

# Les indicateurs en perspective\*

Thierry Marchant<sup>†</sup> Denis Bouyssou<sup>‡</sup> Patrice Perny<sup>§</sup>

Marc Pirlot<sup>¶</sup> Alexis Tsoukias<sup>||</sup> Philippe Vincke<sup>\*\*</sup>

3 juin 2003

## Résumé

Les indicateurs ou indices sont de plus en plus fréquents, aussi bien dans des domaines techniques que dans notre vie quotidienne. Ils sont utilisés pour mesurer l'obésité, la confiance, la santé, la solvabilité d'un client, l'éthique d'un placement, le respect des droits de l'homme, etc. Au delà de la mesure, ils servent souvent à prendre ou à appuyer des décisions. Il est donc important de savoir si ces indicateurs remplissent correctement leur rôle : s'ils mesurent bel et bien ce qu'ils sont sensés mesurer et s'ils peuvent servir d'outils d'aide à la décision.

Dans cet article, nous ne donnons pas de réponse générale, mais, à travers l'analyse de trois indicateurs bien connus, nous montrons que la question est complexe et que de nombreux problèmes peuvent se présenter. Ces trois exemples sont l'indice de développement humain, l'indice de pollution de l'air (ATMO) et le score au décathlon.

Dans la dernière section, nous comparons les indicateurs, comme outils d'aide à la décision, avec l'aide multicritère à la décision qui, elle aussi, cherche à résumer des données complexes et multidimensionnelles.

Les indicateurs sont omniprésents dans notre vie quotidienne : Q.I., Dow Jones, PNB, qualité de l'air, nombre de médecins par habitant, indice de pauvreté, indice de position sociale, indice des prix à la consommation, rate

---

\*Ce texte est en grande partie basé sur le chapitre 4 d'un livre publié par les mêmes auteurs [2]

<sup>†</sup>Université de Gand, [thierry.marchant@rug.ac.be](mailto:thierry.marchant@rug.ac.be)

<sup>‡</sup>CNRS – LAMSADE, Université Paris Dauphine

<sup>§</sup>LIP6, Université Paris 6

<sup>¶</sup>Faculté Polytechnique de Mons

<sup>||</sup>CNRS – LAMSADE, Université Paris Dauphine

<sup>\*\*</sup>Université Libre de Bruxelles

of return, etc. En lisant un journal, on peut avoir le sentiment que ces nombres magiques gouvernent le monde :

Les pays européens avec un ratio déficit/PNB inférieur à 3% pourront faire partie de la zone EURO.

Qualité de l'air ce jour : 7. Il est conseillé aux personnes âgées, aux femmes enceintes et aux jeunes enfants de rester à l'intérieur.

La banque mondiale menace le gouvernement  $x$  de lui retirer son aide s'il ne réussit pas à ramener l'indice  $y$  en dessous du niveau  $z$ .

Remarquez que, bien souvent, la banque mondiale ne fonde pas sa décision uniquement sur des considérations économiques ou financières. Les violations des droits de l'homme sont également fréquemment invoquées pour motiver telle ou telle décision. Mais il faut savoir qu'il existe aussi un indice des droits humains (see e.g. [6]).

Pourquoi ces indicateurs sont-ils si puissants? Probablement parce qu'il est presque unanimement admis qu'ils offrent un reflet fidèle de la réalité, qu'ils sont objectifs. Ceci nous amène à soulever plusieurs questions.

1. Y-a-t-il une réalité, plusieurs réalités ou pas de réalité du tout? Nombreux sont les philosophes qui aujourd'hui considèrent que la réalité n'est pas unique. Chaque individu a sa propre perception du monde et partant sa propre réalité. On peut rétorquer que ces différentes réalités, particulières à chaque individu, ne sont que des visions particulières d'une seule et même réalité. Mais, comme il est impossible ou certainement difficile de considérer la réalité indépendamment de la perception que nous en avons, il n'y a peut-être pas de sens à considérer la réalité *per se*, indépendamment de nous [13]. Par conséquent, un indice pourrait fort bien ne refléter que la réalité de celui qui l'a construit et n'être pertinent que lui.
2. Quelle que soit la réponse à la question précédente, pouvons-nous espérer qu'un indicateur reflète fidèlement la réalité ou une réalité? La réalité est tellement complexe que nous pouvons en douter. Nous devons dès lors nous contenter d'indicateurs ne prenant en compte que certains aspects de la réalité. Nous devons concevoir des indicateurs qui, de la réalité, prennent en compte les facettes ou les aspects qui sont importants ou pertinents pour nous, pour notre problème, dans un contexte donné. Or c'est loin d'être toujours le cas. Par exemple, l'indice de développement humain (IDH), défini par le Programme des Nations Unies pour le Développement (PNUD) pour mesurer le

développement [10], est utilisé sur tous les continents et dans tous les domaines (politique, économie, industrie, etc.). Il nous est permis de douter que, dans toutes ces applications, les intérêts soient les mêmes. Dans le rapport 1997 sur le développement humain (p.21), le PNUD annonce fièrement que

Dans de nombreux pays, l'IDH est utilisé pour classer les collectivités territoriales de différents niveaux afin de repérer celles qui sont les plus défavorisées en termes de développement humain. Plusieurs pays, les Philippines, par exemple, ont recours à de telles analyses dans le cadre de leurs opérations de planification. [...] L'IDH est en particulier utilisé lorsqu'un chercheur a besoin d'un outil de mesure composite du développement. Dans ce cadre, d'autres indicateurs l'accompagnent parfois.

Ceci démontre clairement que de nombreuses personnes utilisent l'IDH de façons très différentes.

De plus, même dans le cas du PNUD, on peut se demander si l'IDH a été construit avec un objectif bien précis. Pourquoi le PNUD a-t-il besoin d'un indice de développement humain? Pour couper les vivres aux états évoluant dans la mauvaise direction? Pour répartir les subsides parmi les pays les plus pauvres (en fonction de quelle clef de répartition)? Pour mettre la pression sur les gouvernements des états les moins bien évalués? Pour prouver que les démocraties occidentales ont les meilleurs systèmes politiques?

3. Supposons un instant que le but d'un indicateur soit clairement défini. Sommes-nous pour autant sûr qu'il "indique" ce que nous voulons? Sommes-nous sûr que les opérations mathématiques utilisées lors du calcul de l'indicateur nous mènent à des conclusions sensées et pertinentes?

Nous allons maintenant analyser en détail trois indicateurs bien connus, dans trois domaines totalement différents : l'indice de développement humain, l'indice de qualité de l'air et le score au décathlon.

## 1 L'indice de développement humain

Dans le rapport 1997 sur le développement humain, p.15 [10], le PNUD écrit :

L'indicateur de développement humain mesure le niveau moyen auquel se trouve un pays donné selon trois critères essentiels du développement humain : longévité, instruction et conditions de vie. Cet indicateur composite comprend donc trois variables : l'espérance de vie, le niveau d'éducation (mesuré, d'une part, par le taux d'alphabétisation des adultes et, d'autre part, par le taux combiné de scolarisation dans le primaire, le secondaire et le supérieur) et le PIB réel par habitant (mesuré en PPA, ou parités de pouvoir d'achat).

La définition précise de l'IDH est présentée à la page 137 du rapport 1997 sur le développement humain. L'IDH est la moyenne arithmétique d'un indicateur d'espérance de vie, d'un indicateur de niveau d'éducation et d'un indicateur de PIB réel corrigé par habitant (PPA). Chacun de ces trois indicateurs est calculé comme suit.

**Indicateur d'espérance de vie (IEV)** Cet indicateur mesure l'espérance de vie à la naissance. Afin de normaliser l'échelle de cet indicateur, une valeur minimale (25 ans) et maximale (85 ans) ont été fixées. L'indicateur est défini par

$$\frac{\text{espérance de vie à la naissance} - 25}{85 - 25}.$$

C'est donc une valeur comprise entre 0 et 1.

