

Validation des modèles de simulation en géographie: nouvelles pratiques et perspectives

J. Raimbault^{1,2,*}

juste.raimbault@polytechnique.edu

¹UPS CNRS 3611 ISC-PIF

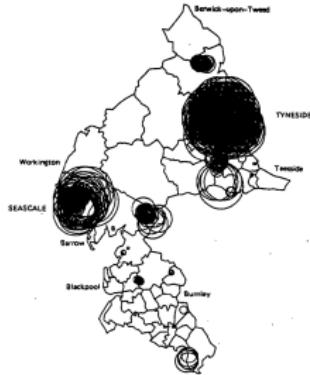
²UMR CNRS 8504 Géographie-cités

Séminaire EpisteMod

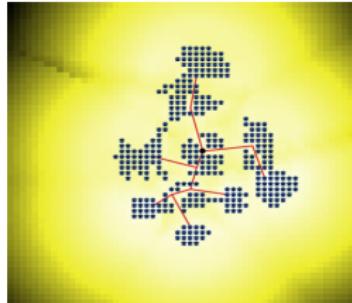
IHPHST

October 9th 2018

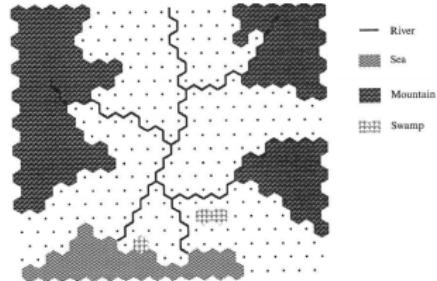
Une longue histoire de modélisation et simulation en TQG



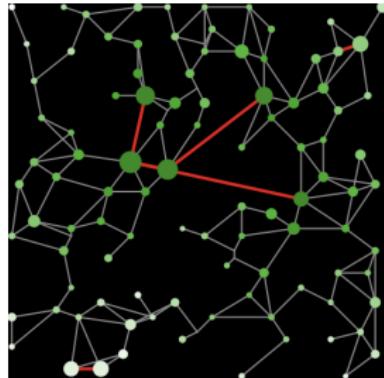
Machine d'analyse géographique
[Openshaw et al., 1987]



Morphogenèse urbaine hybride
[Raimbault et al., 2014]



Modèle Simpop 1 [Sanders et al., 1997]



Modèle SimpopNet [Schmitt, 2014]

Nécessité des modèles de simulation en géographie induite par les multiples complexités de ces systèmes ?

- Complexité ontologique [Pumain, 2003]
- Complexité dynamique: non-ergodicité et dépendance au chemin [Pumain, 2012]
- Complexité et co-évolution
- Complexité et emergence

Succession historique d'épistémologies dans le cas des systèmes de villes
[Varenne, 2017]:

- ① Déduction depuis la théorie (top-down): Christaller
- ② Induction depuis l'empirique (bottom-up): Berry
- ③ Vers une épistémologie abductive (interaction iterative théorique-empirique): Pumain

→ la simulation permet la synthèse

Validation des modèles de simulation

Multiples approches de la “validation” et “verification” des modèles [Rey-Coyrehourcq, 2015]: origine dans l’ingénierie système; statut épistémologique lié aux fonctions du modèle; approches empiriques ad-hoc.

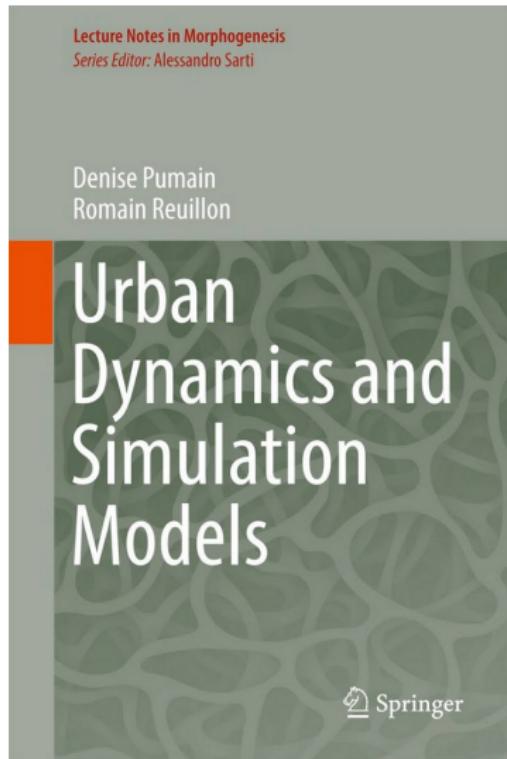
Constat d'un manque d'outils

En pratique: tests unitaires du programme, validation interne, sorties visuelles, reproduction de motifs (quantitatif ou qualitatif), pouvoir prédictif, analyses de sensibilité.

→ en pratique, peu de latitude sur les hypothèses théoriques et de modélisation; peu d’interaction entre les domaines de connaissance.

[Rey-Coyrehourcq, 2015]: vers une validation par l’évaluation (faits stylisés, pattern-oriented modeling, multi-modélisation) en interaction avec la théorie et l’empirique.

Vers de nouvelles pratiques: ERC Geodiversity



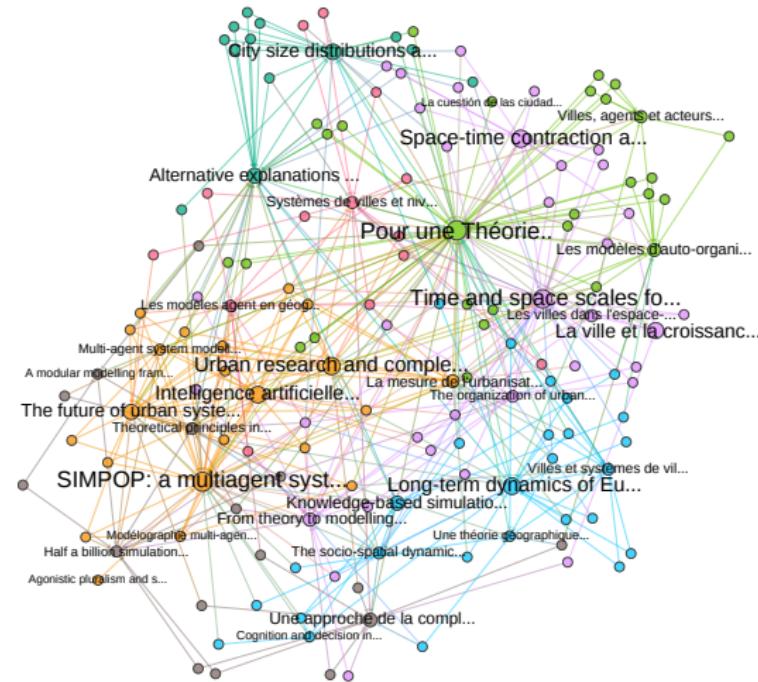
Développement de la théorie évolutive des villes [Pumain, 2018a]

