

Un Cadre Basé sur la Discrépance pour une Comparaison de la Robustesse entre Evaluations Multi-attributs

JUSTE RAIMBAULT^{1,2}

¹ UMR CNRS 8504 Géographie-Cités, Paris, France

² UMR-T IFSTTAR 9403 LVMT, Champs-sur-Marne, France
juste.raimbault@polytechnique.edu

Abstract. Les évaluations multi-objectifs sont un aspect essentiel de la gestion de systèmes complexes, puisque la complexité intrinsèque d'un système est généralement étroitement liée au nombre d'objectifs d'optimisation potentiels. Cependant, une évaluation ne fait pas sens si sa robustesse, au sens de sa fiabilité, n'est pas donnée. Les méthodes statistiques usuelles fournissant une mesure de robustesse sont très dépendantes des modèles sous-jacents. Nous proposons une formulation d'un cadre indépendant du modèle, dans le cas d'indicateurs intégrés et agrégés (évaluation multi-attributs), qui permet de définir une mesure de robustesse relative prenant en compte la structure des données et les valeurs des indicateurs. La méthode est testée sur données urbaines synthétiques associées aux arrondissements de Paris, et à des données réelles de revenus pour l'évaluation de la ségrégation urbaine dans la région métropolitaine du Grand Paris. Les premiers résultats numériques montrent les potentialités de cette nouvelle méthode. De plus, sa relative indépendance au type de système et au modèle pourrait la positionner comme une alternative aux méthodes statistiques classiques d'évaluation de la robustesse.

Keywords: Evaluation Multi-attributs, Robustesse Non-modèle-dépendante, Systèmes Urbains, Discrépance

1 Introduction

1.1 Contexte Général

Les problèmes multi-objectifs sont organiquement liés à la complexité des systèmes sous-jacents. En effet, que ce soit dans le champ des *Systèmes Complexes Industriels*, dans le sens de systèmes conçus par ingénierie, où la construction de Systèmes de Systèmes (SoS) par couplage et intégration induit souvent des objectifs contradictoires [1], ou dans le champ des *Systèmes Complexes Naturels*, au sens de systèmes non désignés, physiques, biologiques ou sociaux, qui présentent des propriétés d'émergence et d'auto-organisation, pour lesquels les objectifs peuvent e.g. être le résultat de l'interaction d'agents hétérogènes (voir [2] pour une revue étendue des types de systèmes concernés par cette approche),

l’optimisation multi-objectifs peut être explicitement introduite pour étudier ou désigner le système, mais régit généralement déjà implicitement les mécanismes internes du système. Le cas des Systèmes Complexes Sociaux-techniques est particulièrement intéressant puisque selon Haken [3], ils peuvent être vus comme des systèmes hybrides embarquant des agents sociaux dans des “artefacts techniques” (parfois jusqu’à un niveau inattendu, créant ce que PICON décrit comme *cyborgs* [4]), et cumulent ainsi la potentialité d’être à l’origine de problèmes multi-objectifs¹. La notion récente d’*éco-quartier* [5] est un exemple typique pour lequel la durabilité implique des objectifs contradictoires. L’exemple des systèmes de transport, dont la conception a glissé durant la seconde moitié du 20ème siècle d’analyses coût-bénéfices à la prise de décision multi-critères, est également typique de tels systèmes [6]. Les systèmes géographiques sont à présent bien étudiés d’un tel point de vue, en particulier grâce à l’intégration des cadres multi-objectifs au sein des Systèmes d’Information Géographiques [7]. Comme dans le cas microscopique des éco-quartiers, la planification et le design urbains mésoscopiques et macroscopiques peuvent être rendus durables grâce aux évaluations par indicateurs [8].

Un aspect crucial de l’évaluation est une certaine notion de sa fiabilité, que nous nommerons ici *robustesse*. Les méthodes statistiques incluent naturellement cette notion puisque la construction et l’estimation de modèles statistiques donne divers indicateurs de la consistance des résultats [9]. Le premier exemple venant à l’esprit est l’application de la loi des grands nombres pour obtenir la *p-valeur* d’une estimation de modèle, qui peut être interprété comme une mesure de confiance en les valeurs estimées. D’autre part, les intervalles de confiance et le *beta-power* sont d’autres indicateurs importants de robustesse statistique. L’inférence bayésienne fournit également des mesures de robustesse quand la distribution des paramètres est estimée de manière séquentielle. Concernant les optimisations multi-objectifs, en particulier par des algorithmes heuristiques (comme par exemple les algorithmes génétiques, ou les solveurs de recherche opérationnelle), la notion de robustesse d’une solution consiste plus en la stabilité de la solution dans l’espace des phases du système dynamique correspondant. Des progrès récents ont été faits vers une formulation unifiée de la robustesse pour les problèmes d’optimisation multi-objectifs, comme dans [10] où les fronts de Pareto robustes sont définis comme des solutions insensibles aux petites perturbations. Dans [11], la notion de degré de robustesse est introduite, formalisée comme une sorte de continuité des autres solutions dans des voisinages successifs d’une solution.

Cependant, il n’existe pas de méthode générique qui permettrait une évaluation de la robustesse de façon indépendante au modèle, i.e. qui serait extraite de la structure des données et des indicateurs mais ne dépendrait pas de la méthode utilisée. Un avantage serait par exemple une estimation *a priori* de la robustesse

¹ Nous désignons ici par *Evaluation Multi-objectifs* toutes les pratiques incluant le calcul de multiples indicateurs d’un système (il peut s’agir d’optimisation multi-objectif pour un design de système, une évaluation multi-objectif d’un système existant, une évaluation multi-attributs ; notre cadre particulier correspondra au dernier cas).

potentielle d'une évaluation et de décider ainsi si elle vaut la peine d'être faite. Nous proposons un cadre répondant à cette contrainte dans le cas particulier des évaluations multi-attributs, i.e. quand le problème est rendu unidimensionnel par agrégation des objectifs. Il est basé sur les données et non sur les modèles, au sens où l'estimation de la robustesse ne dépendra pas de la manière dont les indicateurs sont calculés, tant qu'ils respectent certaines hypothèses détaillées par la suite.