**Indicateur de niveau d'éducation (INE)** C'est une combinaison de deux autres indicateurs : l'indicateur d'alphabétisation des adultes (IAA) et l'indicateur de scolarisation tous niveaux confondus (IS). Le premier est la proportion d'adultes alphabétisés tandis que le second est la proportion d'enfants en âge scolaire qui vont réellement à l'école (primaire, secondaire ou supérieure). L'INE est une moyenne pondérée de l'IAA et de l'IS. Il est égal à

$$\frac{2\text{IAA} + \text{IS}}{3}.$$

**L'indicateur de PIB réel corrigé par habitant en dollars PPA (IPIB)**

Cet indicateur mesure le revenu par habitant. Etant donné que un dollar a plus de valeur pour quelqu'un gagnant \$100 que pour quelqu'un gagnant \$100 000, le revenu est d'abord transformé en utilisant la formule d'Atkinson [1]. La valeur transformée du montant  $y$ , c.à.d.  $W(y)$  est donnée par

$$\left\{ \begin{array}{ll} y & \text{si } 0 < y < y^*, \\ y^* + 2[(y - y^*)^{1/2}] & \text{si } y^* \leq y < 2y^*, \\ y^* + 2(y^*)^{1/2} + 3[(y - 2y^*)^{1/3}] & \text{si } 2y^* \leq y < 3y^*, \\ \vdots & \\ y^* + 2(y^*)^{1/2} + 3(y^*)^{1/3} + \dots + n[(y - (n-1)y^*)^{1/n}] & \text{si } (n-1)y^* \leq y < ny^*. \end{array} \right.$$

Dans cette formule,  $y$  représente le revenu,  $W(y)$ , le revenu transformé et  $y^*$  est fixé à \$5 835 (PPA\$), c.à.d. la moyenne mondiale du revenu annuel par habitant en 1994.

L'échelle des revenus est ensuite normalisée en utilisant les valeurs \$40 000 et \$100 comme maximum et minimum et la formule

$$\frac{\text{revenu transformé} - W(100)}{W(40\,000) - W(100)}.$$

C'est donc également une valeur comprise entre 0 et 1. Notons que  $W(40\,000) = 6\,154$  et  $W(100) = 100$ .

Quelques mots maintenant à propos des données et de leur date de collecte : le rapport sur le développement humain est une publication annuelle (depuis 1990). Bien évidemment, le rapport 1997 ne contient que des données antérieures à 1997. En fait, l'IDH calculé dans le rapport 1997 est considéré par le PNUD comme l'IDH de 1994. En réalité, les choses sont encore un peu plus compliquées puisque l'IDH de l'année  $i$ , publié dans le rapport de l'année  $j$  ( $j > i$ ) est un agrégat de données collectées en l'année  $i$  pour certaines dimensions et en d'autres années, antérieures, pour d'autres dimensions. Dans ce texte, nous n'utilisons que des données figurant dans le rapport sur le développement humain 1997 (en bref RDH97), c.à.d. des données de 1994 ou plus tôt.

Pour illustrer le calcul de l'IDH, calculons celui de la Grèce (RDH97). L'espérance de vie en Grèce est de 77.8 ans. L'IEV est donc égal à  $(77.8 - 25)/(85 - 25) = 0.880$ . L'IAA est égal à 0.967 et l'IS vaut 0.820. Par conséquent,  $INE = (2 \times 0.967 + 0.820)/3 = 0.918$ . Le PIB grec réel par habitant est de \$11 265 et se situe au dessus de  $y^*$  mais sous  $2y^*$ . Le PIB réel ajusté par habitant est donc de \$5 982 (PPA\$) car  $5\,982 = 5\,835 + 2(11\,265 - 5\,835)^{1/2}$ . Dès lors,  $IPIB = (5\,982 - W(100))/(W(40\,000) - W(100)) = (5\,982 - 100)/(6\,154 - 100) = 0.972$ . Finalement, l'IDH de la Grèce est égal à  $(0.880 + 0.918 + 0.972)/3 = 0.923$ .

## 1.1 La normalisation des échelles

Pour pouvoir calculer l'IEV et l'IPIB, des valeurs maximales et minimales ont été fixées, de façon à ce que, après normalisation, l'étendue de l'échelle couvre l'intervalle de 0 à 1. Le choix de ces limites est arbitraire. Pourquoi 25 et 85 ans ? Est-ce que 25 ans correspond à la plus petite valeur observée ? Non, la plus petite valeur observée est 22.6 ans (Rwanda, RDH97). L'IEV est donc négatif pour le Rwanda. En fait, le nombre 25 avait été choisi dans le premier rapport (1990), alors que la plus petite valeur observée était au dessus de 35. A cette époque, personne n'aurait imaginé que l'espérance de vie puisse un jour descendre sous la barre des 25 ans. Pour éviter ce problème, ils auraient pu choisir une valeur plus basse encore : 20 ou 10. La probabilité d'observer une valeur sous la limite inférieure aurait été plus petite. Mais le choix de ces limites n'est pas sans conséquences. Considérons l'exemple suivant.

Supposons que l'INE et l'IPIB aient été calculés pour la Corée du Sud et le Costa Rica (RDH97). Nous connaissons également l'espérance de vie à la naissance pour la Corée du Sud et le Costa Rica (voir Table 1). Si le

	espérance de vie	INE	IPIB
Corée du Sud	71.5	.93	.97
Costa Rica	76.6	.86	.95

TAB. 1 – Limites : espérance de vie, INE et IPIB pour la Corée du Sud et le Costa Rica (RDH97)

maximum et le minimum pour l'espérance de vie sont fixés à 85 et 25, alors l>IDH est de 0.890 pour la Corée du Sud et 0.889 pour le Costa Rica. Mais si le maximum et le minimum pour l'espérance de vie sont fixés à 80 et 25, alors l>IDH est de 0.915 pour la Corée du Sud et 0.916 pour le Costa Rica. Dans le premier cas, le Costa Rica est moins développé que la Corée du Sud tandis que, dans le second, le Costa Rica est plus développé que la Corée du Sud. Ceci nous montre bien que le choix des limites est important.

En fait, quand nous rétrécissons l'étendue de l'espérance de vie de [25,85] à [25,80], nous augmentons la différence entre toute paire de valeurs de l'IEV d'un facteur égal à  $(85 - 25)/(80 - 25)$ . Ceci revient donc à augmenter le poids de l'IEV du même facteur. Dans notre exemple, l'espérance de vie est plus grande au Costa Rica qu'en Corée du Sud. Il n'est donc pas étonnant qu'en donnant plus de poids à l'espérance de vie (en rétrécissant l'étendue de l'échelle), la position du Costa Rica s'améliore par rapport à la Corée du Sud.

Remarquons finalement que, du moins en apparence, aucune limite n'a été fixée pour l'IAA et l'IS. Dès lors, *de facto*, les valeurs 1 et 0 correspondent au maximum et au minimum. Ceci est aussi un choix arbitraire. Il est clair que les valeurs 1 et 0 n'ont pas été observées et ne risquent pas d'être observées avant longtemps. L'étendue réelle de ces échelles est donc plus petite que  $[0,1]$  et on pourrait les normaliser en utilisant d'autres valeurs que 0 et 1.

