

Annexe 4

Changement d'échelle

Ecoscore

Selon les indications de la documentation de l'Ecoscore¹², trois indications sont données

1. « Dans le cadre de l'Eco-score, le score PEF est converti sur une échelle de 0 à 100 afin de faciliter sa lecture. »
2. « Le score est borné dans l'intervalle [0, 100] »
3. « La formule suit une courbe logarithmique (non linéaire) où l'impact sur l'environnement est doublé tous les 20 points. »

Soit x le score PEF (mPt/kg), avec x_{min} la valeur minimale et x_{max} la valeur maximale. Avec l'affirmation 3, il faut donc utiliser une formule avec un logarithme base 2 et un facteur de 20

$$-\frac{\ln(x)}{\ln(2)} \times 20$$

sachant que l'ordre est inverse entre le score PEF (plus la valeur est grande, plus l'impact est important) et l'Eco-score (100 est pour le produit le plus vertueux), ce qui explique le signe négatif.

En lien avec l'affirmation 2, il faut ensuite borner les valeurs par une borne inférieure (x_{lb}) et supérieure (x_{ub}). Nous avons l'impact d'un produit avec un Eco-score de 0 (le plus mauvais) qui est 32 fois (2^5 car il y a 5 fois 20 dans 100) plus important que le produit avec un Eco-score de 100 (le meilleur), donc

$$x_{ub} = 32 \times x_{lb}$$

Il y a trois grandes possibilités pour faire cela :

- On ne peut pas être plus de 32 fois pire que le meilleur : $x_{lb} = x_{min}$, $x_{ub} = 32 \times x_{min}$
- On ne peut pas être plus de 32 fois meilleur que le pire : $x_{lb} = \frac{1}{32} \times x_{max}$, $x_{ub} = x_{max}$
- On enlève les valeurs extrêmes d'une manière symétrique.

L'avantage de cette dernière possibilité est de ne pas dépendre d'une seule valeur (x_{min} ou x_{max}) mais de la distribution d'une série de données. C'est une approche classique en traitement de données, que l'on va garder ici :

$$\begin{aligned}x_{lb} &= x_{P\%} \\x_{ub} &= x_{100-P\%}\end{aligned}$$

avec $P\%$ le percentile qui permet d'avoir $x_{100-P\%} = 32 \times x_{P\%}$. Dans le cas des données agribalyse $P\% \approx 6,4\%$ ce qui signifie que près de 13% des produits agribalyse sont en dehors des bornes de l'intervalle considéré. C'est une proportion importante.

On peut ensuite facilement mettre l'échelle à la valeur de 0 pour $x_{93,6\%}$

$$-\frac{\ln(x)}{\ln(2)} \times 20 + \frac{\ln(x_{93,6\%})}{\ln(2)} \times 20$$

¹² Extraits de <https://docs.score-environnemental.com/methodologie-recette/fonctionnement-general-recette> accédée le 27-01-2021

Ce qui peut s'écrire plus simplement, en limitant les valeurs à l'intervalle [0, 100]

$$\begin{aligned} \{x < x_{6,4\%} &\rightarrow Eco - score = 100 \\ x_{6,4\%} \leq x \leq x_{93,6\%} &\rightarrow Eco - score = \frac{\ln(x_{93,6\%}) - \ln(x)}{\ln(2)} \times 20 \\ x > x_{93,6\%} &\rightarrow Eco - score = 0 \end{aligned}$$