1.2 Approche Proposée

Objectifs comme Intégrales Spatiales Nous supposons que les objectifs peuvent être exprimés comme intégrales spatiales, ce qui devrait s'appliquer à tout système territorial, et nos cas d'application sont des systèmes urbains. Ce n'est pas si restrictif en terme d'indicateurs possibles si l'on utilise les bonnes variables et noyaux intégrés : de façon analogue à la méthode de Regression Géographique Pondérée [12], toute variable spatiale peut être intégrée contre des noyaux réguliers de taille variable et le résultats sera une agrégation spatiale dont la signification dépendra de l'étendue du noyau. Les exemples utilisés par la suite comme des moyennes conditionnelles ou des sommes vérifient parfaitement cette hypothèse. Même un indicateur déjà agrégé dans l'espace peut être interprété comme une intégrale spatiale en utilisant une distribution de Dirac au centroïde de la zone correspondante.

Objectifs Agrégés Linéairement Une seconde hypothèse que nous faisons est que l'évaluation multi-objectifs est effectuée par agrégation linéaire des objectifs, c'est à dire qu'on se place dans le cadre d'un problème d'optimisation multi-attributs. Si $(q_i(\mathbf{x}))_i$ sont les valeurs des fonctions objectifs, on définit alors des poids $(w_i)_i$ afin de construire la fonction de prise de décision $q(\mathbf{x}) = \sum_i w_i q_i(\mathbf{x})$, dont la valeur détermine ensuite la performance d'une solution. Cette approche est analogue aux utilités agrégées en économie et est utilisée dans de nombreux domaines. La subtilité réside dans le choix des poids, i.e. de la forme de la fonction de projection, et différentes solutions ont été développées pour obtenir des poids selon la nature du problème. Récemment, [13] a proposé de comparer la robustesse des différentes techniques d'agrégation par une analyse de sensibilité, effectuée par simulations de Monte-Carlo pour produire des données synthétiques, ce qui permet d'obtenir la distribution des biais pour les différentes techniques, certaines étant significativement plus performantes que d'autres. Toutefois, la quantification de la robustesse dépend toujours des modèles utilisés dans ce travail.

Le reste de cette monographie est organisé de la façon suivante : la section 2 décrit intuitivement puis mathématiquement le cadre proposé ; la section 3 détaille ensuite l'implémentation, la collecte des données pour les cas d'étude et les résultats numériques pour une évaluation intra-urbaine synthétique et un cas réel métropolitain ; la section 4 discute finalement les limitations et les potentialités de la méthode.

2 Description du Cadre

2.1 Description Intuitive

Nous décrivons à présent le cadre proposé pour permettre théoriquement de comparer la robustesse d'évaluation de deux systèmes urbains différents. Ce cadre est une généralisation d'une méthode empirique proposée dans [14] pour accompagner une étude dans un autre contexte effectuant une comparaison du sens et de la pertinence des indicateurs dans un contexte de durabilité. Intuitivement, la base empirique se base sur les principes suivants :

- Les systèmes urbains peuvent être vus selon l'information disponible, i.e. les données brutes décrivant le système. Dans une approche basée sur les données, celles-ci sont la base de notre cadre et la robustesse sera déterminée par leur structure.
- A partir des données sont capturés des indicateurs (fonctions objectifs). Nous supposons qu'un choix d'indicateurs est une intention particulière de traduire des aspects particuliers du système, i.e. de capturer une réalisation d'un "fait urbain" au sens de MANGIN [15] - une sorte de fait stylisé en terme de processus et de mécanismes, ayant différentes réalisations sur des systèmes distincts dans l'espace, dépendant de chaque contexte géographique précis.
- Etant donné plusieurs systèmes et indicateurs associés, un espace commun peut être construit pour les comparer. Dans cet espace, les données représentent plus ou moins bien le système réel, c'est à dire qu'elles sont imprécises en fonction de l'échelle initiale, de la précision effective des données. Nous proposons de capturer exactement ces différents aspects au travers de la notion de discrédance d'un nuage de points, qui est un outil mathématique provenant des théories d'échantillonnage, permettant d'exprimer la façon dont un jeu de données remplit l'espace dans lequel il s'insère [16].

Synthétisant ces contraintes, nous proposons une notion de *Robustesse* d'une évaluation qui capture à la fois, en combinant la fiabilité des données à l'importance relative des indicateurs,

1. *Données manquantes* : une évaluation se basant sur des jeux de données plus raffinés sera naturellement plus robuste.
2. *Importance des indicateurs* : les indicateurs avec plus d'importance relative pèseront plus dans la robustesse totale.

