A Discrepancy-based Framework to Compare Robustness between Multi-Attribute Evaluations

J. Raimbault[1,2]
juste.raimbault@parisgeo.cnrs.fr

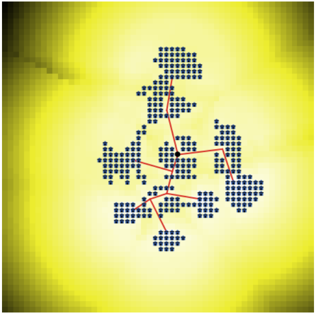
[1]UMR CNRS 8504 Géographie-cités
[2]UMR-T IFSTTAR 9403 LVMT

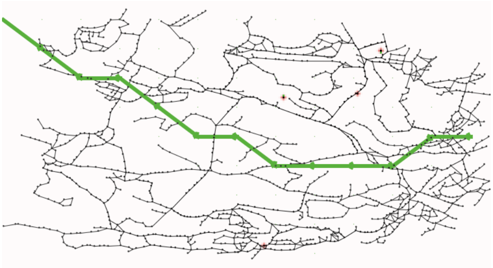
CSD&M 2016 - Paris

14th December 2016

Multi-objective Complex Systems

From morphogenetic factors to optimal design, fundamental multi-objective nature of optimization processes





Source : Left [Raimbault et al., 2014] ; Right [Raimbault and Gonzalez, 2015]

Multi-objective Evaluations of Socio-technical Systems

Systematic multi-objective nature of problems in design of Complex Industrial Systems [Marler and Arora, 2004] and in the study of Complex Natural Systems [Newman, 2011]

→ Design and understanding (i.e. evaluation) of Socio-technical systems at the intersection as hybrid systems [Picon, 2013][Haken and Portugali, 2003]

→ Territorial systems as typical examples : e.g. sustainable urban design [Souami, 2012], multi-criteria decision-making for transportation infrastructures [Bavoux et al., 2005]

Robustness of Evaluations

Reliability of an evaluation is crucial; various approaches to robustness depending on method

→ Naturally included in the construction and estimation of statistical models [Launer and Wilkinson, 2014] (e.g. p-value, beta power, AIC)

→ In multi-objective optimization, diverse methods: sensitivity of Pareto front to perturbations [Deb and Gupta, 2006]; continuity of solutions [Barrico and Antunes, 2006]

→ [Dobbie and Dail, 2013] studies robustness for multi-attribute evaluations as biases depending on weighting techniques

Towards a Generic Robustness Framework

Research Objective : *Investigate a generic data-driven approach to Robustness in Multi-attribute evaluations of Complex Socio-Technical Systems*

→ Model-independence and method-independence; framework based only on data structure (and thus quality) and indicator values

→ Particular case of multi-attribute evaluations, where dimensions are aggregated, in order to obtain a simple measure of robustness

Intuitive and Theoretical Framing

1. Systems are seen from the perspective of raw data available: data-driven approach
2. A choice of indicators captures the realization of an "urban fact" [Mangin and Panerai, 1999], in the sense of stylized process with different spatial realizations
3. Given many systems and indicators, a common space can be build to compare them
4. Discrepancy of data [Dick and Pillichshammer, 2010] in that space captures various aspects linked to robustness: system scale and range, precision of data, missing data
5. Robustness must also capture an error done on indicator computation

Assumptions

Objectives as Kernel Integrals

$\rightarrow$ Kernel functions exist for each objective, computed as their integrals

$\rightarrow$ Reasonable for most systems, as e.g. for territorial systems it is analog to smoothing of Geographically Weighted Regression [Brunsdon et al., 1998]

Linearly Aggregated Objectives

$\rightarrow$ Aggregated objective as linear combination of attributes $q(\vec{x}) = \sum_i w_i q_i(\vec{x})$

$\rightarrow$ Choice of weights at the core of decision-making process; not our scope here, we take simply relative indicator importance $w^L_{i,c} = \frac{\hat{q}_{i,c}}{\sum_c \hat{q}_{i,c}}$

Formal Description (I)