Comparaison

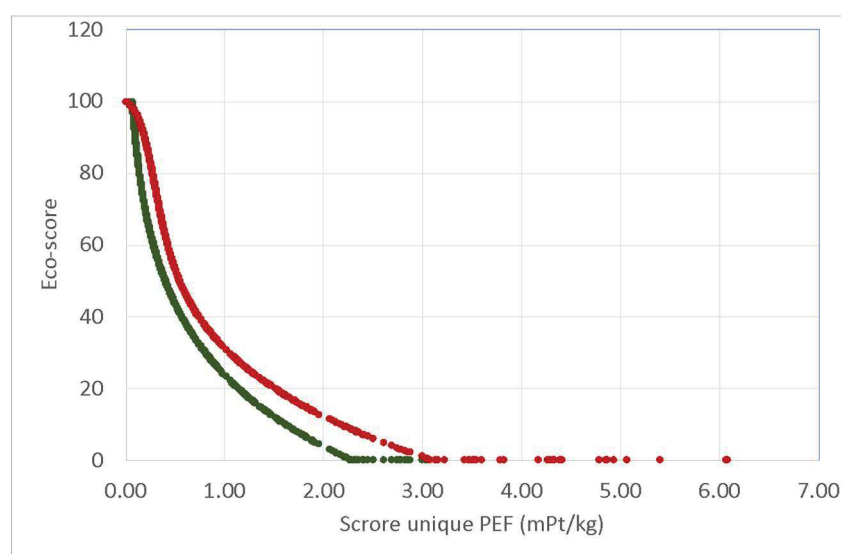
La formule proposée dans la documentation de l'Ecoscore est la suivante (est présentée ici la seconde version, disponible en mai 2021) :

$$100 - \frac{\ln(10x + 1)}{\ln(2 + \frac{1}{100x^4})} \times 20$$

avec x en $\mu\text{Pt}/100\text{g}$. Comme le montre la figure ci-dessous, l'Eco-score obtenu selon les affirmations du texte et celui selon l'équation proposée sur la page web ne donne pas le même résultat. Entre les deux alternatives, la lettre (A,B,C,D ou E) qui est attribuée au produit est différente pour 45% des produits.

Pour avoir quelque chose de similaire sur les Eco-score entre 0 et 50 avec la première version proposée, il ne faut pas utiliser comme borne dans l'équation la valeur $x_{93,6\%} = 2,27 \text{ mPt/kg}$, mais une valeur proche de $3,1 \text{ mPt/kg}$. Dans cette configuration, il n'y aurait que 15% des produits qui changent de lettre. Il n'y a cependant aucune justification envisageable pour ce $3,1 \text{ mPt/kg}$. Cela voudrait dire que l'on considère en dehors des bornes, 3,1% des valeurs pour les scores PEF les plus élevés et 9,9% des valeurs pour les scores PEF les plus bas. Il faudrait un argumentaire pour justifier cette asymétrie. La forme de la relation entre 50 et 100 ne correspond pas à une relation logarithmique qui double tous les 20 (même si elle présente l'avantage de tendre vers la valeur zéro).

Une autre approche serait de dire que le meilleur peut être 128 fois mieux que le pire et ($x_{ub} = 128 \times x_{lb}$), de remplacer les 5 niveaux [A, B, C, D, E] par 7 : [A++, A+, A, B, C, D, E] (sachant que l'affichage pourrait ne pas différencier A++ , A+ et A et seulement afficher A, avec une argumentation à définir...) Dans cette configuration $P\% \approx 1,7 \%$ et $x_{98,3\%} \approx 3,54 \text{ mPt/kg}$. La lettre de l'Eco-score change pour 15% des produits.



a)