- Faits stylisés saillants pour l'ensemble des systèmes de villes principaux
- Construction de modèles de simulation (modèles à visée explicative)
- Outils et méthodes d'exploration des modèles de simulation



Théorie évolutive des villes

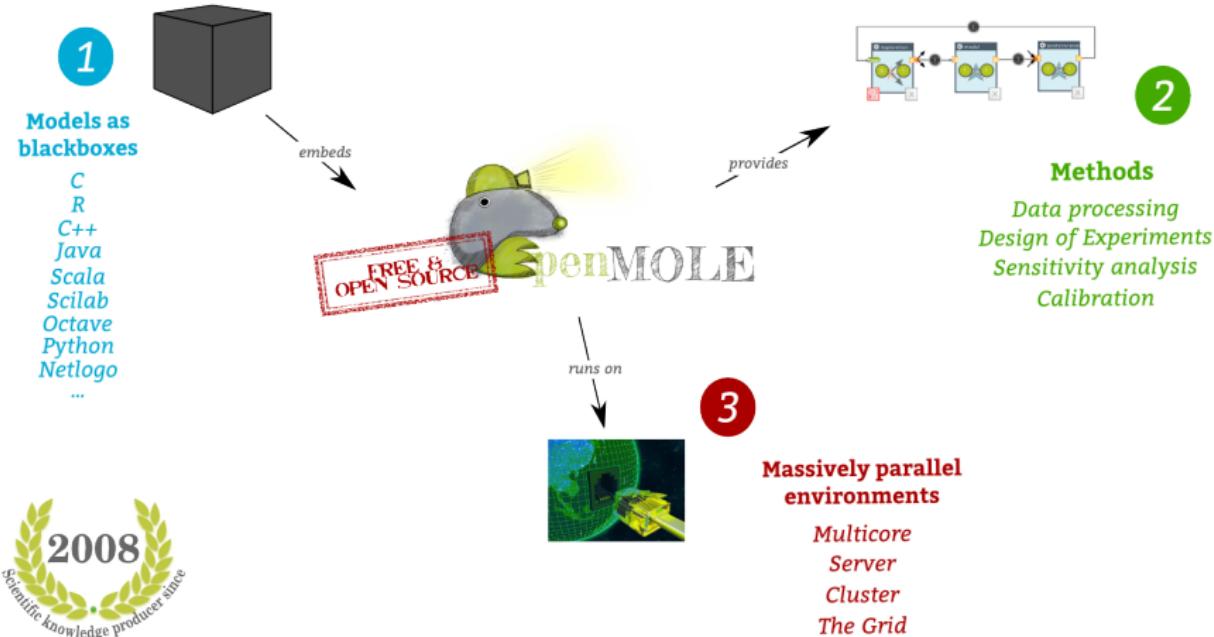
De [Pumain, 1997] à [Pumain, 2018a]: les systèmes de villes comme des systèmes co-évolutifs au sein desquels les interactions sont cruciales



[Raimbault, 2017] Analyse par réseau de citation des publications clés autour de la théorie évolutive des villes

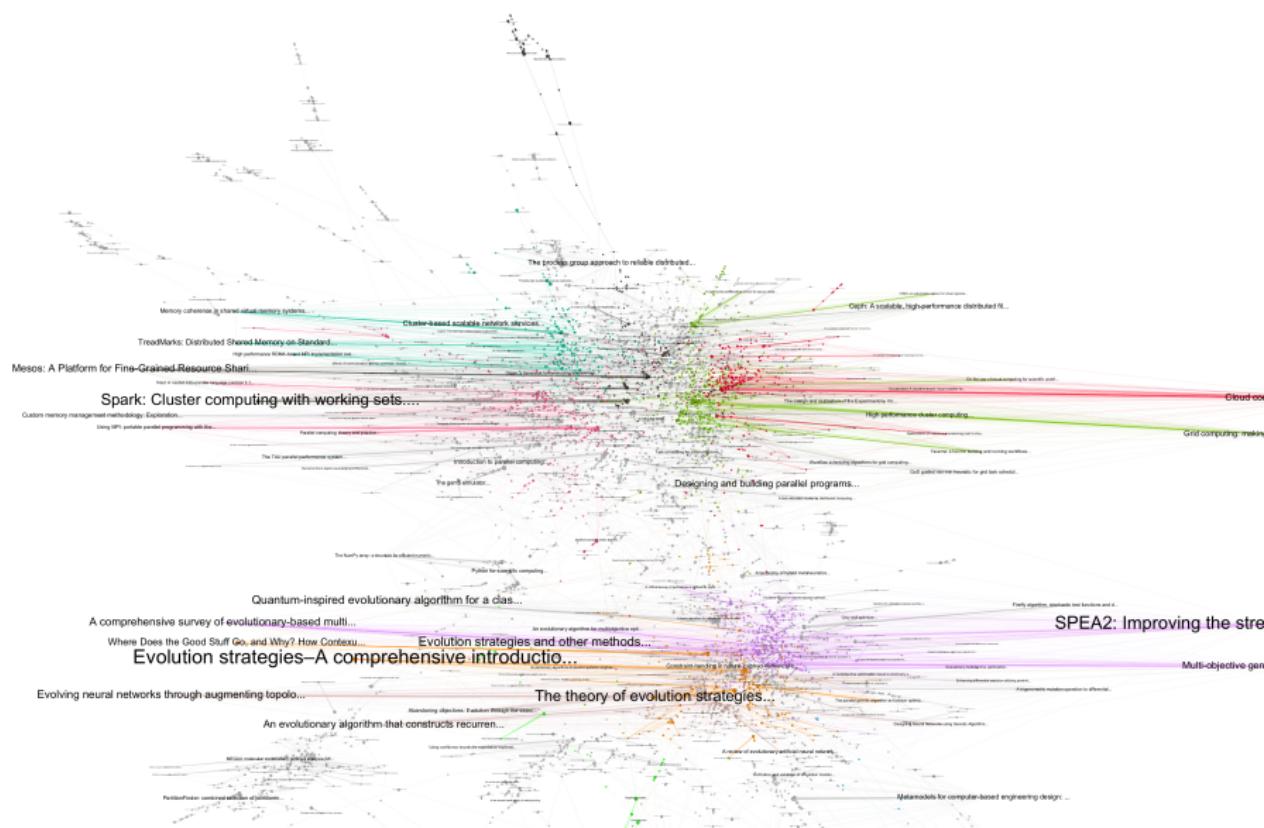
Manifeste d'OpenMOLE

(i) Méthodes d'exploration innovantes; (ii) Passage à l'échelle par calcul distribué; (iii) Pas d'interférence avec le modèle.