## 1.2 La compensation

Les données pour le Gabon et les Iles Salomon (RDH97) sont présentées à la Table 2. Les Iles Salomon ont des performances moyennes à bonnes sur toutes les dimensions. Le Gabon, quant à lui, est légèrement meilleur que les Iles Salomon sur toutes les dimensions sauf l'espérance de vie où il est très mauvais (15 ans de différence). Il nous semble que cette très faible espérance de vie est un signe clair de faible développement humain, même si la situation est meilleure sur les autres dimensions. Néanmoins, l'IDH est égal à 0.56 pour

	espérance de vie	IAA	IS	PIB réel
Gabon	54.1	.63	.60	3 641
Iles Salomon	70.8	.62	.47	2 118

TAB. 2 – Compensation : performances du Gabon et des Iles Salomon (RDH97)

le Gabon et les Iles Salomon. Donc, malgré l'analyse informelle que nous venons de faire, nous devrions conclure que le Gabon et les Iles Salomon ont le même niveau de développement. Ce problème est dû au fait que nous avons utilisé la moyenne arithmétique pour agréger nos données en un nombre. Les faiblesses sur une dimension peuvent donc être compensées par des avantages sur d'autres dimensions. Ceci est probablement désirable, jusqu'à un certain point. Mais des faiblesses extrêmes (comme une espérance vie environ 15 ans plus courte) ne devraient pas pouvoir être compensées par de bonnes performances sur d'autres dimensions.

Nous poursuivons plus loin notre analyse de la compensation. Puisque n'importe quelle faiblesse peut être compensée par un point fort, une diminution de l'espérance de vie de un an peut être compensée par une augmentation du PIB réel corrigé (revenu transformé par la formule d'Atkinson). Calculons cette augmentation. Une diminution de l'espérance de vie de un an mène à une diminution de l'IEV de  $1/(85-25) = 0.016667$ . Pour compenser celle-ci, l'IPIB doit augmenter de la même quantité. Le PIB réel corrigé

doit donc augmenter de  $0.016667(6154 - 100) = 100.9\$$  (souvenez-vous que  $W(40000) = 6154$ ). De même, une diminution de l'espérance de vie de 2 ans peut être compensée par une augmentation du PIB réel corrigé égale à  $2 \times 100.9\$$ . Une diminution de l'espérance de vie de  $n$  années peut être compensée par une augmentation du PIB réel corrigé égale à  $n \times 100.9\$$ . La valeur d'une année d'espérance de vie est donc  $100.9\$$  (transformé par la formule d'Atkinson). Le nombre 100.9 est appelé le taux de substitution entre l'espérance de vie et le PIB réel corrigé.

D'autres taux de substitution peuvent aisément être calculés : entre l'espérance de vie et l'alphabétisation des adultes, il est de  $0.016667(1 - 0)(3/2) = 0.025$ . Pour compenser une diminution de l'espérance de vie de  $n$  années, une augmentation de l'alphabétisation des adultes égale à  $n \times 0.025$  est nécessaire.

Raisonnons maintenant en termes de PIB réel non ajusté. Dans un pays où le PIB réel est de  $13071\$$  (Chypre, RDH97), une diminution de l'espérance de vie de un an peut être compensée par une augmentation du PIB réel de  $21084\$$ . Par contre, dans un pays avec un PIB réel de  $700\$$  (Tchad, RDH97), une diminution de l'espérance de vie de un an peut être compensée par une augmentation du PIB réel de  $100.9\$$ . Par conséquent, une année d'espérance de vie pour une population pauvre a beaucoup moins de valeur qu'une année pour une population riche.

### 1.3 L'indépendance des dimensions

Considérons l'exemple de la Table 3. Les pays  $x$  et  $y$  ont des performances

	espérance de vie	IAA	IS	PIB réel
$x$	30	.80	.65	500
$y$	30	.35	.40	3500

TAB. 3 – Indépendance : performances de  $x$  et  $y$

égales pour l'espérance de vie,  $y$  est nettement moins bon que  $x$  en ce qui concerne le taux d'alphabétisation des adultes mais nettement meilleur sur le plan du revenu. Etant donné que l'espérance de vie est très courte, on peut considérer que le taux d'alphabétisation des adultes n'est pas très important (car il n'y a presque pas d'adultes) mais que le revenu est plus important car il permet d'améliorer la qualité de vie. On peut d'ailleurs espérer que le taux de scolarisation et l'espérance de vie s'améliorent rapidement grâce aux revenus élevés. On pourrait donc conclure que  $y$  est plus développé que  $x$ . Notre conclusion est confirmée par l'IDH : 0.30 pour  $x$  et 0.34 pour  $y$ .



Comparons maintenant deux autres pays,  $w$  et  $z$  semblables en tous points à  $x$  et  $y$  sauf que l'espérance de vie est égale à 70 ans pour  $w$  et  $z$  (voir Table 4). Dans ces conditions, le taux d'alphabétisation des adultes de  $z$  est vraiment très mauvais par rapport à celui de  $w$ . La population adulte est très importante et son analphabétisme est un sérieux problème. Même si  $y$  emploie son revenu élevé pour tenter d'améliorer l'éducation, il faudra des décades pour voir des changements significatifs. Par contre, le faible revenu

	life expectancy	IAA	IS	real GDP
$w$	70	.80	.65	500
$z$	70	.35	.40	3 500

TAB. 4 – Independence : performances of  $w$  and  $z$

de  $w$  ne semble pas être un problème de développement majeur puisque l'espérance de vie et l'éducation atteignent un bon niveau. Il n'est donc pas déraisonnable de conclure que  $w$  est plus développé que  $z$ . Mais si nous calculons d'IDH, nous obtenons 0.52 pour  $w$  et 0.56 pour  $z$  ! Ceci ne doit pas nous surprendre. Il n'y a pas de différences entre  $x$  et  $y$  d'une part et  $w$  et  $z$  d'autre part si ce n'est l'espérance de vie. Mais les différences d'espérance de vie entre, d'un part,  $x$  et  $w$  et, d'autre part,  $y$  et  $z$  sont égales. Elles déterminent donc la même augmentation de l'IDH pour  $w$  et  $z$ , comparés respectivement à  $x$  et  $y$ .

Quand nous utilisons une somme ou une moyenne arithmétique, pondérée ou non, pour agréger plusieurs dimensions, des performances identiques de deux objets (par exemple deux pays) sur une ou plusieurs dimensions n'ont pas d'influence sur la façon dont ces objets se comparent. Les performances identiques peuvent être modifiées dans l'une ou l'autre direction, pour autant qu'elles restent identiques, sans affecter la comparaison entre les objets. Cette propriété, inhérente aux sommes et moyennes arithmétiques, est appelée *indépendance des dimensions*. Elle n'est pas toujours désirable, notamment quand nous comparons des pays sur base de l'espérance de vie, l'éducation et le revenu.

#### 1.4 La construction des échelles

Ce point a déjà été abordé dans la Section 1.1 (Normalisation des échelles). Mais la construction d'une échelle, c'est plus que la normalisation. Prenons par exemple le PIB réel. Avant de normaliser cette échelle, on transforme le PIB au moyen de la formule d'Atkinson. Le but de cet ajustement est évident : un dollar de plus pour quelqu'un qui en gagne 40 000, c'est

négligeable ; mais un dollar de plus pour celui qui en gagne 100, c'est important. La formule d'Atkinson reflète cela. Mais, pourquoi choisir  $y^* = \$5\,835$  ? Et pourquoi la formule d'Atkinson ? D'autres formules et d'autres valeurs de  $y^*$  pourraient convenir. Une fois de plus, ce sont des choix arbitraires, du moins partiellement. Et il serait simple de trouver un exemple montrant qu'avec un autre choix (une autre valeur de  $y^*$ , p.ex.), tout aussi défendable, on arriverait à un autre rangement des pays.