2.2 Description Formelle

Indicateurs Soit $(S_i)_{1 \leq i \leq N}$ un nombre fini de systèmes territoriaux géographiquement disjoints, que nous supposons décrits par les données brutes et des indicateurs intermédiaires, donnés par $S_i = (\mathbf{X}_i, \mathbf{Y}_i) \in \mathcal{X}_i \times \mathcal{Y}_i$ avec $\mathcal{X}_i = \prod_k \mathcal{X}_{i,k}$ tel que chaque sous-espace contient des matrices réelles : $\mathcal{X}_{i,k} = \mathbb{R}^{n_{i,k}^X \times p_{i,k}^X}$ (de la même façon pour \mathcal{Y}_i). Nous définissons également une fonction d'indice ontologique $I_X(i, k)$ (resp. $I_Y(i, k)$) prenant des valeurs entières qui coïncident si

et seulement si les deux variables ont même ontologie au sens de [17], c'est à dire qu'elles sont supposées représenter le même objet réel. On distingue les “données brutes” \mathbf{X}_i à partir desquelles les indicateurs sont calculés généralement par des fonctions déterministes explicites, des “indicateurs intermédiaires” \mathbf{Y}_i qui sont déjà intégrés et peuvent être par exemple les sorties de modèles élaborés simulant certains aspects du système urbain. Nous définissons l'espace caractéristique du “fait urbain” par

$$(\mathcal{X}, \mathcal{Y}) \stackrel{\text{def}}{=} \left(\prod \tilde{\mathcal{X}}_c \right) \times \left(\prod \tilde{\mathcal{Y}}_c \right) = \left(\prod_{\mathcal{X}_{i,k} \in \mathcal{D}_{\mathcal{X}}} \mathbb{R}^{p_{i,k}^X} \right) \times \left(\prod_{\mathcal{Y}_{i,k} \in \mathcal{D}_{\mathcal{Y}}} \mathbb{R}^{p_{i,k}^Y} \right) \quad (1)$$

avec $\mathcal{D}_{\mathcal{X}} = \{\mathcal{X}_{i,k} | I(i, k) \text{ distincts}, n_{i,k}^X \text{ maximal}\}$ (de même pour \mathcal{Y}_i). Il s'agit en fait de l'espace abstrait sur lequel les indicateurs sont intégrés. Les indices c introduit par définition correspondent aux différents indicateurs au sein des différents systèmes. Cette espace est l'espace minimal commun à tous les systèmes permettant une définition commune des indicateurs pour tous.

Soit $\mathbf{X}_{i,c}$ les données projetées canoniquement sur le sous-espace correspondant, bien définies pour tout i et tout c . Nous faisons donc l'hypothèse clé que tous les indicateurs sont calculés par intégration contre un noyau donné, i.e. pour tout c il existe H_c espace de fonctions à valeurs réelles sur $(\tilde{\mathcal{X}}_c, \tilde{\mathcal{Y}}_c)$, tel que pour tout $h \in H_c$:

1. h est “suffisamment” régulière (distribution tempérée par exemple)
2. $q_c = \int_{(\tilde{\mathcal{X}}_c, \tilde{\mathcal{Y}}_c)} h$ est une fonction décrivant le “fait urbain” (l'indicateur en lui-même)

Des exemples typiques de noyaux peuvent être :

- Une moyenne des lignes de $\mathbf{X}_{i,c}$ est calculée par $h(x) = x \cdot f_{i,c}(x)$ où $f_{i,c}$ est la densité de la distribution de la variable sous-jacente.
- Un taux d'éléments du jeu de données respectant une condition donnée C , $h(x) = f_{i,c}(x) \chi_C(x)$.
- Pour des variables déjà agrégées \mathbf{Y} , une distribution de Dirac permet des les exprimer également comme des intégrales de noyaux.

Agrégation La détermination des poids est en fait le point crucial des processus de prise de décision multi-attributs, et de nombreuses méthodes sont disponibles (voir [18] pour une revue dans le cas particulier de la gestion de l'énergie durable). Définissons les poids pour l'agrégation linéaire. Nous supposons les indicateurs normalisés, i.e. $q_c \in [0, 1]$, pour une construction plus simple des poids relatifs. Pour i, c et $h_c \in H_c$ donnés, le poids $w_{i,c}$ est simplement constitué par l'importance relative de l'indicateur $w_{i,c}^L = \frac{\hat{q}_{i,c}}{\sum_c \hat{q}_{i,c}}$ où $\hat{q}_{i,c}$ est un estimateur de q_c pour les données $\mathbf{X}_{i,c}$ (i.e. la valeur calculée effectivement). On peut noter que cette étape n'est pas contraignante et que cela peut être étendu à tout ensemble d'attribution de poids, en prenant par exemple $\tilde{w}_{i,c} = w_{i,c} \cdot w'_{i,c}$ si \mathbf{w}' sont les poids fixés par le preneur de décisions. Nous nous concentrons sur l'influence relative des attributs et pour cela choisissons cette forme simple pour les poids.

Estimation de la Robustesse La scène est à présent prête pour permettre d’estimer la robustesse d’une évaluation faite par la fonction d’agrégation. Pour cela, nous appliquons un théorème d’approximation d’intégrale similaire aux méthodes introduites dans [19], puisque la forme intégrée des indicateurs permet justement de bénéficier de tels résultats théoriquement puissants. Soit $\mathbf{X}_{i,c} = (\mathbf{X}_{i,c,l})_{1 \leq l \leq n_{i,c}}$ et $D_{i,c} = \text{Disc}_{\tilde{\chi}_c, L^2}(\mathbf{X}_{i,c})$ la discrépance du jeu de données² [20]. Avec $h \in H_c$, on a la borne supérieure sur l’erreur d’approximation de l’intégrale

$$\left\| \int h_c - \frac{1}{n_{i,c}} \sum_l h_c(\mathbf{X}_{i,c,l}) \right\| \leq K \cdot \|h_c\| \cdot D_{i,c}$$

où K est une constante indépendante des points de données et des fonctions objectifs. Cela donne directement

$$\left\| \int \sum w_{i,c} h_c - \frac{1}{n_{i,c}} \sum_l w_{i,c} h_c(\mathbf{X}_{i,c,l}) \right\| \leq K \sum_c |w_{i,c}| \|h_c\| \cdot D_{i,c}$$

En supposant l’erreur réalisée de manière raisonnable (scénario du “pire de cas” pour la connaissance de la valeur théorique de la fonction agrégée), nous prenons cette borne supérieure comme une approximation de sa magnitude. De plus, la normalisation des indicateurs implique que $\|h_c\| = 1$. Nous proposons alors de comparer les bornes d’erreurs entre deux évaluations. Elle dépendent seulement de la distribution des données (équivalence à la *robustesse statistique*) et des indicateurs choisis (sorte de *robustesse ontologique*, i.e. est-ce que les indicateurs ont un sens réel dans le contexte choisi et est-ce que leur valeur fait sens), et sont un moyen de combiner ces deux types de robustesse dans une seule valeur.