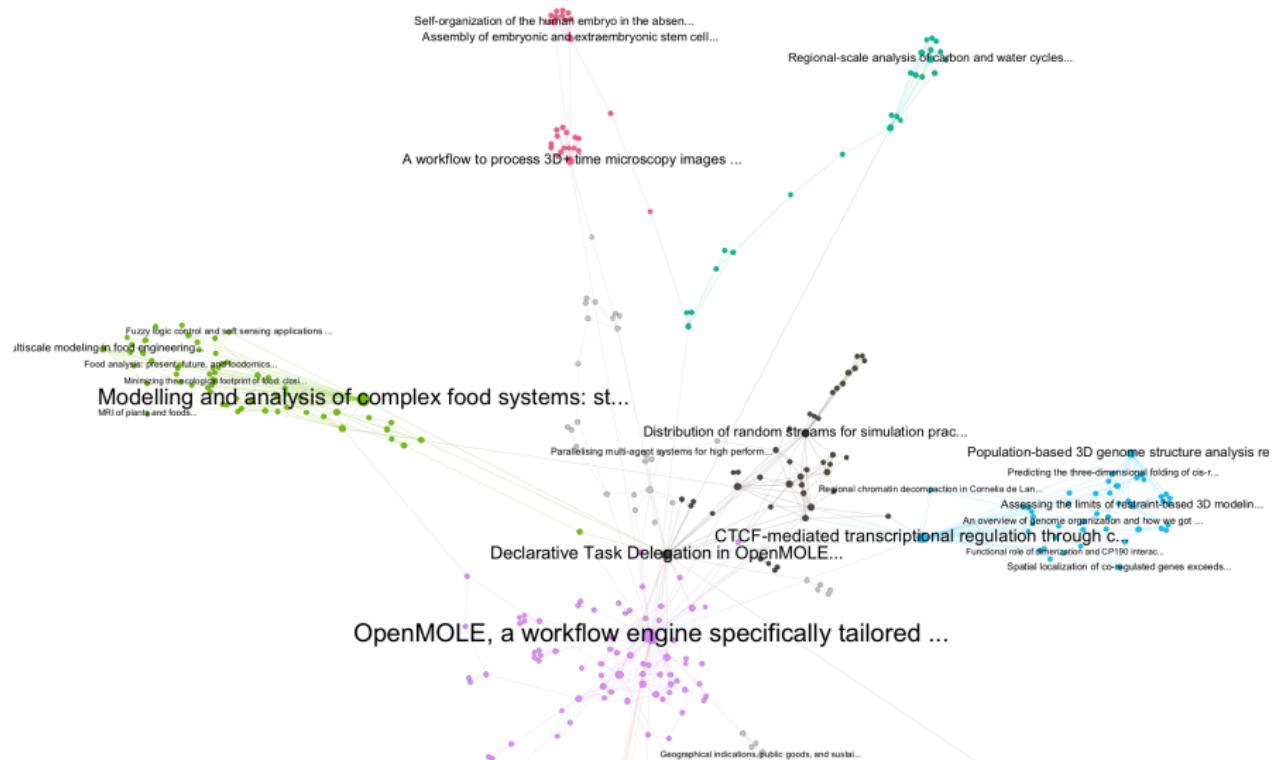


Environnement scientifique d'OpenMOLE

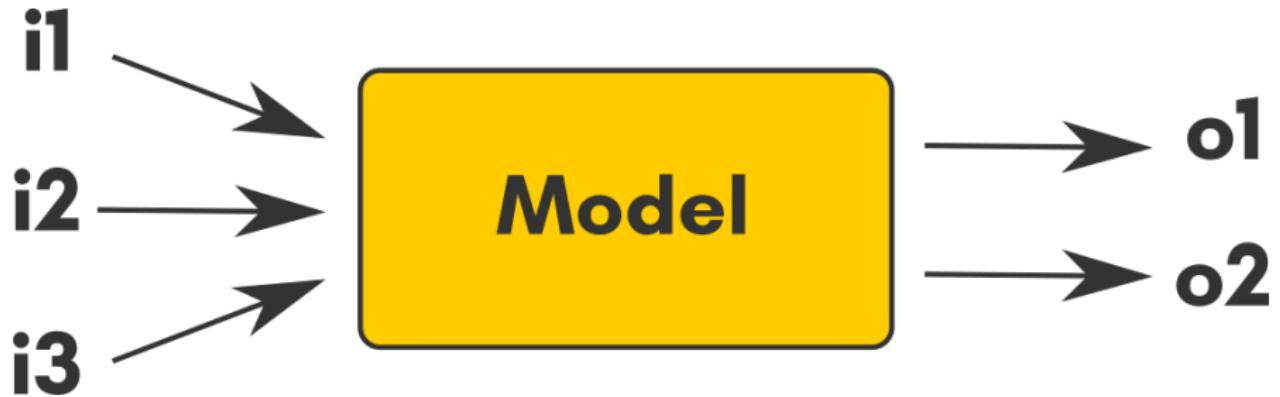
[Raimbault, 2018a]



Environnement scientifique immédiat



Modèles de simulation



Une interface web ergonomique

Mozilla Firefox

localhost:46857/app +

localhost:46857/app Rechercher

New project Run

Model Exploration Tutorial | Calibrate

results 4.00E+0

Calibrate.oms 0.94E+0

Objective.oms 0.29E+0

fitnessVSEval.png 15.17E+0

Calibrate.oms

```
4
5 // Execute the workflow
6 // Define the population (10) and the number of generations (100).
7 // Define the inputs and their respective variation bounds.
8 // Define the objectives to minimize.
9 // Assign 1 percent of the computing time to reevaluating
10 // parameter settings to eliminate over-evaluated individuals.
11 val nsga2 =
12   NSGA2(
13     mu = 50,
14     genome = Seq(
15       diffusion in (0.0, 99.0),
16       evaporation in (0.0, 99.0)),
17     objectives = Seq(deltaFood),
18     stochastic = Stochastic(seed = seed, aggregation = Seq(median))
19   )
20
21 val evolution =
22   SteadyStateEvolution(
23     algorithm = nsga2,
24     evaluation = ants -- objective,
25     parallelism = 10,
26     termination = 100
27   )
28
29 // Define a hook to save the Pareto frontier
30 val savePopulationHook = SavePopulationHook(evolution, workDirectory / "results")
31
32 // Plug everything together to create the workflow
33 (evolution hook savePopulationHook)
34
35
36
37
```

built the 01/05/2018 18:51:28

C
R
C++
Java
Scala
Scilab
Octave
Python
Netlogo
...