Notons encore que le fait de n'avoir pas corrigé ou transformé l'espérance de vie, le taux d'alphabétisation des adultes ou le taux de scolarisation est aussi un choix arbitraire. On pourrait défendre le point de vue selon lequel augmenter l'espérance de vie d'un an dans un pays où elle est de 30 ans est un grand progrès tandis que ce n'est qu'un léger progrès dans un pays où l'espérance de vie est de 70 ans. Certains pourraient même avancer qu'augmenter l'espérance de vie au delà de 80 ans n'est plus une amélioration. Cela engendre de telles dépenses de soins de santé qu'il y a moins de ressources pour d'autres domaines importants comme l'éducation, la politique en faveur de l'emploi, etc.

## 1.5 Quelques aspects statistiques

Considérons les quatre indicateurs formant l'IDH, d'un point de vue statistique. L'indicateur d'espérance de vie est la moyenne sur une population et sur une certaine période de la durée de vie des individus composant cette population. Il est bien connu que les moyennes, même si elles peuvent nous rendre service, ne peuvent refléter la variété des observations. Un pays où la totalité des individus vit jusque 50 ans a une espérance de vie de 50 ans. De même, un pays où une partie de la population (rurale, pauvre ou d'une certaine race) meurt jeune et où une autre vit jusque 80 ans peut aussi avoir une espérance de vie de 50 ans.

Notons que ce type de moyenne est assez particulier. Ce n'est pas la même moyenne, par exemple, que lorsque, disposant de plusieurs mesures du poids d'un objet, nous considérons la moyenne de ces mesures comme une bonne estimation du poids réel. Le poids d'un objet existe réellement (si la réalité existe). Par contre, même si la réalité existe, la moyenne de la durée de vie ne correspond à rien de réel. C'est la durée de vie d'un humain moyen ou idéal, comme si nous, les humains, étions des copies imparfaites de cet humain idéal.

Jusqu'au 19<sup>ème</sup> siècle, les deux sortes de moyennes étaient désignées par des noms différents : moyenne proportionnelle (différentes mesures d'un seul objet) et valeur commune (plusieurs objets, chacun mesuré une seule fois).

Elles étaient considérées comme des concepts différents. Au 19ème siècle, l'astronome et statisticien belge Quetelet (1796-1894) inventa le concept d'*homme moyen* et unifia les deux types de moyenne [4].

Pour vous convaincre que le concept d'homme moyen est étrange (quoiqu'éventuellement utile), considérons un pays dont tous les habitants seraient des triangles rectangles de différentes tailles et formes (cet exemple est emprunté à [15]). Pour simplifier l'exemple, supposons qu'il n'y a que deux sortes de triangles rectangles (voir Fig. 1), en proportions égales. Un statisticien veut déterminer le triangle rectangle moyen (l'habitant moyen) et calcule, à cet effet, la longueur moyenne de chaque côté. Il obtient un triangle avec des côtés de longueurs égales à 4, 8 et 9, c.à.d. un triangle qui n'est pas rectangle. En effet,  $4^2 + 8^2 \neq 9^2$ . Le triangle rectangle moyen n'est donc pas un triangle rectangle ! L'angle qui paraît droit est en fait un angle d'environ 91 degrés. Dans le même esprit, Quetelet a mesuré la taille moyenne des humains, dans toutes les dimensions, y compris la taille du foie, du cœur, des poumons et des autres organes. Ce qu'il obtint était un humain moyen dans lequel il était impossible de placer les organes moyens. Ils prenaient trop de place !

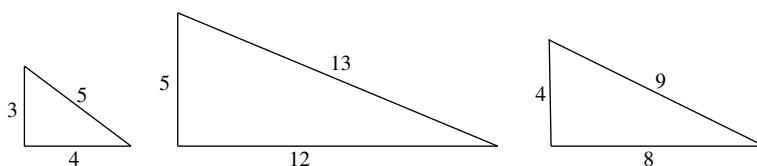


FIG. 1 – Deux triangles rectangles et leur moyenne

L'indicateur d'alphabétisation des adultes est assez différent de celui de l'espérance de vie. Il est simplement égal au nombre d'adultes alphabétisés, divisé par la taille de la population adulte (ce qui permet des comparaisons entre pays). On peut donc penser que ce n'est pas une moyenne mais plutôt une fréquence relative. En fait, cela dépend de notre interprétation. Si nous considérons qu'un IAA de 0.60 signifie que 60% de la population est alphabétisée, alors ce n'est pas une moyenne. Si, nous considérons que 0.60 signifie que l'habitant moyen est alphabétisé à 60%, alors c'est une moyenne. Et cette dernière interprétation n'est pas moins sensée que le calcul de l'espérance de vie. Si nous définissons une variable dont la valeur est 1 pour un adulte alphabétisé et 0 pour un analphabète et que nous calculons la moyenne de cette variable sur la population sur une certaine période (comme pour l'espérance de vie), ce que nous obtenons n'est rien d'autre que l'indicateur d'alphabétisation des adultes.

Nous pouvons faire la même analyse avec l'indicateur de scolarisation et le PIB réel corrigé. Ce sont des nombres qui sont mesurés au niveau du pays et puis divisés par la taille de la population. Mais ils peuvent aussi être interprétés au niveau de l'individu, comme des moyennes.

Qu'en est il alors de l'IDH ? Comment devons nous le considérer ? Selon le rapport sur le développement humain, l'IDH

[...] mesure le niveau moyen auquel se trouve un pays donné  
[...]

De plus, l'IDH contient un indicateur (l'IEV) qui ne peut être interprété que en ayant présente à l'esprit l'idée d'homme moyen de Quetelet. Il en résulte, si nous voulons être cohérents, que l'IAA, l'IPIB et l'IDH doivent eux aussi être interprétés comme des moyennes, à la lumière du concept d'homme moyen. L'IDH décrit donc, d'une certaine façon, le niveau de développement humain de l'homme moyen.

## 2 L'indice de qualité de l'air

En réaction à l'augmentation alarmante de la pollution de l'air, principalement dans les villes, au cours des dernières décades, plusieurs gouvernements et organisations internationales ont établi des normes pour la concentration de différents polluants dans l'air (par exemple, le Clean Air Act aux Etats-Unis). Généralement, ces normes spécifient, pour chaque polluant, une concentration à ne pas dépasser. Naturellement, ces normes ne sont que des normes et sont par conséquent souvent dépassées. Par conséquent, puisqu'un air de qualité n'est pas garanti par ces normes, différents systèmes de surveillance ont été développés pour fournir aux gouvernements mais aussi aux citoyens des informations à propos de la qualité de leur air.

Deux exemples de tels systèmes sont le *Air Quality Index (AQI)*<sup>1</sup>, développé par la US Environmental Protection Agency ([9] ou <http://www.epa.gov/oar/oaqps/psihold.html>), et l'indice *ATMO*, développé par le ministère français pour l'environnement (<http://www.ademe.fr/jda/indatmo.htm>). Ces deux indices sont fort semblables et nous allons étudier l'indice ATMO.

L'indice ATMO est basé sur la concentration de 4 polluants majeurs : le dioxyde de soufre (SO<sub>2</sub>), le dioxyde d'azote (NO<sub>2</sub>), l'ozone (O<sub>3</sub>) et les particules solides . Pour chaque polluant, un sous-indice est calculé et l'indice ATMO final est défini comme étant égal au sous-indice le plus élevé. Voici comment chaque sous-indice est calculé. Pour chaque polluant, la concentration est convertie en un nombre sur une échelle de 1 à 10. Le niveau 1

<sup>1</sup>This is a revised version (1999) of the older Pollutant Standard Index.

correspond à un air de qualité excellente ; les niveaux 5 et 6 se trouvent juste de part et d'autre des normes à long terme de l'Union Européenne ; le niveau 8 correspond aux normes UE à court terme et le niveau 10 indique un air dangereux.