Nous définissons ainsi un *ratio de robustesse* pour comparer la robustesse de deux évaluations par

$$R_{i,i'} = \frac{\sum_c w_{i,c} \cdot D_{i,c}}{\sum_c w_{i',c} \cdot D_{i',c}} \quad (2)$$

L’interprétation intuitive de cette définition est que l’on compare la robustesse des évaluations en comparant la plus grande erreur faite dans chaque cas selon la structure des données et l’importance relative.

En construisant une relation d’ordre sur les évaluations en comparant la position du ratio par rapport à un, il est clair qu’on obtient un ordre complet sur l’ensemble des évaluations possibles. Ce ratio devrait en théorie permettre de comparer n’importe quelle évaluation d’un système urbain. Afin de garder un

² La discrépance est définie comme la norme- L^2 de la discrépance locale qui est pour des points de données normalisés $\mathbf{X} = (x_{ij}) \in [0, 1]^d$, une fonction de $\mathbf{t} \in [0, 1]^d$ comparant le nombre de points compris dans le volume de l’hypercube correspondant, donné par $\text{disc}(\mathbf{t}) = \frac{1}{n} \sum_i \mathbb{1}_{\prod_j x_{ij} < t_j} - \prod_j t_j$. C’est une mesure de la manière dont le nuage de points couvre l’espace.

sens ontologique à cela, il devrait être utilisé pour comparer des sous-systèmes disjoints avec une proportion raisonnable d’indicateurs en commun, ou le même sous-système avec des indicateurs différents. On peut noter que cela fournit un moyen de tester l’influence des indicateurs sur une évaluation, en analysant la sensibilité du ratio à leur suppression. Au contraire, la détermination d’un nombre “minimal” d’indicateurs faisant chacun varier le ratio fortement pourrait être un moyen d’isoler des paramètres essentiels régissant le sous-système.

3 Résultats

Implémentation Le pré-traitement des données géographiques est fait via QGIS [21] pour des raisons de performances. L’implémentation du coeur est faite en R [22] pour la flexibilité de la gestion des données et du traitement statistique. De plus, le package `DiceDesign` [23] conçu pour les expériences numériques et l’échantillonnage, permet un calcul efficient et direct des discrécances. Enfin, tout aussi important, l’ensemble du code source est disponible de manière ouverte sur le dépôt `git` du projet³ pour permettre la reproductibilité et la réutilisation [24].

3.1 Implémentation sur des Données Synthétiques

Nous proposons dans un premier temps d’illustrer l’implémentation par une application à des données et indicateurs synthétiques, pour des indicateurs de qualité de vie intra-urbaine pour la ville de Paris.

Données Brutes Le cas virtuel se base sur des données géographiques réelle, en particulier pour les arrondissements parisiens. Nous utilisons les données disponibles par le projet OpenStreetMap [25] qui fournit déjà des données précises en haute définition pour de nombreux aspects urbains. Nous utilisons le réseau de rues et la position des bâtiments dans la ville de Paris. Les limites des arrondissements, utilisées pour agréger et extraire les features lorsqu’on travaille sur un seul district, sont aussi pris de la même source. Nous utilisons les centroïdes des polygones des bâtiments et les segments du réseau de rues. Le jeu de données brutes consiste d’environ 200k bâtiments et 100k segments de rues.

Indicateurs synthétiques Nous travaillons sur chaque arrondissement de Paris (du 1er au 20ème) comme un système urbain évalué. Des données synthétiques aléatoires sont associées aux features spatiales, chaque arrondissement pouvant alors être évalué de manière stochastique, et des répétitions permettent d’obtenir le comportement statistique moyen des indicateurs jouets et des ratios de robustesse. Les indicateurs choisis doivent être calculés comme des indicateurs résidentiels et du réseau de rues. Pour montrer différents exemples, nous implémentons deux kernels moyens et une moyenne conditionnelle, tous liés à la durabilité environnementale et la qualité de vie, chacun devant être maximisés.

³ à <https://github.com/JusteRaimbault/RobustnessDiscrepancy>

On peut noter que ces indicateurs ont un sens réel mais pas de raison particulière d'être agrégés, ils sont ici choisis pour l'aspect pratique du modèle jouet et de la génération de données synthétiques. Avec $a \in \{1 \dots 20\}$ le nombre d'arrondissements, $A(a)$ l'aire spatiale correspondante à chacun, $b \in B$ les coordonnées des bâtiments et $s \in S$ les segments de rues, nous prenons

- Le complémentaire de la distance journalière moyenne au travail en voiture par individu, approché par, avec $n_{cars}(b)$ nombre de voiture dans le bâtiment (généré aléatoirement en associant des voiture à bâtiments proportionnel au taux de motorisation attendu α_m 0.4 à Paris), d_w distance des individus à leur travail (généré à partir du bâtiment vers un point aléatoire distribué uniformément dans l'étendue spatiale du jeu de données), et d_{max} le diamètre de l'aire de Paris, $\bar{d}_w = 1 - \frac{1}{|b \in A(a)|} \cdot \sum_{b \in A(a)} n_{cars}(b) \cdot \frac{d_w}{d_{max}}$
- Le complémentaire des flots moyens de voitures des rues dans la zone, approché par, avec $\varphi(s)$ flot relatif dans le segment de rue s , généré par le minimum entre 1 et une distribution log-normale ajustée pour avoir 95% de masse plus petite que 1, ce qui mimique la distribution hiérarchique de l'utilisation des rues (qui correspond à la centralité de chemin), et $l(s)$ longueur du segment, $\bar{\varphi} = 1 - \frac{1}{|s \in A(a)|} \cdot \sum_{s \in A(a)} \varphi(s) \cdot \frac{l(s)}{\max(l(s))}$
- Longueur relative de rues piétonnes \bar{p} , calculé via une dummy variable aléatoire uniforme ajustée pour obtenir une proportion fixée de segments piétonne.