Modèle NetLogo

```
val model =  
  NetLogo6Task(  
    workDirectory / "Fire.nlogo",  
    List("setup", "while [any? turtles] [go]")) set (  
      inputs += seed,  
      outputs += (seed, density),  
      inputs += density mapped "density",  
      outputs += burned mapped "burned-trees"  
    )
```

Code R

```
val i = Val[Int]

val rTask =
  RTask("""
    source("function.R")
    function(i)""") set (
      resources += workDirectory / "function.R",
      inputs += i
    )
```

Code Python en package

Package

```
care -o python.tgz.bin python matrix.py data.csv 1 out.csv
```

Run

```
val pyTask =  
    CARETask(  
        workDirectory / "../python.tgz.bin",  
        "python matrix.py data.csv 2 out.csv"  
    )
```

Méthodes

- Reconstruction de données
- Estimation de paramètres
- Analyse de sensibilité
- Étude de robustesse
- Optimisation

Conçues pour passer à l'échelle, prendre en compte la stochasticité, être utilisable sur n'importe quel modèle et environnement de calcul.

Méthodes: scripting

```
DirectSampling(  
    evaluation = myModel,  
    sampling =  
        LHS(  
            500,  
            diffusion in (10.0, 100.0),  
            evaporation in (10.0, 100.0)  
        )  
)
```

Environnements de calcul: passage à l'échelle

Prototypes locaux, passage à l'échelle transparent: zero déploiement, pas de connaissance technique requise, pas d'installation préalable.

```
val cluster = SLURMEnvironment("login", "cluster.domain.org")  
  
DirectSampling(  
    evaluation = myModel on cluster,  
    sampling =  
        LHS(  
            500,  
            diffusion in (10.0, 100.0),  
            evaporation in (10.0, 100.0)  
        )  
)
```

Environnements supportés

Environnements pris en charge à l'heure actuelle:

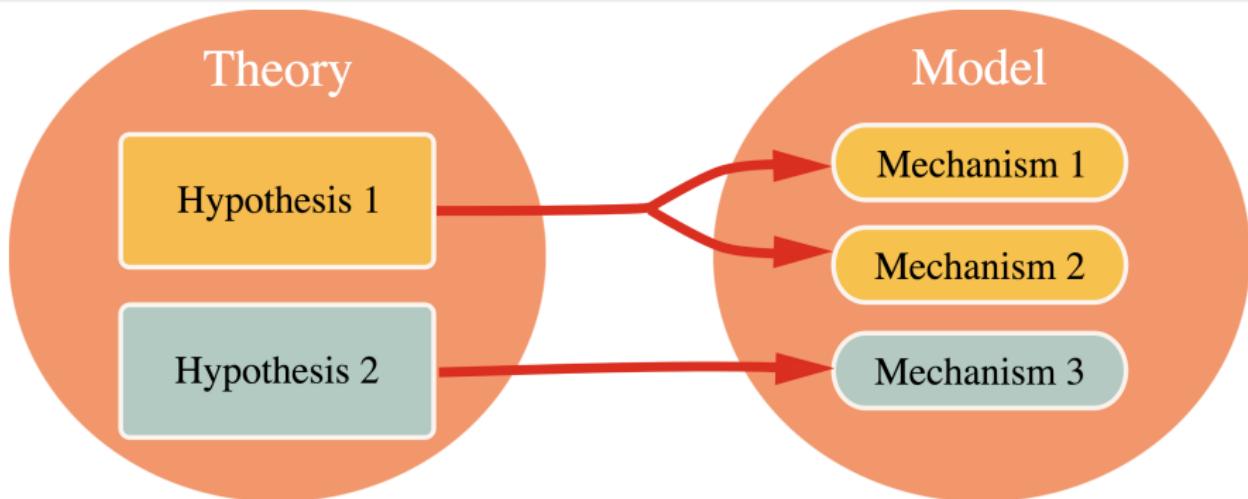
- Multi-thread
- Delegation through SSH
- PBS
- SLURM
- Condor
- SGE
- OAR
- EGI Grid

Cadre théorique et méthodes (algorithmes) pour accompagner le processus de modélisation

Un processus de modélisation:

- ① Traçable: comprendre les choix effectués
- ② Reproductible: vérification des raisons de ceux-ci
- ③ Réutilisable: étudier des choix alternatifs

Modèles descriptifs

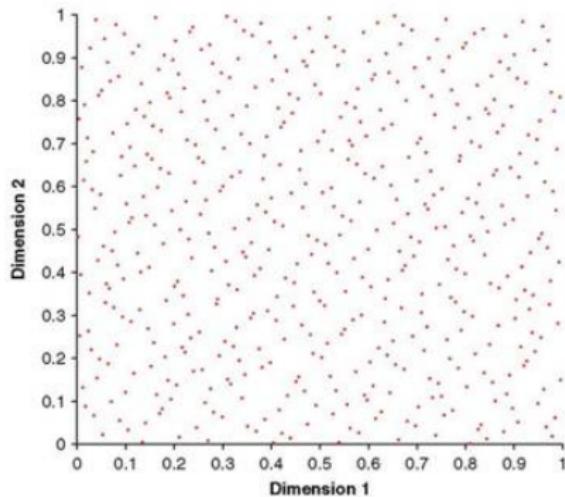


Construire et évaluer une théorie impliquant des effets causaux par sa capacité à (re-)produire des motifs ou données.

Evaluation: Comment s'assurer de

- ① La suffisance des mécanismes ?
- ② La nécessité des mécanismes ?
- ③ L'unicité des mécanismes ?