Pour illustrer, supposons que les sous-indices soient ceux de la Table 5. L'indice ATMO qui en résulte est la valeur la plus élevée, c.à.d. 8, indi-

polluant	CO <sub>2</sub>	SO <sub>2</sub>	O <sub>3</sub>	poussière
sous-indice	3	3	2	8

TAB. 5 – Sous-indices de l'indice ATMO

quant un air de mauvaise qualité. Dans les paragraphes qui suivent, nous présentons certains problèmes particuliers à l'indice ATMO.

## 2.1 La monotonie

Supposons que, suite à un trafic intense, à l'absence de vent et à une journée fort ensoleillée, le sous-indice pour l'ozone passe de 3 à 8, pour l'air décrit à la Table 5. Ceci correspond sans aucun doute à un air de moins bonne qualité. Aucun polluant n'a diminué ; par contre, l'un d'eux a fortement augmenté. Nous nous attendons par conséquent à ce que l'indice ATMO reflète cette détérioration. Et pourtant, l'indice ATMO n'a pas changé. Le maximum des quatre sous-indices est toujours 8.

Ceci nous enseigne que certains changements, même importants, ne sont pas pris en compte par l'indice ATMO. Dans notre exemple, le changement était très important puisque nous sommes passés d'un air presque parfait à un air très toxique. Remarquons encore que si, quelques jours plus tard, le sous-indice pour l'ozone repasse à 3, l'indice ATMO ne change toujours pas alors que la qualité de l'air s'améliore considérablement. Ceci nous montre que l'indice ATMO n'est pas monotone : certains changements, dans les deux directions, ne sont pas reflétés par l'indice.

## 2.2 La non compensation

Considérons maintenant l'indice ATMO pour deux airs différents ( $x$  et  $y$ ) décrits à la Table 6.

L'air  $x$  est parfait sur toutes les dimensions excepté une : il dépasse légèrement les normes UE à long terme. L'air  $y$  n'est bon sur aucune dimension. Il est de qualité moyenne sur toutes les dimensions et proche des normes UE à long terme pour trois dimensions. L'indice ATMO est égal à

polluant	CO <sub>2</sub>	SO <sub>2</sub>	O <sub>3</sub>	poussière
$x$	1	1	6	1
$y$	5	4	5	5

TAB. 6 – Sous-indice pour  $x$  and  $y$

6 pour l’air  $x$  et à 5 pour l’air  $y$ . Par conséquent, la qualité de l’air  $x$  est considérée comme inférieure à celle de l’air  $y$ . Au contraire de ce que nous avons observé avec l’IDH, aucune compensation n’a lieu entre les différentes dimensions. Le faible désavantage de  $x$  (6 comparé à 5, pour l’ozone) n’est pas compensé par ses importants points forts (1 comparé à 4 ou 5, pour le dioxyde de carbone, le dioxyde d’azote et la poussière). Dans le cas du développement humain, la compensation entre dimensions était trop forte tandis que nous avons ici un autre extrême : pas de compensation du tout, ce qui n’est probablement pas mieux.

### 2.3 La signifiante

Oublions un instant nos critiques à l’encontre de l’indice ATMO et supposons qu’il fonctionne bien. Considérons l’affirmation “L’indice ATMO de ce jour (6) est le double de celui d’hier (3).” Que signifie-t-elle ? Nous allons montrer qu’elle n’est pas signifiante (qu’elle ne signifie rien) dans un certain sens. Revenons à la définition des sous-indices. Pour un polluant donné, la concentration est mesurée en  $\mu\text{g}/\text{m}^3$ . Les valeurs des concentrations sont ensuite transformées en nombre entre 1 et 10. Ceci est fait de façon plus ou moins arbitraire. Par exemple, au lieu de choisir les valeurs 5-6 pour les normes UE à long terme et 8 pour celles à court terme, on pourrait choisir 6-7 et 9. L’indice fonctionnerait aussi bien. La valeur même de l’indice ne nous fournit aucune information pertinente. Ce qui nous intéresse, c’est le fait que nous soyons en dessous ou au dessus de certaines normes, elles-mêmes liées aux effets des polluants sur la santé.

Mais, si nous changeons la définition des sous-indices, les valeurs des indices d’aujourd’hui et d’hier sont différentes ; par exemple 7 et 4. Or il n’est plus vrai que 7 soit le double de 4. Dès lors, puisque la véracité de l’affirmation “L’indice ATMO de ce jour est le double de celui d’hier” dépend d’un choix arbitraire, nous disons que cette affirmation n’est pas signifiante (voir [11] pour une introduction à la théorie de la signifiante). Au contraire, l’affirmation “Le sous-indice ATMO pour l’ozone de ce jour (6) est supérieur à celui d’hier,” est signifiante. Toute transformation raisonnable des valeurs de concentrations en nombres entre 1 et 10 nous conduirait à

la même conclusion : le sous-indice d'aujourd'hui est plus grand que celui d'hier. Par "transformation raisonnable," nous entendons une transformation qui préserve l'ordre : une concentration ne peut pas être transformée en une valeur d'indice inférieure à une autre valeur correspondant à une concentration inférieure. Des concentrations de 110 and 180  $\mu\text{g}/\text{m}^3$  peuvent être transformées en 3 et 6, ou 4 et 6, ou 2 et 4 mais pas 4 et 2.

Plus subtil maintenant. "L'indice ATMO de ce jour (6) est supérieur à celui d'hier (3)." Cette affirmation est-elle signifiante ? Dans le paragraphe précédent nous avons vu que le côté arbitraire de la construction de l'échelle de 1 à 10 pour un sous-indice n'est pas problématique quand nous voulons comparer deux valeurs du même sous-indice. Mais si nous voulons comparer des valeurs de deux sous-indices différents, ce n'est plus vrai. Un 3 pour un sous-indice peut-être plus dangereux qu'un 6 pour un autre. Bien entendu, les échelles ont été construites avec grand soin : 5 correspond aux normes UE à long terme et 8 à celles à court terme. Ceci a pour but de rendre les sous-indices commensurables. Les comparaisons devraient donc être significatives. Mais pouvons-nous vraiment assurer qu'un 5 (ou la concentration correspondante en  $\mu\text{g}/\text{m}^3$ ) soit équivalent pour deux sous-indices différents ? Équivalent en quels termes ? Certains polluants ont peut-être principalement des effets à court terme tandis que d'autres sont surtout nocifs à long terme. Ils peuvent avoir des effets sur parties différentes de l'organisme. Devons-nous comparer les effets en terme d'inconfort, de mortalité après  $n$  années, de coûts de soins de santé, ... ?

### 3 Le score au décathlon

Le décathlon est une compétition sportive qui se déroule en 10 épreuves : les courses sur 100, 400 et 1500 mètres, le 110 mètres haie, les lancers du javelot, du disque et du poids, les sauts en hauteur, en longueur et à la perche. Il a été introduit aux jeux olympiques de Stockholm en 1912 et est généralement étalé sur 2 ou 3 jours.

#### 3.1 Un peu d'histoire

Pour déterminer le vainqueur d'un décathlon, on calcule un score pour chaque athlète : la somme des scores obtenus pour chaque discipline. Celui qui a le meilleur score gagne. Les scores par discipline ne sont pas juste des temps et des distances. Cela n'aurait évidemment aucun sens d'additionner des temps et des distances. Jusqu'en 1908, les scores par discipline étaient simplement la place obtenue par l'athlète dans la discipline correspondante.

Par exemple, si un athlète faisait le troisième meilleur saut en hauteur, son score pour la discipline “saut en hauteur” était 3. Le vainqueur du décathlon était alors celui avec le score total le plus faible.