Comme les données synthétiques sont stochastiques, les simulations sont lancées pour chaque quartier $N = 50$ fois, ce qui était un compromis raisonnable entre convergence statistique et temps nécessaire au calcul. La table 1 montre les résultats (moyennes et déviations standard) des valeurs des indicateurs et le calcul du ratio de robustesse. Les déviations standard obtenues confirment que ce nombre de simulations donnent des résultats consistants. Les indicateurs obtenus en fixant un ratio fixe montre peu de variabilité, ce qui peut être une limite de cette approche jouet. On obtient toutefois le résultat intéressant que la majorité des arrondissements donne des évaluations plus robustes que le 1er arrondissement, ce qui était attendu par la taille et la fonction de ce quartier: il s'agit en effet d'un petit quartier avec de grand bâtiment administratifs, ce qui implique moins d'éléments spatiaux et pour cela une évaluation moins robuste selon la définition qu'on en a donnée.

3.2 Application à un Cas Réel: Ségrégation Métropolitaine

Le premier exemple avait pour but de montrer les potentialités de la méthode mais était purement synthétique, ne pouvant pour cela fournir pas de conclusion concrète ni d'implications pour la gouvernance. Nous proposons maintenant de l'appliquer à des données réelles dans le cas de la ségrégation métropolitaine.

Données Nous travaillons sur les données de revenus, disponible pour la France à un niveau intra-urbain (unités statistiques élémentaires IRIS) pour l'année 2011

Arrdt	$< \bar{d}_w > \pm \sigma(\bar{d}_w)$	$< \bar{\varphi} > \pm \sigma(\bar{\varphi})$	$< \bar{p} > \pm \sigma(\bar{p})$	$R_{i,1}$
1er	0.731655 ± 0.041099	0.917462 ± 0.026637	0.191615 ± 0.052142	1.000000 ± 0.000000
2ème	0.723225 ± 0.032539	0.844350 ± 0.036085	0.209467 ± 0.058675	1.002098 ± 0.039972
3ème	0.713716 ± 0.044789	0.797313 ± 0.057480	0.185541 ± 0.065089	0.999341 ± 0.048825
4ème	0.712394 ± 0.042897	0.861635 ± 0.030859	0.201236 ± 0.044395	0.973045 ± 0.036993
5ème	0.715557 ± 0.026328	0.894675 ± 0.020730	0.209965 ± 0.050093	0.963466 ± 0.040722
6ème	0.733249 ± 0.026890	0.875613 ± 0.029169	0.206690 ± 0.054850	0.990676 ± 0.031666
7ème	0.719775 ± 0.029072	0.891861 ± 0.026695	0.209265 ± 0.041337	0.966103 ± 0.037132
8ème	0.713602 ± 0.034423	0.931776 ± 0.015356	0.208923 ± 0.036814	0.973975 ± 0.033809
9ème	0.712441 ± 0.027587	0.910817 ± 0.015915	0.202283 ± 0.049044	0.971889 ± 0.035381
10ème	0.713072 ± 0.028918	0.881710 ± 0.021668	0.210118 ± 0.040435	0.991036 ± 0.038942
11ème	0.682905 ± 0.034225	0.875217 ± 0.019678	0.203195 ± 0.047049	0.949828 ± 0.035122
12ème	0.646328 ± 0.039668	0.920086 ± 0.019238	0.198986 ± 0.023012	0.960192 ± 0.034854
13ème	0.697512 ± 0.025461	0.890253 ± 0.022778	0.201406 ± 0.030348	0.960534 ± 0.033730
14ème	0.703224 ± 0.019900	0.902898 ± 0.019830	0.205575 ± 0.038635	0.932755 ± 0.033616
15ème	0.692050 ± 0.027536	0.891654 ± 0.018239	0.200860 ± 0.024085	0.929006 ± 0.031675
16ème	0.654609 ± 0.028141	0.928181 ± 0.013477	0.202355 ± 0.017180	0.963143 ± 0.033232
17ème	0.683020 ± 0.025644	0.890392 ± 0.023586	0.198464 ± 0.033714	0.941025 ± 0.034951
18ème	0.699170 ± 0.025487	0.911382 ± 0.027290	0.188802 ± 0.036537	0.950874 ± 0.028669
19ème	0.655108 ± 0.031857	0.884214 ± 0.027816	0.209234 ± 0.032466	0.962966 ± 0.034187
20ème	0.637446 ± 0.032562	0.873755 ± 0.036792	0.196807 ± 0.026001	0.952410 ± 0.038702

Table 1. Résultats numériques des simulations pour chaque arrondissement avec $N = 50$ répétitions. Chaque valeur des indicateurs factice est donnée par sa moyenne sur les répétitions et la déviation standard associée. Le ratio de robustesse est calculé par rapport au premier arrondissement (choix arbitraire). Un ratio inférieur à 1 signifie que la borne de l'intégrale est plus petite pour le premier système, i.e. que l'évaluation est plus robuste pour celui-ci. A cause de la petite taille du 1er arrondissement, on s'attend que la majorité des arrondissements aient un ratio plus petit que 1, ce qui se confirme dans les résultats même lorsque l'on ajoute la déviation standard.

sous la forme de résumé statistiques (déciles uniquement si la zone est peuplée suffisamment pour assurer l'anonymat), fournies par l'INSEE⁴. Les données sont associées à l'étendue géographique des unités statistiques, permettant le calcul d'indicateurs d'analyse spatiale.