Approche classique: Design of Experiments

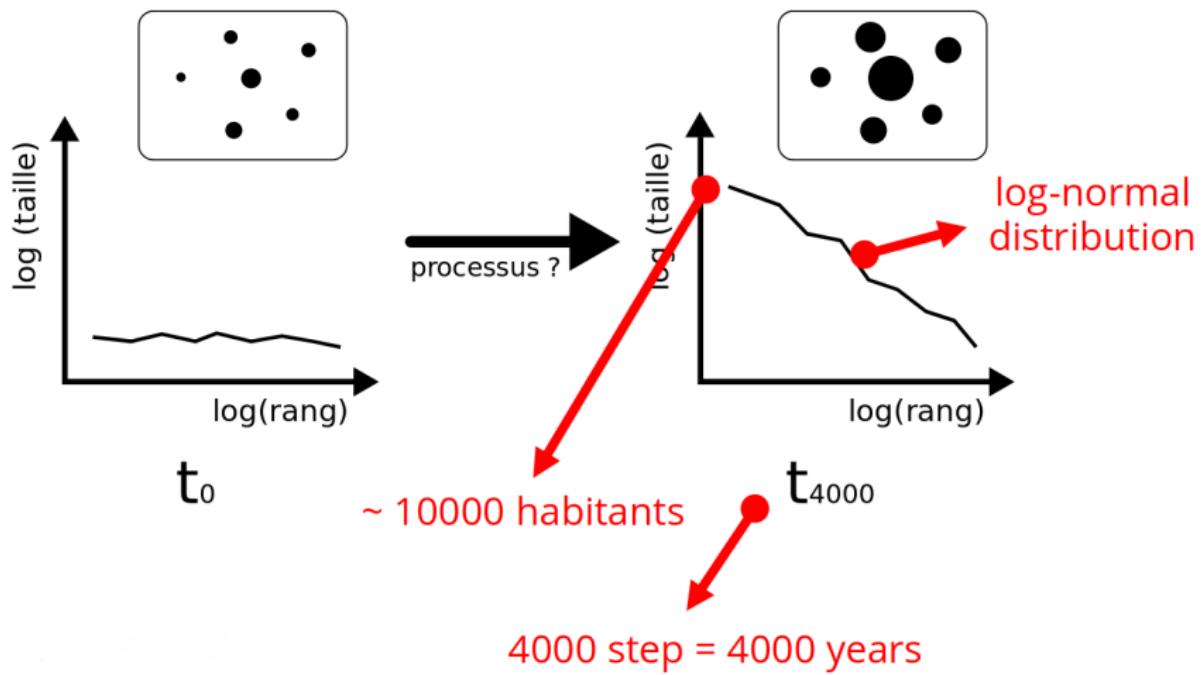


→ énorme quantité de données produites; le problème devient une question de fouille de données; espace des paramètres laissé majoritairement inexploré.

Méthode proposée [Schmitt et al., 2015]

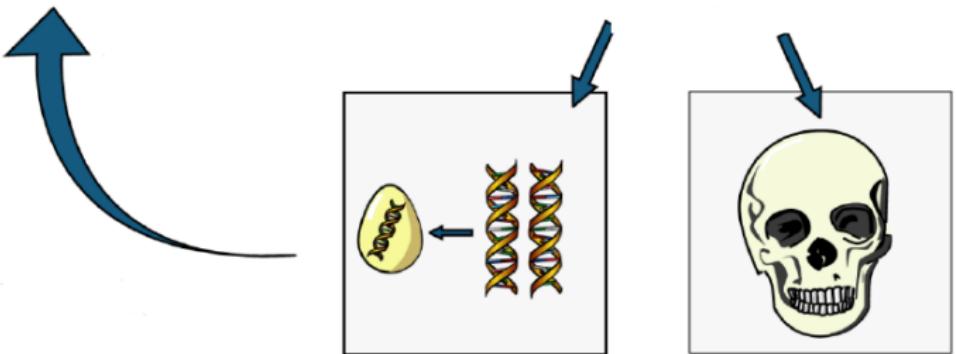
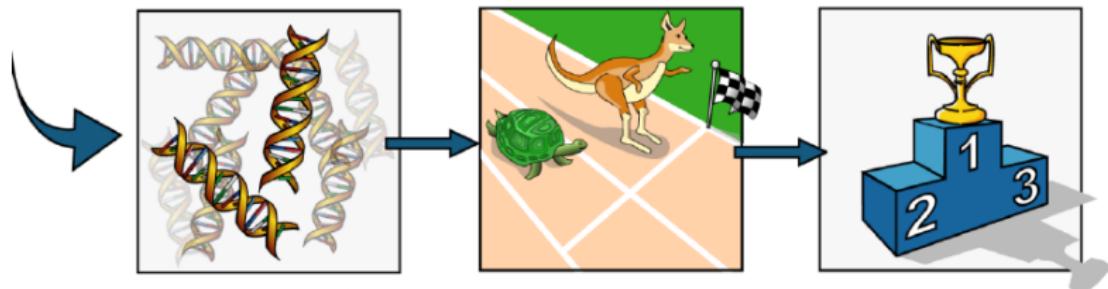
Approche inverse: des sorties aux paramètres

Formalisation thématique d'indicateurs:



Calibration

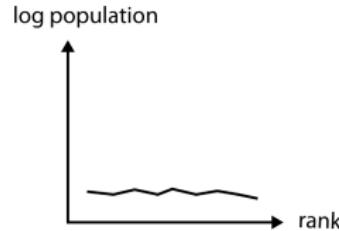
Algorithme génétique pour la calibration



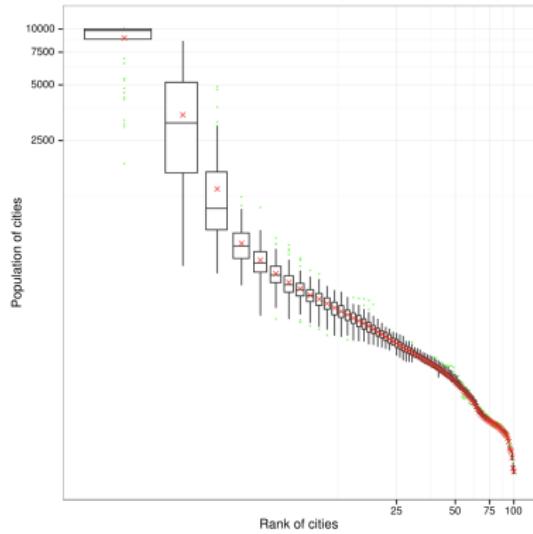
Résultats de calibration

Pas de compromis entre les trois objectifs.