Territorial Systems : Data $S_i = \mathbf{X}_i \in \mathscr{X}_i$ with $\mathscr{X}_i = \prod_k \mathscr{X}_{i,k}$

System space

$$\mathscr{X} \underset{def}{=} \left(\prod \tilde{\mathscr{X}}_c\right) = \left(\prod_{\mathscr{X}_{i,k} \in \mathscr{D}_{\mathscr{X}}} \mathbb{R}^{p^X_{i,k}}\right)$$

Objectives : H_c space of real-valued functions on $\tilde{\mathscr{X}}_c$, such that :

1. $h_c \in H_c$ are "enough" regular (tempered distributions e.g.)
2. $q_c = \int_{\tilde{\mathscr{X}}_c} h$ is a function describing the "urban fact" (the indicator in itself)
3. Normalized kernels $h_c(\vec{x}) \in [0,1]$

Formal Description (II)

Integral approximation theorem gives upper bound on error, linked to data discrepancy [Niederreiter, 1972][Varet, 2010]

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

which propagates to the linear aggregation

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

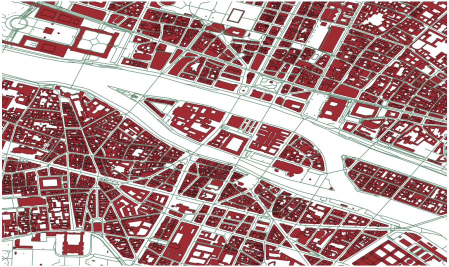
$\rightarrow$ A relative *Robustness Ratio* can thus be defined between two evaluations :

$$R_{i,i'} = \frac{\sum_c w_{i,c} \cdot D_{i,c}}{\sum_c w_{i',c} \cdot D_{i',c}} \tag{1}$$

Implementation on Synthetic Data

Using OpenStreetMap roads and buildings data, construction of synthetic indicators on Paris districts (car daily use, car flows in streets, relative length of pedestrian streets), computed in Monte-carlo simulations

$\rightarrow$ Minimal ratio (relative to 1st Arr.) in 15th Arr. (0.92 ± 0.03), intuitively expected



Example of raw data

Metropolitan Segregation

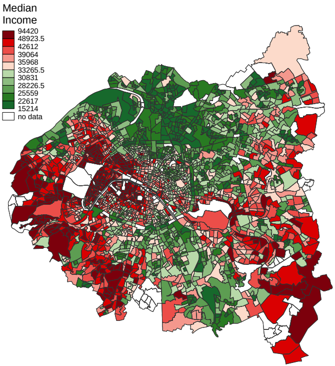
Application to metropolitan Segregation on Ile-de-France, Insee income data (2011)

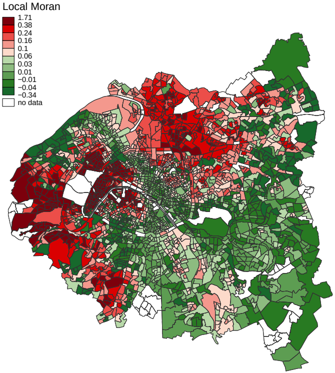
Indicators :

- Spatial autocorrelation Moran index
- Dissimilarity index $d = \frac{1}{\sum_{ij} w_{ij}} \sum_{ij} w_{ij} \left| \tilde{X}_i - \tilde{X}_j \right|$
- Complementary of distribution entropy $\varepsilon = 1 + \frac{1}{\log(N)} \sum_i \frac{X_i}{\sum_k X_k} \cdot \log\left(\frac{X_i}{\sum_k X_k}\right)$

Metropolitan Segregation: Indicators

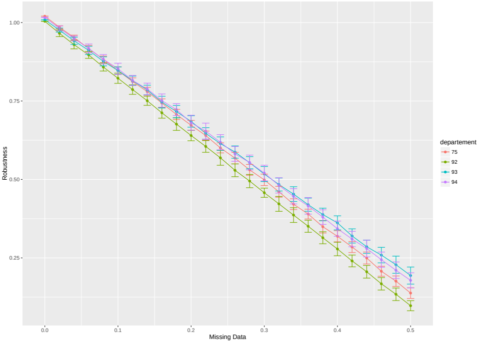
Example of Segregation maps (linking median income with local Moran)