Indicateurs Nous utilisons ici trois indicateurs de ségrégation intégrés sur une zone géographique. Supposons la zone divisée en unités couvrantes \mathcal{S}_i pour $1 \leq$

⁴ <http://www.insee.fr>

$i \leq N$ avec pour centroïdes (x_i, y_i) . Chaque unité a des caractéristiques de population P_i et de revenu médian X_i

We use here three indicators of segregation integrated on a geographical area. Let assume the area divided into covering units \mathcal{S}_i for $1 \leq i \leq N$ with centroids (x_i, y_i) . Each unit has characteristics of population P_i and median income X_i . We define spatial weights used to quantify strength of geographical interactions between units i, j , with d_{ij} euclidian distance between centroids : $w_{ij} = \frac{P_i P_j}{(\sum_k P_k)^2} \cdot \frac{1}{d_{ij}}$ if $i \neq j$ and $w_{ii} = 0$. The normalized indicators are the following

- Spatial autocorrelation Moran index, defined as weighted normalized covariance of median income by $\rho = \frac{N}{\sum_{ij} w_{ij}} \cdot \frac{\sum_{ij} w_{ij} (X_i - \bar{X})(X_j - \bar{X})}{\sum_i (X_i - \bar{X})^2}$
- Dissimilarity index (close to Moran but integrating local dissimilarities rather than correlations), given by $d = \frac{1}{\sum_{ij} w_{ij}} \sum_{ij} w_{ij} |\tilde{X}_i - \tilde{X}_j|$
with $\tilde{X}_i = \frac{X_i - \min(X_k)}{\max(X_k) - \min(X_k)}$
- Complementary of the entropy of income distribution that is a way to capture global inequalities $\varepsilon = 1 + \frac{1}{\log(N)} \sum_i \frac{X_i}{\sum_k X_k} \cdot \log\left(\frac{X_i}{\sum_k X_k}\right)$

Numerous measures of segregation with various meanings and at different scales are available, as for example at the level of the unit by comparison of empirical wage distribution with a theoretical null model [26]. The choice here is arbitrary in order to illustrate our method with a reasonable number of dimensions.

Results We apply our method with these indicators on the Greater Paris area, constituted of four *départements* that are intermediate administrative units. The recent creation of a new metropolitan governance system [27] underlines interrogations on its consistence, and in particular on its relation to intermediate spatial inequalities. We show in Fig. 1 maps of spatial distribution of median income and corresponding local index of autocorrelation. We observe the well-known West-East opposition and district disparities inside Paris as they were formulated in various studies, such as [28] through the analysis of real estate transactions dynamics. We then apply our framework to answer a concrete question that has implications for urban policy : *how are the evaluation of segregation within different territories sensitive to missing data ?* To do so, we proceed to Monte Carlo simulations (75 repetitions) during which a fixed proportion of data is randomly removed, and the corresponding robustness index is evaluated with renormalized indicators. Simulations are done on each *department* separately, each time relatively to the robustness of the evaluation of full Greater Paris. Results are shown in Fig. 2. All areas present a slightly better robustness than the reference, what could be explained by local homogeneity and thus more fiable segregation values. Implications for policy that can be drawn are for example direct comparisons between areas : a loss of 30% of information on 93 area corresponds to a loss of only 25% in 92 area. The first being a deprived area, the inequality is

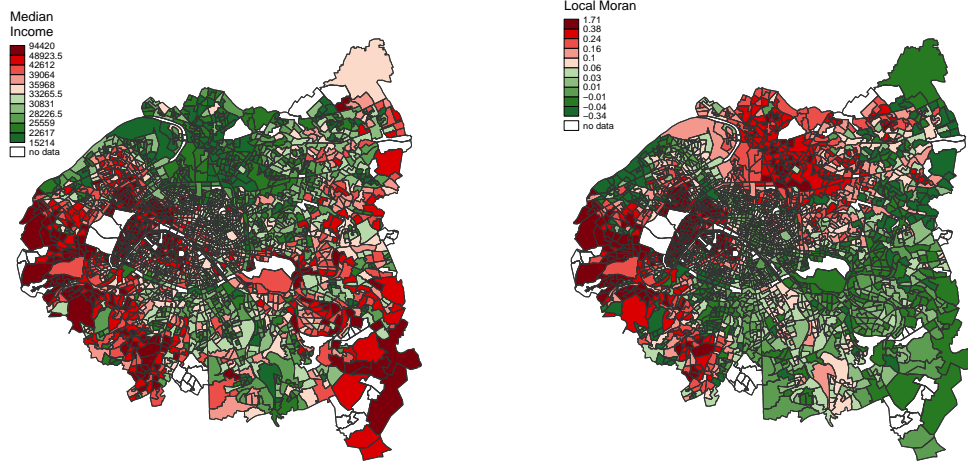


Fig. 1. Maps of Metropolitan Segregation. Maps show yearly median income on basic statistical units (IRIS) for the three departments constituting mainly the Great Paris metropolitan area, and the corresponding local Moran spatial autocorrelation index, defined for unit i as $\rho_i = N / \sum_j w_{ij} \cdot \frac{\sum_j w_{ij} (X_j - \bar{X})(X_i - \bar{X})}{\sum_i (X_i - \bar{X})^2}$. The most segregated areas coincide with the richest and the poorest, suggesting an increase of segregation in extreme situations.

increased by this relative lower quality of statistical information. The study of standard deviations suggest further investigations as different response regimes to data removal seem to exist.