Searched pattern



Produced pattern

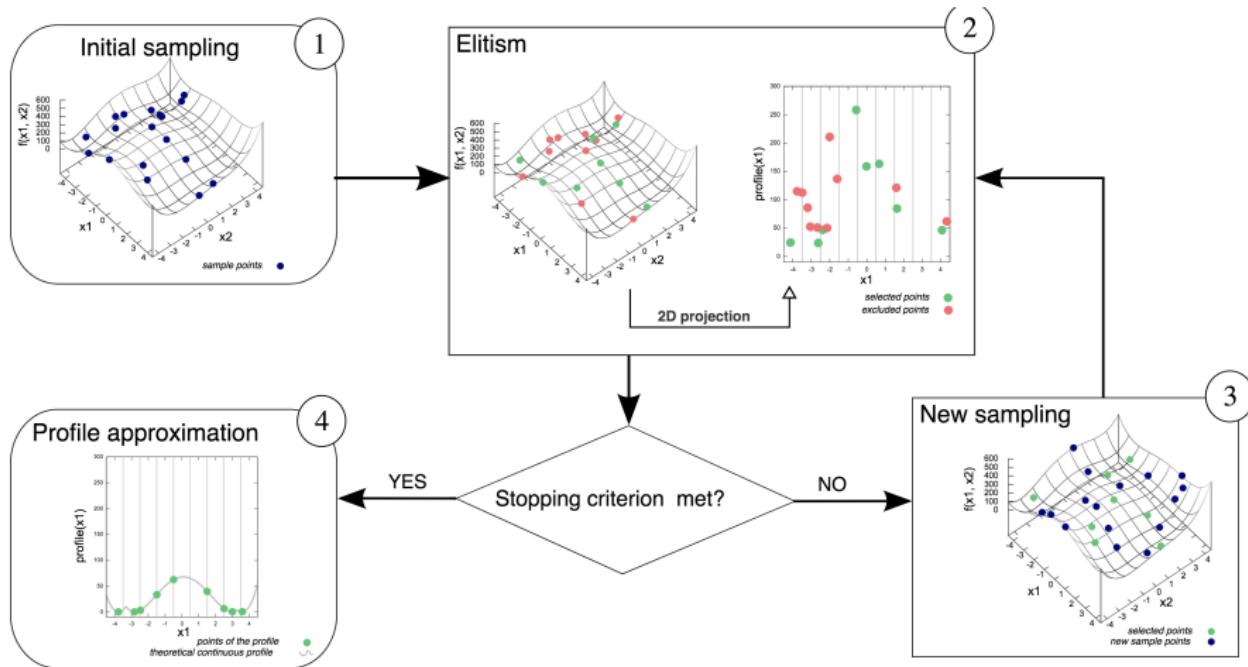


Performances: gestion de la stochasticité (gain x100), calcul distribué (gain x1000)

Nouvel algorithme

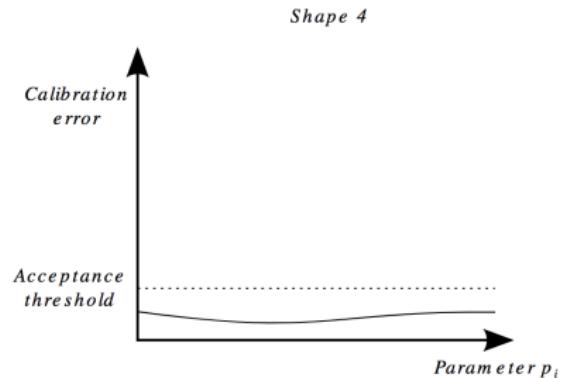
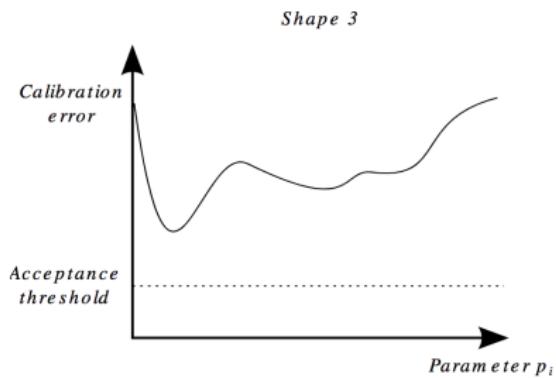
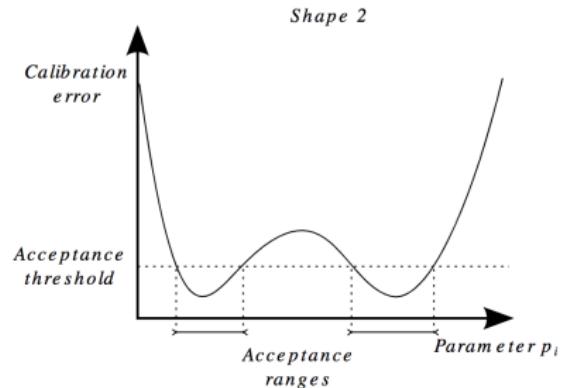
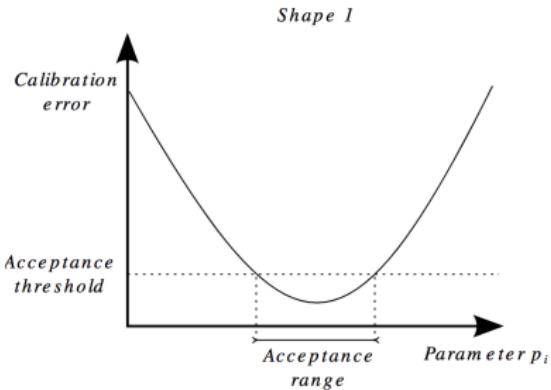
- ① Déetecte si un paramètre est nécessaire
- ② Contraint mieux les bornes des paramètres
- ③ Façon indirecte d'identifier des extensions possibles