Le principal problème avec ces scores par discipline, c’est qu’ils sont une représentation très imparfaite des performances des athlètes. Supposons qu’un athlète arrive un dixième de seconde avant le suivant dans un 100 mètres. Ils ont les places  $i$  et  $i + 1$  (p.ex. 3<sup>ème</sup> et 4<sup>ème</sup>). La différence des scores qu’ils reçoivent est donc 1. Supposons maintenant que la différence entre les deux athlètes soit de une seconde, leurs places étant inchangées. La différence entre leurs scores reste 1 alors qu’on pourrait souhaiter une différence plus importante. C’est pourquoi d’autres façons de calculer les scores par discipline ont été utilisées depuis 1908 [3, 16]. On utilise depuis lors des tables permettant la conversion des temps ou des distances en score par discipline. Avec ces tables, les meilleurs résultats sont associés avec les scores les plus élevés, au contraire de ce qui se faisait avant 1908. Le vainqueur du décathlon est donc celui avec le score total le plus élevé.

De nombreuses versions de ces tables ont vu le jour. Plusieurs d’entre elles, en usage entre 1934 et 1962, sont basées sur l’idée qu’améliorer la performance d’un athlète d’une certaine quantité (p.ex. 5cm au saut en longueur) est d’autant plus difficile que la performance est bonne. Cela mérite donc plus de points. La forme générale de ces tables, pour les distances, est illustrée à la figure 2 (table convexe). Pour les temps, la forme est évidemment différente puisqu’une amélioration est une diminution du temps.

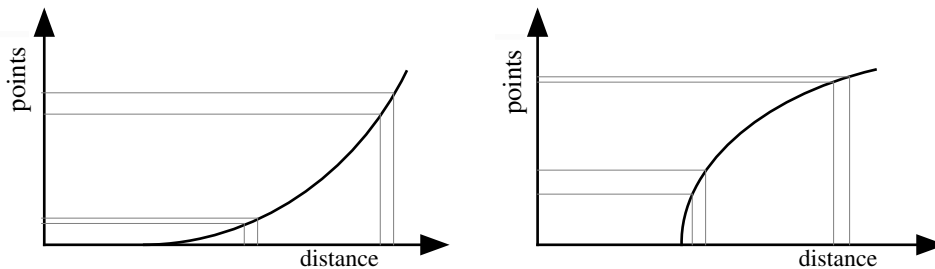


FIG. 2 – Tables du décathlon pour les distances : forme générale d’une table convexe (gauche) et concave (droite)

Les tables convexes sont à l’origine d’un problème important. Si un athlète décide de se concentrer sur quelques épreuves (par exemple les 4 courses) et de s’entraîner beaucoup plus pour ces épreuves là que pour les



autres, il sera avantagé. En effet, il va gagner beaucoup de points pour ces épreuves puisqu'il va effectuer d'excellentes performances. Par ailleurs, il effectuera des performances médiocres dans les autres disciplines mais cela lui fera perdre peu de points puisque la pente des courbes est plus faible dès qu'on s'éloigne des meilleures performances. Au total, il aura gagné plus de points qu'il n'en aura perdu. Ces tables convexes encouragent donc les athlètes à se spécialiser, ce qui est tout à fait contraire à l'esprit du décathlon.

C'est pour cette raison que, à partir de 1962, différentes tables concaves (voir Figure 2) ont été en vigueur. Celles-ci encouragent vivement l'excellence dans toutes les disciplines.

Un exemple d'une table réelle, en usage en 1998, est présenté à la Figure 3. Remarquez un nouveau changement : cette table n'est plus concave, elle est presque linéaire mais légèrement convexe.

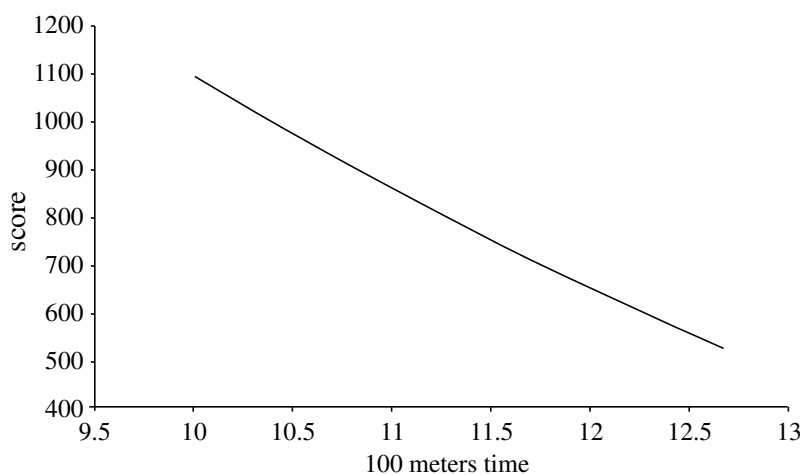


FIG. 3 – La table du 100 mètres en 1998, sous forme de courbe

Il y aurait de nombreux points intéressants à discuter à propos du décathlon.

- Comment fixe-t-on les valeurs minimales et maximales ? Elles peuvent fortement influencer le score comme nous l'avons vu avec l'IDH (Section 1.1). Bien entendu, les maximas doivent être, d'une façon ou une autre liés aux records du monde mais tout le monde sait que les athlètes prennent un malin plaisir à faire tomber ces records
- Pourquoi additionner les scores par discipline ? D'autres opérations pourraient éventuellement convenir. La multiplication, par exemple, favoriserait les athlètes ayant des performances égales dans toutes

les disciplines, ce qui est facile à montrer avec un petit exemple. Considérons qu'il n'y a que trois disciplines et que, pour chacune d'entre elles, le score par discipline est un nombre entre 0 et 10. Un athlète  $x$  obtient un score de 8 dans chaque discipline tandis que  $y$  obtient 9, 8 et 7. Si nous additionnons les scores, nous obtenons 24 pour  $x$  et pour  $y$ . Si nous les multiplions,  $x$  gagne avec 512 points contre 504 pour  $y$ .

– ...

Nous allons maintenant porter notre attention sur le rôle du score global au décathlon.

### 3.2 Rôle du score global au décathlon

A première vue, le rôle du score global est évident : désigner le vainqueur. Mais nous allons voir que la réalité est plus complexe et subtile. Et ceci n'est pas sans rapport avec les changements fréquents de définitions du score au cours de l'histoire du décathlon. Mais revenons-en au début de cette histoire. Le but de ceux qui imaginèrent le premier score global était probablement de simplement désigner le vainqueur mais nous pouvons être sûr qu'après la toute première compétition, un autre rôle fit son apparition. Beaucoup de gens ont probablement utilisé le score global pour évaluer et comparer les performances des athlètes. Par exemple, un tel athlète a un score proche de celui du vainqueur et peut donc être considéré comme un bon athlète tandis que tel autre a un score beaucoup moins bon que celui du vainqueur et est par conséquent considéré comme un nettement moins bon athlète.

Peu après (probablement après la deuxième compétition), un troisième rôle est apparu. Le score a très certainement été utilisé pour observer et mesurer l'évolution des athlètes. Par exemple, tel athlète a amélioré son score par rapport à la compétition précédente. Ou encore, le score de  $x$  dans cette compétition est meilleur que celui de  $y$  dans la compétition précédente. Remarquons que cette comparaison n'est pas signifiante (dans le sens introduit à la Section 2.3); supposons qu'un athlète gagne une compétition avec un score égal à 16. Dans la compétition suivante, à cause d'un plat de moules avariées, il fait une contre-performance mais ses principaux concurrents sont absents ou ont aussi de mauvais résultats. Il peut donc encore gagner le décathlon avec un score égal ou même supérieur à celui du décathlon précédent bien que ses performances soient moins bonnes.

