RX-03

Un indicateur officiel donnait la même valeur à deux situations différentes

Retour d'expérience. Une colonne d'un fichier public, reprise telle quelle, portait une butée invisible.

Famille	*Généricité*	*Statut*
Retours d'expérience	GÉNÉRIQUE	EXPLORÉ — production du corpus, non validée par une décision
Écrit le	*Licence*	
15 septembre 2026	CC BY-SA 4.0	

L'attendu

Publier l'humidité des sols au Domaine du Mons depuis 1958, à partir de la réanalyse SIM2 de Météo-France. Le fichier fournit deux colonnes : `SWI`, l'humidité du sol ramenée à sa réserve utile, et `SSWI_10J`, cette même humidité ramenée à la normale du jour, en écarts-types. La seconde est précisément l'indicateur sur lequel les services de l'État constatent une sécheresse des sols, au seuil de −1,5. Il n'y avait qu'à la lire.

Le déroulé

Le tableau des épisodes de sécheresse les plus longs, une fois affiché, portait une ligne étrange : décembre 1985, **−8,2 écarts-types**.

Un écart de huit sigmas n'existe pas dans une série de soixante-dix ans. Sous hypothèse normale, c'est une chance sur 10^{15}. Vérification dans la série brute :

Date	SWI	SSWI_10J
16 décembre 1978	0,73	**−8,169**
16 décembre 1985	0,569	**−8,169**

Deux états de sol nettement différents, **le même chiffre à la troisième décimale**. Ce n'est pas une valeur extrême, c'est une butée : l'indicateur sature, et au-delà d'un certain point il cesse de distinguer quoi que ce soit.

La raison est structurelle. En décembre, sur ce point, le sol est presque toujours à sa capacité au champ. La distribution de référence du jour y est étroite et bornée d'un côté. Toute standardisation appliquée à une distribution qui n'a presque pas de largeur produit des écarts-types énormes, puis butte sur la limite numérique de l'ajustement.

L'effet sur les résultats publiés

L'indicateur alimentait trois choses : le compte annuel de jours de sécheresse, la plus longue suite de l'année, et la liste des épisodes. Deux des six épisodes les plus longs étaient des épisodes d'hiver fabriqués par la butée. Le classement des étés, lui, reposait sur le SWI brut et n'était pas touché — c'est ce qui a sauvé la conclusion principale.

Remplacement : l'écart est recalculé par **rang normalisé**. La valeur du jour est située dans l'échantillon de toutes les journées de 1958 à 2020 situées à moins de dix jours de la même date dans l'année — environ 1 300 valeurs — et ce rang est converti en écart-type. L'indicateur est alors **borné par l'échantillon lui-même** : avec 1 300 valeurs, il ne peut pas dépasser ±3,4.

	`SSWI_10J` repris	Rang normalisé
Valeur extrême de la série	−8,17	−3,37
Valeurs identiques pour des sols différents	oui	non
Épisodes d'hiver aberrants	2 sur 6	0
Classement des étés les plus secs	inchangé	inchangé

Et la valeur devient vérifiable. « −3,4 σ le 14 septembre 2026 » se traduit en une phrase qu'on peut contrôler à la main : *aucune des 1 323 journées de référence de la mi-septembre n'a été plus sèche*. C'est exact ; la plus sèche jusqu'ici était le 21 septembre 2019, à 0,193 contre 0,167.

Les leçons

1. **Une valeur impossible est un cadeau ; une valeur presque plausible ne l'est pas.** −8,2 sautait aux yeux. Si la butée avait été à −3,5, rien ne l'aurait signalée, et deux

faux épisodes seraient restés dans le tableau. Le test qui a tranché n'est pas « la valeur est-elle grande ? » mais **« deux entrées différentes donnent-elles la même sortie ? »** — une question qu'on peut poser à n'importe quel indicateur, en triant la colonne et en cherchant les doublons exacts.

2. **Un indicateur dérivé se vérifie contre la grandeur dont il dérive.** Le SSWI se contrôle contre le SWI ; c'est en les affichant côte à côte que l'anomalie est apparue. Un indicateur publié sans la grandeur brute qui le nourrit n'est pas contrôlable.

3. **Officiel ne veut pas dire sans butée.** La colonne vient de Météo-France, elle est juste pour l'usage auquel elle est destinée — le suivi de la sécheresse en saison végétative, où la distribution a de la largeur. Elle est reprise ici hors de cet usage, sur l'année entière. **Le défaut n'est pas dans la source, il est dans la reprise.**

4. **Préférer une statistique bornée par les données à une statistique bornée par un modèle.** Un rang ne peut pas sortir de l'échantillon. Un ajustement paramétrique, si. À qualité d'information égale, la première est plus sûre en bout de distribution — et elle s'explique en une phrase à quelqu'un qui n'est pas statisticien.

Les changements

- Toute colonne dérivée reprise d'une source extérieure est **testée pour la saturation** avant usage : tri de la colonne, recherche de valeurs identiques à la précision publiée sur des entrées différentes, et examen des extrêmes au regard de la taille de la série.
- Les indicateurs standardisés sont **recalculés à partir de la grandeur brute** quand la série de référence le permet, par une méthode bornée par l'échantillon, et la méthode est publiée avec la donnée.
- Une grandeur publiée s'accompagne de la grandeur brute dont elle dérive, dans le même fichier.
- Toute valeur en écarts-types affichée sur le site est traduisible en une phrase de rang vérifiable à la main.

Portée

Générique. La butée guette tout indice standardisé appliqué hors de sa saison ou de son domaine : SPI de précipitations sur un mois sec, z-scores de séries bornées, indices de forme sur des distributions à queue courte, scores normalisés d'un petit échantillon.

Elle guette aussi, hors du climat, tout score composite plafonné en amont sans que le plafond soit documenté.

Une valeur en butée ne se signale pas comme erreur. Elle se signale comme record.

Fiches liées

RX-01 · RX-02 · PA-02

CoMons, Collectif du Mons — RX-03, Un indicateur officiel donnait la même valeur à deux situations différentes. Fiche du Grain, CC BY-SA 4.0. Version en ligne : comons.actitude.org/grain-rx-03-indicateur-en-butee.html