Algorithme des profils

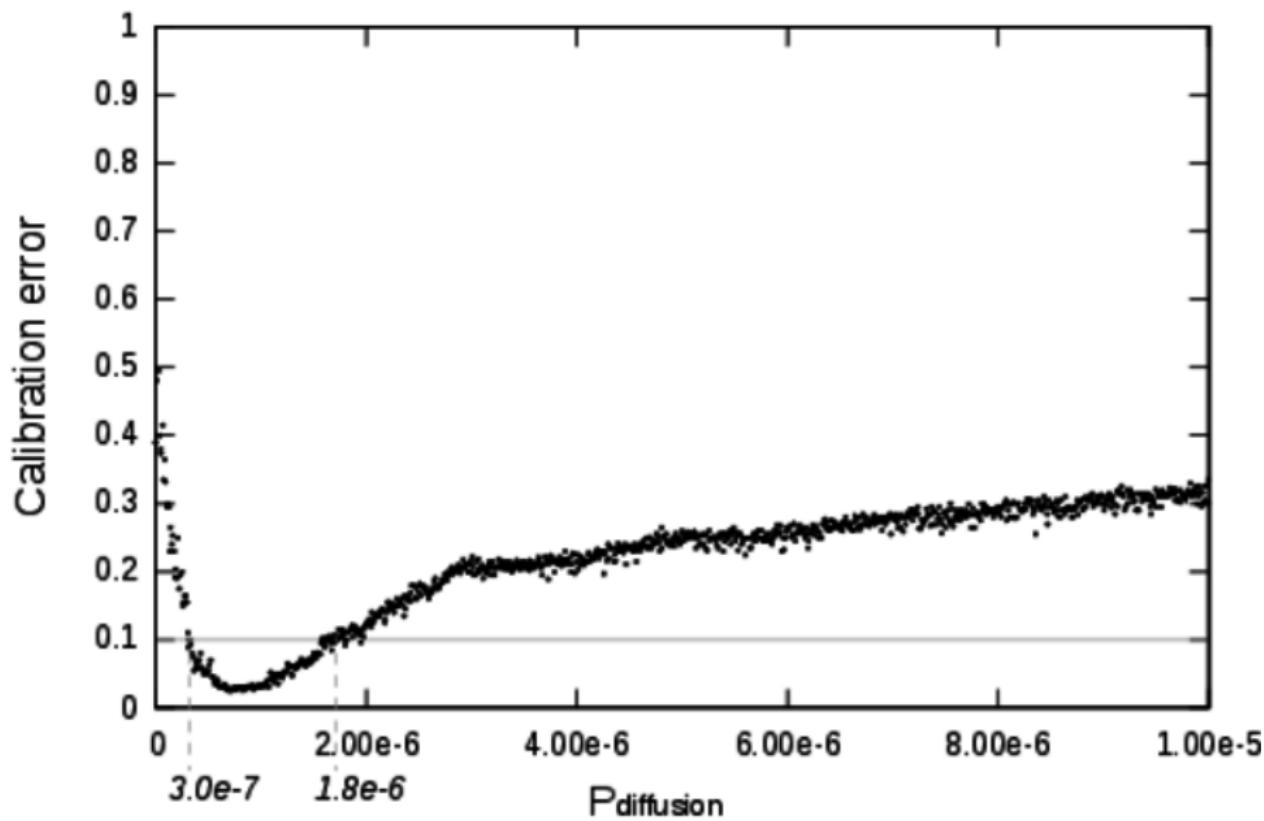


Meilleure calibration pour des pas fixés le long d'une dimension

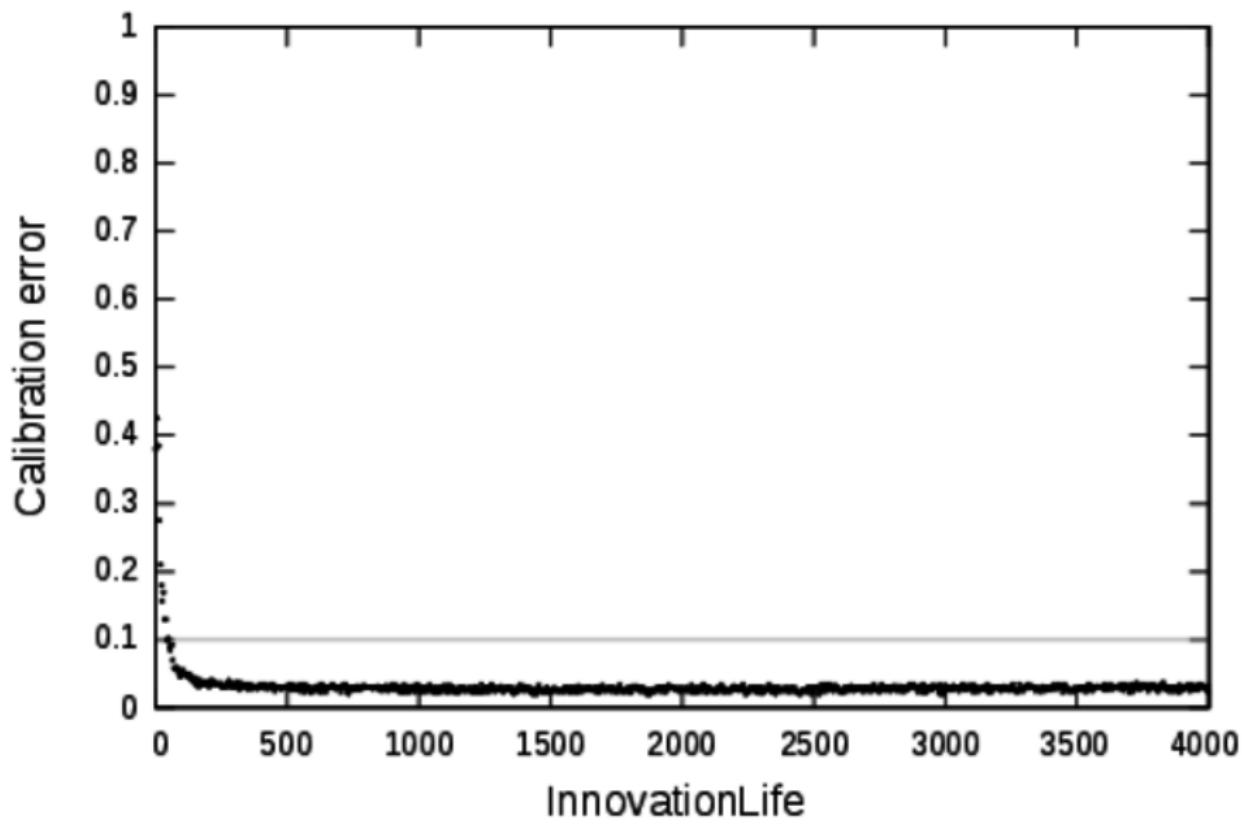
Algorithme des profils



Résultats



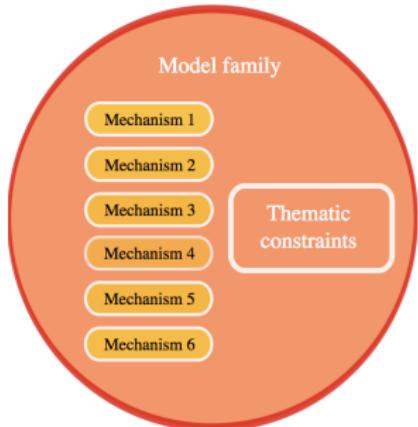
Résultats



Unicité [Cottineau, 2014]

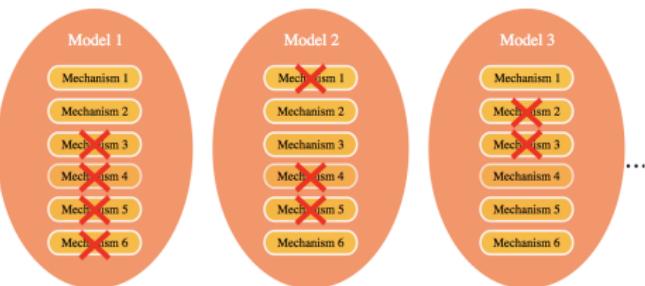
Automatisation de la confrontation entre hypothèses/mécanismes alternatifs

