant -> cela sera automatiquement implementé dans le code, donc ne pourra pas arriver…

Je propose de supprimer les 2 paragraphes suivants :

Cette information peut être directement attachée au paramètre de la table contrôlée par le biais d’une colonne supplémentaire : quality\_code (QC)

Ces contrôles n’éliminent pas de données, mais génèrent des codes qualité douteux ou non. Ainsi, toute donnée déclarée douteuse reste dans la table contrôlée mais génère un code qualité QC=3. À la suite de contrôles, le gestionnaire de la base peut valider la donnée douteuse (QC passe de 3 à 2) ou la modifier.

Et de les remplacer par :

Il sera possible de détecter les agrégations douteuses, ainsi que le nombre de modifications automatiques ou incidents de la période couvrant l’agrégation

NB : il y a un seul problème : le compteur va avoir un poid différent suivant le niveau ou les incidents sont repérés : par exemple si 12 mesures sont erronées (sur 12 \* 5 mn = 1 heure), il y aura un compteur modification valant 12 au niveau mois, ce qui est different de 12 incidents sur des jours differents. Il est proposé de ne compter les modifications/incidents qu’une fois lors de l’agrégation au niveau supérieur.

### Contrôles dans agg\_hour

**Contrôle simple :**

|  |  |  |
| --- | --- | --- |
| **Paramètre** | **Nom du paramètre en base** | **Contrôle associé** |
| Durée d’insolation |  | < à 3mn (offset éventuel du capteur) entre 21UTC et 3UTC |
| Rayonnement global horaire |  | < = à 10 J/cm2 (offset éventuel du capteur) entre 21UTC et 3UTC |
| Heures des extrêmes |  | Les heures des extrêmes horaires doivent appartenir à l’intervalle [h-59mn, h] |
| Rayonnement UV |  | <= 0.02J/cm2 entre 21UTC et 3UTC |
| Précipitations |  | Cumul de pluie en 1H <= seuil à définir (général ou par stations) |

**Principe du contrôle temporel :**

* P(h) la valeur du paramètre p à l’observation h
* P(h-1) et p(h+1) les valeurs de ce paramètre aux observations des heures précédentes et suivantes. On calcule une valeur estimée du paramètre p\*(h) = ((p(h-1) +p(h+1))/2

L’écart en valeur absolue entre valeur estimée et valeur observée | p\*(h) – p(h) | doit être inférieur à un seuil fixe pour chaque paramètre.

|  |  |  |
| --- | --- | --- |
| **Paramètre** | **Nom du paramètre en base** | **Contrôle associé** |
| Température |  | Seuil fixe : écart < 10°C |
| Températures extrêmes de l’heure |  | Seuil fixe : écart < 10°C |
| Pression mer |  | Seuil fixe : écart <5hPa |
| Pression de la mer minimale |  | Seuil fixe : écart <5hPa |
| Vent moyen |  | Seuil fixe : écart < 10m/s |
| Rafales |  | Seuil fixe : écart < 15m/s |
| Humidité |  | Seuil fixe : écart < 50% |
| Humidités extrêmes de l’heure |  | Seuil fixe : écart < 50% |

**Contrôles de capteurs bloqués :**

* On ne peut pas avoir plus de N valeurs consécutives (horaires) égales.
* N est dépendant du paramètre

|  |  |  |
| --- | --- | --- |
| **Capteur bloqué** | **Paramètre contrôlé** | **N** |
| Anémomètre bloqué | Rafales  Vent moyen | N = 12  N = 24 |
| Girouette bloquée | Direction du vent | N=12 |
| Thermomètre bloqué | Température | N=12 |

**Contrôle de cohérence inter-paramètres**

|  |  |
| --- | --- |
| **Paramètres contrôlés** | **Contrôles** |
| Températures extrêmes | Erreur si TX(H) – TN(H) > 10° |
| Vent / Rafales et direction | DD = 0 => FF=0 |
| Vent et rafales | FF(H) – 1.1m/s <= Rafales(H) |
| Pression mer | Erreur si PMERMIN > PMER |
| Pression mer | Erreur si :  Abs(PMER-PMERMIN) > 5hPa |

Ce qui serait intéressant serait d’avoir une table spécifique aux contrôles douteux avec un nom spécifique pour chaque contrôle. Cela permettrait de plus facilement connaître l’origine du problème. Ex : Si FF est mis à douteux, on ne saura pas si cela vient d’une cohérence inter-paramètres ou d’une variation temporelle douteuse.

## Actions suite à modification

On va avoir deux types d’actions :

* Annulation de la donnée. Dans ce cas la donnée, et les données liées (voir trigger plus après dans ce document) seront mises à null, et traitée comme une valeur non fournie aux niveaux supérieurs d’agrégation. Aussi le nombre de modification sera incrementée. Une entrée sera créée dans une table historique « fitration », la date du jour, le no du poste, le niveau d’agregation, le nom du parametre, la valeur rejetée, et le no de la règle rejetant cette valeur.
* Détection d’une valeur inconsistante : Cette valeur sera gardée dans la base de données ( ?? Peut-on faire autre chose ??). Dans ce cas elle sera utilisée dans les calculs d’agrégation (il faudra voir pour les valeurs extremes…). Aussi le nombre d’incident sera incrementé. Une entrée sera créée dans une table « historique » incident, la date du jour, le no du poste, le niveau d’agregation, le nom du parametre, la valeur rejetée, et le no de la règle rejetant cette valeur.