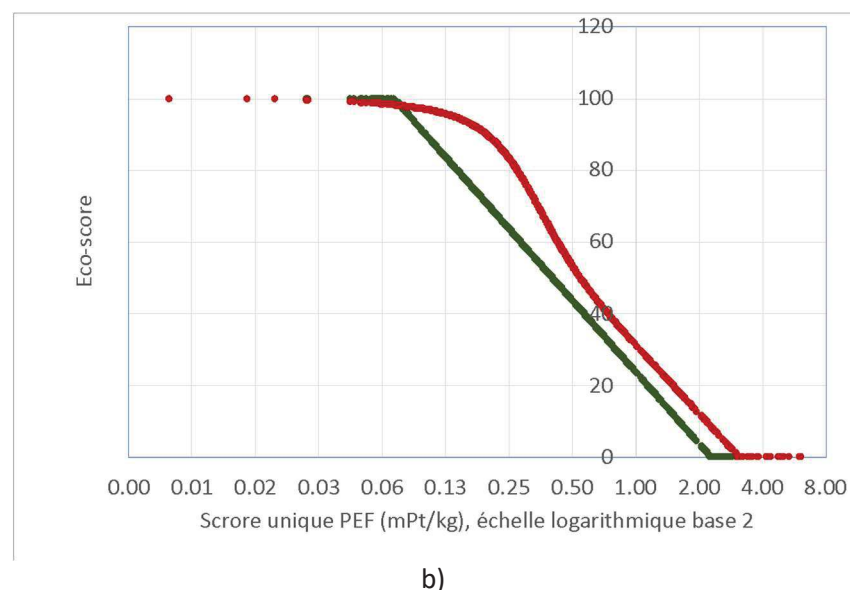


Figure B. Eco-score en fonction des valeurs de score unique PEF pour les 2480 données Agribalyse : en vert, avec la formule construite selon les indications du texte, en rouge selon l'équation proposée sur la page web. a) échelle normale pour l'axe des abscisses, b) échelle logarithmique base 2 pour l'axe des abscisses.

Changement d'équation pour les boissons

La formule diffère pour les boissons. Pour celles-ci, la formule est $-36 * \ln(100x + 1) + 150$ et l'eau et les sodas sont retirés de l'affichage. Un score de 100 est ainsi beaucoup plus vite atteint. Changer la relation de changement d'échelle pour une catégorie donnée revient à construire ici à un affichage intra-catégorie pour celle-ci et non pas inter-catégorie. La note d'une boisson ne peut alors pas être comparée à la note d'un aliment.

On peut imaginer un affichage qui ne ferait que de l'inter-catégorie ou un qui ne ferait que de l'intra-catégorie. Mais si l'objectif de l'affichage est de permettre des comparaisons inter- et intra-catégories, il semble difficile **d'argumenter un affichage dont l'objectif change selon la catégorie (pour certains inter-, pour d'autres intra-catégorie) comme cela est le cas ici**. Ce choix très particulier nécessite une argumentation qui n'est pas présente dans la documentation de l'Ecoscore et d'une transparence vis-à-vis des consommateurs.

Remarque finale

Dans tous les cas, il faut argumenter vis-à-vis de la construction de l'équation :

- Le choix des bornes (ce qui a été négligé dans la proposition initiale) en évitant de les déterminer en fonction des points extrêmes, mais bien selon l'ensemble des données disponibles.
- L'échelle logarithmique base 2, passé d'une échelle linéaire à une échelle logarithmique peut être discutée vis-à-vis de la perception du consommateur, en s'appuyant sur la loi de Fechner par exemple, mais cet aspect doit être approfondi.
- Un affichage à 5 niveaux (ce qui signifie que le meilleur est 32 fois mieux que le pire).

ADEPALE et ATLA

L'ADEPALE et l'ATLA utilisent une autre formule pour passer à une échelle logarithmique :

$$100 - 50 \times \log_{10} \left(\frac{x}{6,09} \times 100 \right)$$

ce qui permet d'obtenir un score entre 0 et 100 pour les valeurs entre les bornes $x_{ub} = 6,09$ et $x_{lb} = 0,0609$ mPt/kg respectivement.

Pour avoir une relation en logarithme base 10 entre 0 et 100, une manière plus générale d'écrire cette formule est :

$$\begin{aligned} \{x < x_{lb} \rightarrow Eco - score = 100 \\ x_{lb} \leq x \leq x_{ub} \rightarrow Eco - score = \frac{\log_{10}(x_{ub}) - \log_{10}(x)}{\log_{10}(x_{ub}) - \log_{10}(x_{lb})} \times 100 \\ x > x_{ub} \rightarrow Eco - score = 0 \end{aligned}$$

Il serait préférable de ne pas déterminer les bornes par un point unique, ici le plus grand impact. Choisir les bornes pour représenter par exemple 95% du jeu de données ($x_{ub} = x_{97,5\%}$, $x_{lb} = x_{2,5\%}$) et ainsi s'affranchir des valeurs extrêmes semble une alternative plus robuste.