4 Discussion

4.1 Applicability to Real situations

Implications for Decision-making The application of our method to concrete decision-making can be thought in different ways. First in the case of a comparative multi-attribute decision process, such as the determination of a transportation corridor, the identification of territories on which the evaluation may be flawed (i.e. has a poor relative robustness) could allow a more refined focus on these and a corresponding revision of datasets or an adapted revision of weights. In any case the overall decision-making process should be made more reliable. A second direction lays in the spirit of the real application we have proposed, i.e. the sensitivity of evaluation to various parameters such as missing data. If a decision appears as reliable because data have few missing points, but the evaluation is very sensitive to it, one will be more careful in the interpretation of

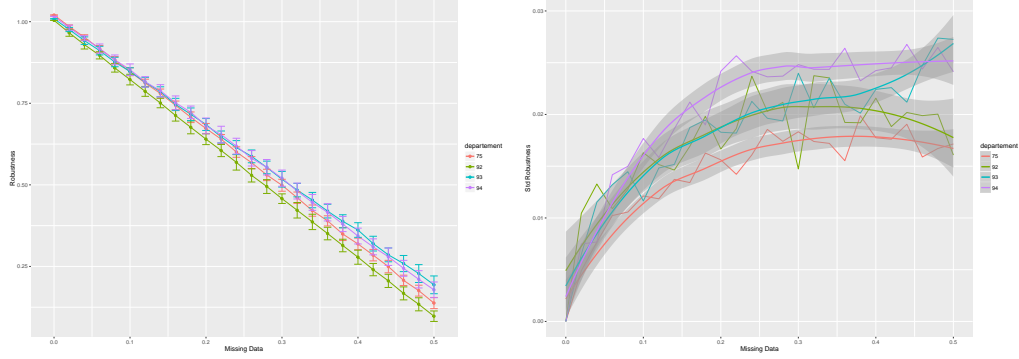


Fig. 2. Sensitivity of robustness to missing data. *Left.* For each department, Monte Carlo simulations ($N=75$ repetitions) are used to determine the impact of missing data on robustness of segregation evaluation. Robustness ratios are all computed relatively to full metropolitan area with all available data. Quasi-linear behavior translates an approximative linear decrease of discrepancy as a function of data size. The similar trajectory of poorest departments (93,94) suggest the correction to linear behavior being driven by segregation patterns. *Right.* Corresponding standard deviations of robustness ratios. Different regimes (in particular 93 against others) unveil phase transitions at different levels of missing data, meaning that the evaluation in 94 is from this point of view more sensitive to missing data.

results and taking the final decision. Further work and testing will however be needed to understand framework behavior in different contexts and be able to pilot its application in various real situations.

Integration Within Existing Frameworks The applicability of the method on real cases will directly depend on its potential integration within existing framework. Beyond technical difficulties that will surely appear when trying to couple or integrate implementations, more theoretical obstacles could occur, such as fuzzy formulations of functions or data types, consistency issues in databases, etc. Such multi-criteria framework are numerous. Further interesting work would be to attempt integration into an open one, such as e.g. the one described in [29] which calculates various indices of urban segregation, as we have already illustrated the application on metropolitan segregation indexes.

Availability of Raw Data In general, sensitive data such as transportation questionnaires, or very fine granularity census data are not openly available but provided already aggregated at a certain level (for instance French Insee Data are publicly available at basic statistical unit level or larger areas depending on variables and minimal population constraints, more precise data is under restricted access). It means that applying the framework may imply complicated data research procedure, its advantage to be flexible being thus reduced through additional constraints.

4.2 Validity of Theoretical Assumptions

A possible limitation of our approach is the validity of the assumption formulating indicators as spatial integrals. Indeed, many socio-economic indicators are not necessarily depending explicitly on space, and trying to associate them with spatial coordinates may become a slippery slope (e.g. associate individual economic variables with individual residential coordinates will have a sense only if the use of the variable has a relation with space, otherwise it is a non-legitimate artifact). Even indicators which have a spatial value may derive from non-spatial variables, as [30] points out concerning accessibility, when opposing integrated accessibility measures with individual-based non necessarily spatial-based (e.g. individual decisions) measures. Constraining a theoretical representation of a system to fit a framework by changing some of its ontological properties (always in the sense of real meaning of objects) can be understood as a violation of a fundamental rule of modeling and simulation in social science given in [31], that is that there can be an universal “language” for modeling and some can not express some systems, having for consequence misleading conclusion due to ontology breaking in the case of an over-constrained formulation.

4.3 Framework Generality

We argue that the fundamental advantage of the proposed framework is its generality and flexibility, since robustness of the evaluations are obtained only through data structure if ones relaxes constraints on the value of weight. Further work should go towards a more general formulation, suppressing for example the linear aggregation assumption. Non-linear aggregation functions would require however to present particular properties regarding integral inequalities. For example, similar results could search in the direction of integral inequalities for Lipschitzian functions such as the one-dimensional results of [32].

Conclusion

We have proposed a model-independent framework to compare the robustness of multi-attribute evaluations between different urban systems. Based on data discrepancy, it provide a general definition of relative robustness without any assumption on model for the system, but with limiting assumptions that are the need of linear aggregation and of indicators being expressed through spatial kernel integrals. We propose a toy implementation based on real data for the city of Paris, numerical results confirming general expected behavior, and an implementation on real data for income segregation on Greater Paris metropolitan areas, giving possible insights into concrete policy questions. Further work should be oriented towards sensitivity analysis of the method, application to other real cases and theoretical assumptions relaxation, i.e. the relaxation of linear aggregation and spatial integration.