On peut imaginer qu'après quelque temps, les organisateurs d'épreuves de décathlon se sont rendu compte du problème posé par ces nouveaux rôles. C'est probablement pourquoi ils ont abandonné la somme des rangs pour

utiliser des tables convexes. Ces tables ont rendu significatives, jusqu'à un certain point, les comparaisons de scores de différents athlètes dans différentes compétitions. A la même époque, le score global a commencé à être utilisé comme outil de suivi pour les entraînements. Auparavant, il ne pouvait être utilisé que pendant les compétitions puisqu'il était basé sur les rangs (ou les places). Il n'a alors pas fallu attendre longtemps avant qu'un entraîneur s'en serve comme outil stratégique, conseillant à son athlète de se concentrer sur telle ou telle épreuve afin de gagner un maximum de points. Pour cette raison, en 1962, les organisateurs ont décidé de donner un nouveau rôle au score global : stimuler l'excellence dans toutes les disciplines. Ceci a eu lieu grâce à l'introduction de tables concaves. Mais il est presque certain que le score est encore utilisé à des fins stratégiques, même si on peut espérer que les effets soient moins pervers.

Pour en terminer avec le décathlon, notons que le dernier rôle ne remplace aucun des précédents. Le score continue à remplir tous ces rôles simultanément. Il sert à récompenser des performances égales dans toutes les disciplines mais aussi à évaluer les performances d'un athlète. Même si nous ne considérons que ces deux rôles (les autres pouvant être vus comme des effets secondaires), il est étonnant de voir comme ils sont peu compatibles.

## 4 Les indicateurs et l'aide multicritère à la décision

Généralement, dans un processus d'aide multicritère à la décision (voir, p.ex., [7, 12, 14]), un décideur désire ordonner les éléments d'un ensemble d'alternatives du meilleur au moins bon, ou simplement choisir le meilleur élément. A cette fin, il identifie plusieurs points de vue, dimensions ou critères qui lui paraissent pertinents. Chaque alternative est caractérisée par une performance sur chaque critère : on appelle souvent ceci la matrice d'évaluation ou le tableau des performances. Une méthode d'aide multicritère à la décision est alors éventuellement utilisée pour ranger les alternatives, en tenant compte des préférences du décideur.

Quand nous construisons un indicateur, nous identifions aussi un certain nombre de dimensions. Chaque objet est également caractérisé par une performance sur chaque dimension et une valeur finale de l'indicateur est calculée, valeur que nous pouvons ensuite utiliser pour ranger les objets. L'analogie entre les indicateurs et l'aide multicritère à la décision est frappante : tous deux agrègent de l'information multi-dimensionnelle à propos d'un ensemble d'objets.

L'analogie va plus loin : les indicateurs ne sont pas souvent présentés

comme des modèles d'aide à la décision mais, pourtant, c'est souvent ce qu'ils sont. On les présente comme un moyen efficace de synthétiser l'information. Mais si nous avons besoin d'informations, c'est souvent précisément pour prendre des décisions éclairées.

Mais il y a une différence importante : quand on construit un indicateur, très fréquemment, il n'y a pas de problème de décision bien défini, pas de décideur identifié et, a fortiori, pas de préférences. On pourrait, pour pallier cette absence de préférences, considérer que les préférences sont celles des utilisateurs potentiels. Ceci est possible, jusqu'à un certain point car, très souvent, les préférences des utilisateurs vont dans la même direction pour chaque dimension prise séparément. Par exemple, pour chaque dimension de l'indice ATMO, chacun préfère une concentration aussi faible que possible. Mais ceci n'est certainement pas vrai pour les préférences globales. De plus, même si les préférences unidimensionnelles vont dans la même direction, elles ne sont pas forcément identiques. Ceux qui sont peu sensibles à un polluant particulier accorderont plus de valeur à une dimension de la concentration intervenant à haute concentration qu'à basse concentration. Par contre, ceux qui sont fort sensibles, accorderont peut-être la même valeur aux deux.

**L'apport de la théorie du mesurage** L'absence de préférences introduit une différence cruciale, comme nous allons le voir. En aide à la décision, de nombreuses études et concepts sont inspirés de la théorie du mesurage (voir p.ex. [8]). Celle-ci a pour objet la façon dont nous mesurons des objets (comment nous attribuons un nombre à un objet) de façon à refléter une relation entre ces objets. Par exemple, comment attribuer des nombres à des objets physiques de façon à refléter la relation "plus lourd que" ? C'est-à-dire, comment attribuer un nombre (appelé poids) à chaque objet de façon telle que le nombre associé à  $x$  (poids de  $x$ ) ne soit supérieur au nombre associé à  $y$  (poids de  $y$ ) que si  $x$  est plus lourd que  $y$  (ce que l'on peut observer en plaçant  $x$  et  $y$  sur les deux plateaux d'une balance). Des propriétés supplémentaires peuvent être souhaitables. Par exemple, dans le cas du poids, on peut désirer que le nombre attribué à  $x$  et  $y$  pris ensemble soit la somme de leur poids individuels.

Un autre exemple est celui de la distance. Comment attribuer des nombres à des points de l'espace de façon à refléter la relation "plus distant que", par rapport à un point de référence ? Contrairement à l'exemple précédent, celui-ci a plusieurs dimensions (généralement deux ou trois :  $x$ ,  $y$  ou  $x$ ,  $y$ ,  $z$  ou l'altitude, la longitude et la latitude, etc.). Chaque objet ou point est caractérisé par une performance (coordonnée) sur chaque dimension et on

cherche à agréger ces performances en un indicateur : la distance par rapport au point de référence. Ce problème est au cœur de la géométrie et la solution n'est pas unique. Le plus souvent, nous choisissons la distance Euclidienne (en faisant l'hypothèse que le chemin le plus court entre deux points est la ligne droite). Parfois, une distance Gaussienne semble plus appropriée (quand nous voulons aller d'un point à l'autre à la surface de la terre, le chemin le plus court n'est pas une droite, à moins d'être une taupe). Dans d'autres cas encore, on peut choisir la distance de Manhattan (entre deux points sur l'île de Manhattan, à moins de voler, le chemin le plus court n'est ni une droite, ni une courbe mais une succession de lignes droites se croisant à angles droits). Et on pourrait continuer les exemples.

Tant que nous ne considérons que des caractéristiques physiques, (plus grand que, plus chaud que, plus rapide que, ...), le problème est assez simple : on fait des mesures sans trop de problèmes depuis l'antiquité sans l'aide de la théorie du mesurage. Mais dès que nous cherchons à refléter d'autres types de relations, les choses se gâtent. Comment attribuer des nombres à des personnes ou à des alternatives de façon à refléter la relation "plus aimable que", "plus développé que", "préférable à" ou "plus risqué que" ? Dans ces cas là, la théorie du mesurage peut être d'une grande aide même si elle ne peut résoudre tous les problèmes.

En aide à la décision, c'est la relation "préférable à" qui nous intéresse. Une fois que nous avons mesuré les alternatives de façon à refléter cette relation, il est assez simple de manipuler des nombres. On fait souvent l'hypothèse qu'une relation de préférence sur l'ensemble des alternatives existe mais n'est pas bien connue et on essaie de mesurer les alternatives de façon à découvrir cette relation de préférence. Parfois, on suppose que cette relation n'existe pas entièrement, que les préférences peuvent se construire et évoluer au cours du processus d'aide à la décision, mais que certaines parties ou caractéristiques de la relation existent quand même a priori. La théorie du mesurage peut alors être utilisée pour développer ou analyser une méthode d'aide à la décision.

Par contre, beaucoup d'indices sont construits sans supposer qu'une relation sur l'ensemble des objets existe a priori ou sans essayer de refléter une relation pré-existante. Il semble, au contraire, dans de nombreux cas que le but de l'indice soit précisément de construire ou créer cette relation. Par conséquent, la théorie du mesurage ne nous apprend pas grand chose sur les indicateurs. Elle perd son pouvoir lorsqu'il n'y a pas de relation à refléter.