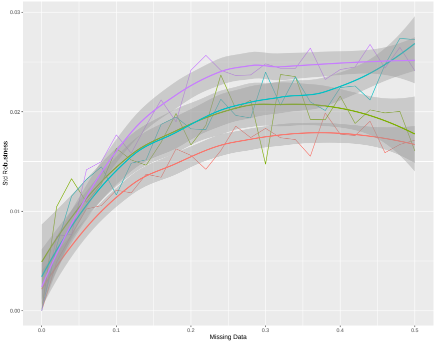




Metropolitan Segregation: Results

Framework Application : robustness comparison between administrative areas and sensitivity to missing data





(Left) Robustness ratio compared to full region for each area, as a function of missing data proportion; (Right) Standard deviations

Discussion

Applicability

→ Application to decision-making procedures : adding robustness as a dimension

→ Assumptions validity ranges : some indicators may difficultly be viewed as spatial integrals (as some accessibility measures [Kwan, 1998])

→ Availability of raw data

Further Developments

→ Application to existing open frameworks (e.g. [Tivadar et al., 2014])

→ More general formulation, first to non-linear aggregation (e.g. for Lipschitzian functions [Dragomir, 1999])

→ Multi-dimensional measure of robustness using discrepancy along different dimensions

Conclusion

→ An original data-driven approach to robustness in multi-attribute evaluations of complex socio-technical systems; independent of model and method

→ Future insights from application to diverse system types and fields should refine the framework

→ Importance of interdisciplinarity (linking here statistics with computational modeling), crucial to study Complex Systems

- Special thanks to J. Keutchayan (Ecole Polytechnique de Montréal) for suggesting the original idea of using discrepancy

- All code and data available at `https://github.com/JusteRaimbault/RobustnessDiscr`

- Paper preprint available [Raimbault, 2016] at `http://arxiv.org/abs/1608.00840`

Reserve Slides

Reserve Slides

Kernel Examples

Typical concrete example of kernels can be :

- A mean of rows of $\mathbf{X}_{i,c}$ is computed with $h(x) = x \cdot f_{i,c}(x)$ where $f_{i,c}$ is the density of the distribution of the assumed underlying variable.
- A rate of elements respecting a given condition C, $h(x) = f_{i,c}(x)\chi_{C(x)}$
- For already aggregated variables $\mathbf{Y}$, a Dirac distribution allows to express them also as a kernel integral.

Implementation

- Preprocessing of geographical data is made through QGIS [QGis, 2011]
- Core implementation of the framework is done in R [Team, 2000] for the flexibility of data management and statistical computations
- package `DiceDesign`[Franco et al., 2009] for computation of discrepancies

Synthetic Indicators (I)

- Complementary of the average daily distance to work with car per individual, approximated by, with $n_{cars}(b)$ number of cars in the building (randomly generated by associated of cars to a number of building proportional to motorization rate α_m 0.4 in Paris), d_w distance to work of individuals (generated from the building to a uniformly generated random point in spatial extent of the dataset), and d_{max} the diameter of Paris area, $\bar{d}_w = 1 - \frac{1}{|b \in A(a)|} \cdot \sum_{b \in A(a)} n_{cars}(b) \cdot \frac{d_w}{d_{max}}$

Synthetic Indicators (II)

- Complementary of average car flows within the streets in the district, approximated by, with $\varphi(s)$ relative flow in street segment s, generated through the minimum of 1 and a log-normal distribution adjusted to have 95% of mass smaller than 1 what mimics the hierarchical distribution of street use (corresponding to betweenness centrality), and $l(s)$ segment length, $\bar{\varphi} = 1 - \frac{1}{|s \in A(a)|} \cdot \sum_{s \in A(a)} \varphi(s) \cdot \frac{l(s)}{\max(l(s))}$
- Relative length of pedestrian streets $\bar{p}$, computed through a randomly uniformly generated dummy variable adjusted to have a fixed global proportion of segments that are pedestrian.