Acknowledgments

The author would like to thank Julien Keutchayan (Ecole Polytechnique de Montréal) for suggesting the original idea of using discrepancy, and anonymous reviewers for the useful comments and insights.

References

1. R Timothy Marler and Jasbir S Arora. Survey of multi-objective optimization methods for engineering. *Structural and multidisciplinary optimization*, 26(6):369–395, 2004.
2. MEJ Newman. Complex systems: A survey. *arXiv preprint arXiv:1112.1440*, 2011.
3. Herman Haken and Juval Portugali. The face of the city is its information. *Journal of Environmental Psychology*, 23(4):385–408, 2003.
4. Antoine Picon. *Smart cities: théorie et critique d’un idéal auto-réalisateur*. B2, 2013.
5. Taoufik Souami. *Ecoquartiers: secrets de fabrication*. Scrineo, 2012.
6. Jean-Jacques Bavoux, Francis Beaucire, Laurent Chapelon, and Pierre Zembri. *Géographie des transports*. Paris, 2005.
7. Stephen J Carver. Integrating multi-criteria evaluation with geographical information systems. *International Journal of Geographical Information System*, 5(3):321–339, 1991.
8. Anne Jégou, Vincent Augiseau, Cécile Guyot, Cécile Judéaux, François-Xavier Monaco, Pierre Pech, et al. L’évaluation par indicateurs: un outil nécessaire d’aménagement urbain durable?. réflexions à partir de la démarche parisienne pour le géographe et l’aménageur. *Cybergeo: European Journal of Geography*, 2012.
9. Robert L Launer and Graham N Wilkinson. *Robustness in statistics*. Academic Press, 2014.
10. Kalyanmoy Deb and Himanshu Gupta. Introducing robustness in multi-objective optimization. *Evolutionary Computation*, 14(4):463–494, 2006.
11. C. Barrico and C.H. Antunes. Robustness analysis in multi-objective optimization using a degree of robustness concept. In *Evolutionary Computation, 2006. CEC 2006. IEEE Congress on*, pages 1887–1892, 2006.
12. Chris Brunsdon, Stewart Fotheringham, and Martin Charlton. Geographically weighted regression. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(3):431–443, 1998.
13. Melissa J Dobbie and David Dail. Robustness and sensitivity of weighting and aggregation in constructing composite indices. *Ecological Indicators*, 29:270–277, 2013.
14. A. Ali, I. Carneiro, L. Dussarps, F. Guédel, E Lamy, J. Raimbault, L. Viger, V. Cohen, T. Aw, and S. Sadeghian. Les eco-quartiers lus par la mobilité : vers une évaluation intégrée. Technical report, Ecole des Ponts ParisTech, June 2014.
15. David Mangin and Philippe Panerai. *Projet urbain*. Parenthèses, 1999.
16. Josef Dick and Friedrich Pillichshammer. *Digital nets and sequences: Discrepancy Theory and Quasi-Monte Carlo Integration*. Cambridge University Press, 2010.
17. Pierre Livet, Jean-Pierre Muller, Denis Phan, and Lena Sanders. Ontology, a mediator for agent-based modeling in social science. *Journal of Artificial Societies and Social Simulation*, 13(1):3, 2010.

18. Jiang-Jiang Wang, You-Yin Jing, Chun-Fa Zhang, and Jun-Hong Zhao. Review on multi-criteria decision analysis aid in sustainable energy decision-making. *Renewable and Sustainable Energy Reviews*, 13(9):2263–2278, 2009.
19. Suzanne Varet. *Développement de méthodes statistiques pour la prédiction d'un gabarit de signature infrarouge*. PhD thesis, Université Paul Sabatier-Toulouse III, 2010.
20. H Niederreiter. Discrepancy and convex programming. *Annali di matematica pura ed applicata*, 93(1):89–97, 1972.
21. DT QGIS. Quantum gis geographic information system. *Open Source Geospatial Foundation Project*, 2011.
22. R Core Team. R language definition, 2000.
23. Jessica Franco, Delphine Dupuy, Olivier Roustant, and Astrid Jourdan. Dicedesign-package. *Designs of Computer Experiments*, page 2, 2009.
24. Karthik Ram. Git can facilitate greater reproducibility and increased transparency in science. *Source code for biology and medicine*, 8(1):7, 2013.
25. Jonathan Bennett. *OpenStreetMap*. Packt Publishing Ltd, 2010.
26. Rémi Louf and Marc Barthelemy. Patterns of residential segregation. *arXiv preprint arXiv:1511.04268*, 2015.
27. Frédéric Gilli and Jean-Marc Offner. *Paris, métropole hors les murs: aménager et gouverner un Grand Paris*. Sciences Po, les presses, 2009.
28. Marianne Guérois and Renaud Le Goix. La dynamique spatio-temporelle des prix immobiliers à différentes échelles: le cas des appartements anciens à paris (1990-2003). *Cybergeo: European Journal of Geography*, 2009.
29. Mihai Tivadar, Yves Schaeffer, André Torre, and Frédéric Bray. Oasis—un outil d'analyse de la ségrégation et des inégalités spatiales. *Cybergeo: European Journal of Geography*, 2014.
30. Mei-Po Kwan. Space-time and integral measures of individual accessibility: a comparative analysis using a point-based framework. *Geographical analysis*, 30(3):191–216, 1998.
31. Arnaud Banos. Pour des pratiques de modélisation et de simulation libérées en géographie et shs. *Thèse d'Habilitation à Diriger des Recherches, UMR CNRS 8504 Géographie-Cités, ISCPIF*, Décembre 2013.
32. SS Dragomir. The ostrowski's integral inequality for lipschitzian mappings and applications. *Computers & Mathematics with Applications*, 38(11):33–37, 1999.