**Les indicateurs et la réalité** Comme nous venons de le voir, un indicateur n'aide pas à découvrir la réalité, c.à.d. une relation pré-existante. Il institue la réalité [4]. Ceci est particulièrement clair avec le score du décathlon. Entre 1908 et 1962, le score avait pour but d'évaluer les performances et de les comparer. Etant donné que l'essentiel, pour un athlète, c'est de gagner (quoiqu'en pensât de Coubertin), le score a été considéré comme la vraie mesure de la performance. Il est devenu la définition de la performance. Tout athlète qui n'en était pas convaincu avait intérêt à changer d'avis et à agir en conséquence s'il voulait continuer la compétition. Ceci n'est pas propre au décathlon. De nombreux gouvernements tentent probablement d'avoir un bon IDH pour leur pays afin de conserver les subsides internationaux ou de légitimer leur autorité aux yeux de leur population ou des gouvernements étrangers. Certaines mairies, désireuses d'attirer des résidents à haut pouvoir d'achat, prétendent avoir un air de bonne qualité. La meilleure façon pour ces communes de donner de la crédibilité à leur affirmation est de présenter un bon indice ATMO (ou tout autre indice dans un pays autre que la France).

On pourrait être tenté de rejeter tout indice qui ne refléterait pas la réalité ou qui, de façon nécessairement arbitraire, instituerait une réalité. Toutefois, ceux-ci ne sont pas forcément inutiles. Un indicateur peut être considéré comme une sorte de langage, basé sur des conventions plus ou moins arbitraires et nous aidant à communiquer plus ou moins efficacement à propos de certains sujets ou à effectuer certaines tâches. Par "efficacement," nous entendons simplement "plus efficacement que sans langage;" pas nécessairement de la façon la plus efficace qui soit. Comme tout langage, il n'est pas toujours précis et laisse de la place pour des ambiguïtés et des contradictions avec lesquelles il faut apprendre à vivre. Si les créateurs du décathlon avaient attendu qu'une théorie bien fondée leur dise comment désigner le vainqueur (théorie qui ne viendra peut-être jamais), il est probable qu'aucun décathlon n'aurait jamais été organisé.

Mais ceci ne signifie pas que tout indicateur est bon. Les ambiguïtés et contradictions sont certainement souhaitables en poésie sinon nous ne pourrions jamais goûter ces vers :

Peut-il se reposer celui qui dort  
Il ne voit pas la nuit ne voit pas l'invisible  
Il a de grandes couvertures  
Et des coussins de sang sur des coussins de boue  
Sa tête est sous les toits et ses mains sont fermées  
Sur les outils de la fatigue

Il dort pour éprouver sa force  
La honte d'être aveugle dans un si grand silence  
  
Aux rivages que la mer rejette  
Il ne voit pas les poses silencieuses  
Du vent qui fait entrer l'homme dans ses statues  
Quand il s'apaise  
  
Bonne volonté du sommeil  
D'un bout l'autre de la mort.<sup>2</sup>

Mais, en matière d'aide à la décision, les ambiguïtés et contradictions doivent être évitées tant que faire se peut. Quand certains éléments de préférence sont connus avec certitude, n'importe quel indicateur devrait les refléter.

**Retour à l'aide multicritère à la décision** Dans un processus d'aide à la décision, les préférences du décideur ne sont pas complètement connues a priori. Sinon, on n'aurait pas besoin d'aide à la décision. S'appuyer uniquement sur la théorie du mesurage est, dès lors, impossible. Dans la plupart des processus d'aide à la décision, comme avec les indicateurs, une certaine part d'arbitraire est inévitable, que ce soit au niveau du choix de l'analyste, des critères ou de la méthode d'agrégation, pour ne citer que quelques exemples.

Mais à la différence de certains indicateurs qui sont construits en l'absence de tout problème de décision, la plupart des processus d'aide à la décision sont liés à un problème décisionnel plus ou moins bien défini. Certains éléments de préférence sont donc nécessairement présents. Par conséquent, si certains objets sont mesurés (si on leur associe des nombres) pendant le processus d'aide à la décision, la théorie du mesurage peut être utilisée pour garantir que le modèle construit pendant ce processus ne contredise aucun des éléments connus, qu'il les reflète tous et que toutes les conclusions qui peuvent être tirées de ces éléments le soient effectivement.

## 5 Conclusion

Dans ce texte, nous avons analysé trois indicateurs différents. Nous avons montré qu'ils présentent tous certaines faiblesses : peu ou trop de compensation, non-monotonie, incapacité de prendre en compte des dépendances entre dimensions, etc. Ces problèmes ne sont pas liés à ces trois indicateurs

---

<sup>2</sup>Paul Eluard, La vie immédiate [5]

particuliers. A peu près n'importe quel indicateur pourrait être critiqué de façon similaire. Les rangements induits par ces indicateurs sont donc, dans certains cas, en porte-à-faux avec les rangements que nous voudrions qu'ils induisent (comme dans tous nos exemples). Ceci est dû, en grande partie, au fait qu'ils sont construits de façon assez arbitraire sans tenter de tenir compte des préférences des utilisateurs. Ceci ne doit pas nous étonner puisqu'ils sont généralement construits sans référence à un problème particulier ou à un utilisateur.

Par ailleurs, nous avons montré qu'un modèle d'aide à la décision, et donc un indicateur, ne peut et ne doit pas forcément refléter la réalité ou, à tout le moins, que la réalité.

## Références

- [1] Atkinson, A.B. (1970) "On the measurement of inequality" *Journal of Economic Theory* 2, pp. 244–263. 4
- [2] Bouyssou, D., Marchant, T., Perny, P., Pirlot, M., Tsoukiàs, A. and Vincke, P. (2000) *Evaluation and decision models : a critical perspective*, Kluwer. 1
- [3] De Jongh, A. (1992) *Théorie du mesurage, agrégation des critères et application au décathlon*, Masters thesis, SMG, Université Libre de Bruxelles, Brussels. 16
- [4] Desrosières, A. (1995) "Réfléter ou instituer : L'invention des indicateurs statistiques", *Rapport technique 129/J310*, INSEE, Paris. 11, 22
- [5] Eluard, P. (1967) *La vie immédiate*, NRF, Gallimard (première parution en 1932) 23
- [6] Horn, R.V. (1993) *Statistical indicators*, Cambridge University Press, Cambridge. 2
- [7] Keeney, R.L. and Raiffa, H. (1976) *Decisions with multiple objectives : Preferences and value tradeoffs*, Wiley, New York. 19
- [8] Krantz, D.H., Luce, R.D., Suppes, P. and Tversky, A. (1971) *Foundations of measurement, Vol. 1 : Additive and polynomial representations*, Academic Press, New York. 20
- [9] Ott, W.R. (1978) *Environmental indices : Theory and practice*, Ann Arbor Science, Ann Arbor. 12
- [10] Programme des Nations Unies pour le Développement (1997) *Rapport mondial sur le développement humain*, Economica, Paris 3



- [11] Roberts, F.S. (1979) *Measurement theory, with applications to Decision Making, Utility and the Social Sciences*, Addison-Wesley 14
- [12] Roy, B. (1985). *Méthodologie multicritère d'aide à la décision*, Economica, Paris, 1985. 19
- [13] Roy, B. (1990) “Science de la décision ou science de l’aide à la décision?”, *Cahier du LAMSADE No 97*, Université Paris-Dauphine, Paris. 2
- [14] Vincke, Ph. (1989). *L’Aide Multicritère à la Décision*, Editions de l’Université de Bruxelles-Editions Ellipses, Bruxelles, 1989. 19
- [15] Warusfel, A. (1961) *Les nombres et leurs mystères*, Points Sciences, Seuil, Paris. 11
- [16] Zarnowsky, F. (1989) *The decathlon — A colorful history of track and field’s most challenging event*, Leisure Press, Champaign. 16

The number after a reference indicates on which page it is cited.