Synthetic Numerical Results

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

References I

Barrico, C. and Antunes, C. (2006).
Robustness analysis in multi-objective optimization using a degree of robustness concept.
In *Evolutionary Computation, 2006. CEC 2006. IEEE Congress on*, pages 1887–1892.

Bavoux, J.-J., Beaucire, F., Chapelon, L., and Zembri, P. (2005).
Géographie des transports.
Paris.

Brunsdon, C., Fotheringham, S., and Charlton, M. (1998).
Geographically weighted regression.
Journal of the Royal Statistical Society: Series D (The Statistician), 47(3):431–443.

References II

Deb, K. and Gupta, H. (2006).
Introducing robustness in multi-objective optimization.
Evolutionary Computation, 14(4):463–494.

Dick, J. and Pillichshammer, F. (2010).
Digital nets and sequences: Discrepancy Theory and Quasi–Monte Carlo Integration.
Cambridge University Press.

Dobbie, M. J. and Dail, D. (2013).
Robustness and sensitivity of weighting and aggregation in constructing composite indices.
Ecological Indicators, 29:270–277.

References III

Dragomir, S. (1999).
The ostrowski's integral inequality for lipschitzian mappings and applications.
Computers & Mathematics with Applications, 38(11):33–37.

Franco, J., Dupuy, D., Roustant, O., and Jourdan, A. (2009).
Dicedesign-package.
Designs of Computer Experiments, page 2.

Haken, H. and Portugali, J. (2003).
The face of the city is its information.
Journal of Environmental Psychology, 23(4):385–408.

References IV

Kwan, M.-P. (1998).
Space-time and integral measures of individual accessibility: a comparative analysis using a point-based framework.
Geographical analysis, 30(3):191–216.

Launer, R. L. and Wilkinson, G. N. (2014).
Robustness in statistics.
Academic Press.

Mangin, D. and Panerai, P. (1999).
Projet urbain.
Parenthèses.

Marler, R. T. and Arora, J. S. (2004).
Survey of multi-objective optimization methods for engineering.
Structural and multidisciplinary optimization, 26(6):369–395.

References V

Newman, M. (2011).
Complex systems: A survey.
arXiv preprint arXiv:1112.1440.

Niederreiter, H. (1972).
Discrepancy and convex programming.
Annali di matematica pura ed applicata, 93(1):89–97.

Picon, A. (2013).
Smart cities: théorie et critique d'un idéal auto-réalisateur.
B2.

QGis, D. (2011).
Quantum gis geographic information system.
Open Source Geospatial Foundation Project.

References VI

Raimbault, J. (2016).
A discrepancy-based framework to compare robustness between multi-attribute evaluations.
Forthcoming in Proceedings of CSDM 2016. arXiv preprint arXiv:1608.00840.

Raimbault, J., Banos, A., and Doursat, R. (2014).
A hybrid network/grid model of urban morphogenesis and optimization.
In *Proceedings of the 4th International Conference on Complex Systems and Applications (ICCSA 2014), June 23-26, 2014, Université de Normandie, Le Havre, France; M. A. Aziz-Alaoui, C. Bertelle, X. Z. Liu, D. Olivier, eds.: pp. 51-60.*

References VII

Raimbault, J. and Gonzalez, J. (May 2015).
Application de la morphogénèse de réseaux biologiques à la conception optimale d'infrastructures de transport.
In *Rencontres du Labex Dynamites*.

Souami, T. (2012).
Ecoquartiers: secrets de fabrication.
Scrineo.

Team, R. C. (2000).
R language definition.

Tivadar, M., Schaeffer, Y., Torre, A., and Bray, F. (2014).
Oasis–un outil d'analyse de la ségrégation et des inégalités spatiales.
Cybergeo: European Journal of Geography.

References VIII

Varet, S. (2010).
Développement de méthodes statistiques pour la prédiction d'un gabarit de signature infrarouge.
PhD thesis, Université Paul Sabatier-Toulouse III.