Thematic hypothesis



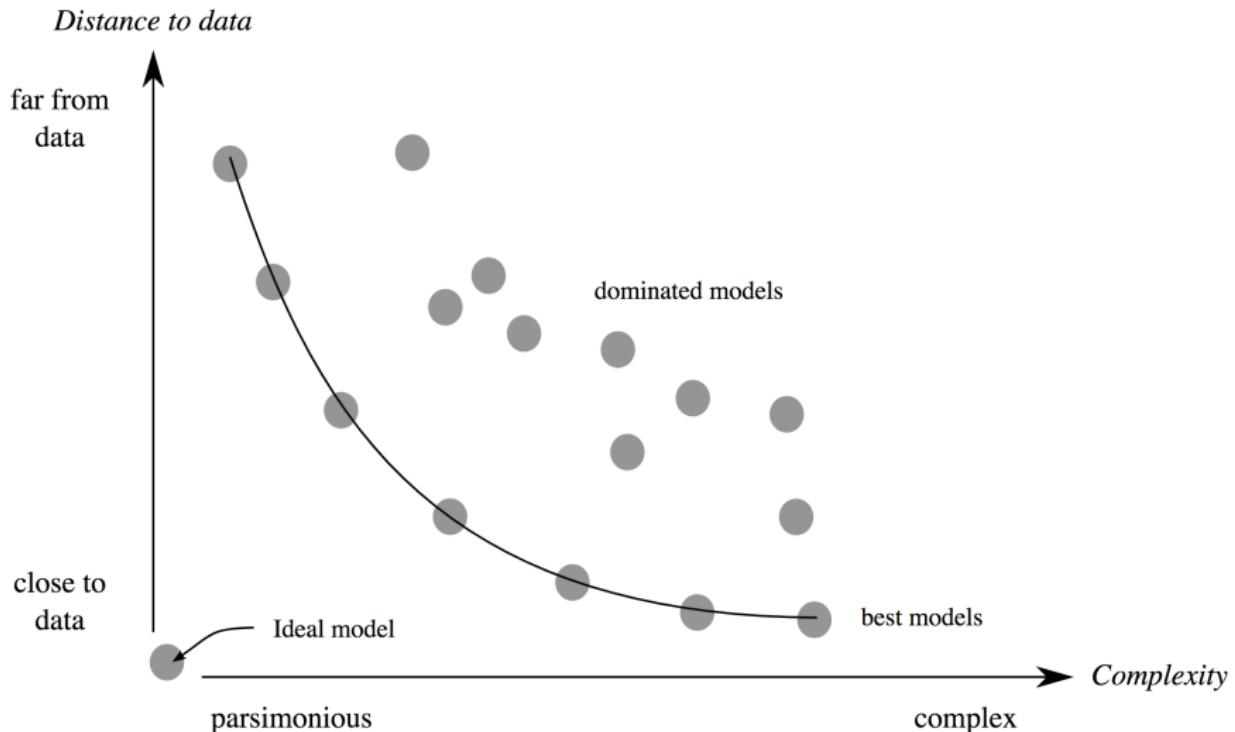
Generates

Candidate models



New calibration algorithm
designed to calibrate
a model family

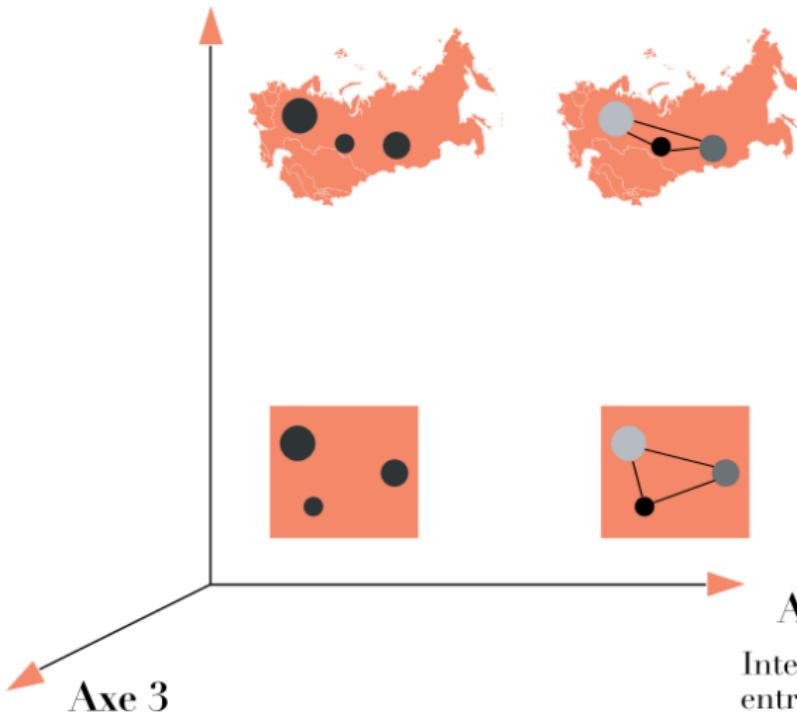
Objectif



Multi-modélisation (64 modèles)

Axe 2

Interactions entre
les agents et
l'environnement

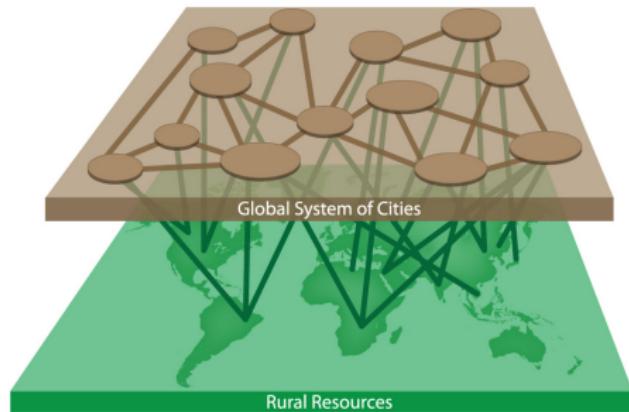


(c) Clémentine Cottineau, UMR Géographie-Cités, P.A.R.I.S., 2014

Exemple d'hypothèses concurrentes

Mécanisme d'échange: marché vs centralisé

Croissance des villes: interaction interurbaines vs situation environnementale



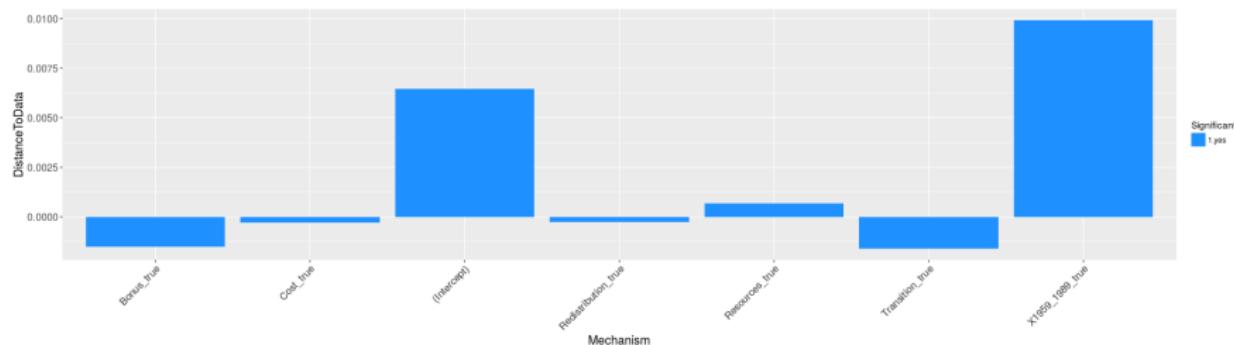
Calibration de la famille de modèles

Meilleur jeu de paramètres pour l'ensemble des 64 modèles

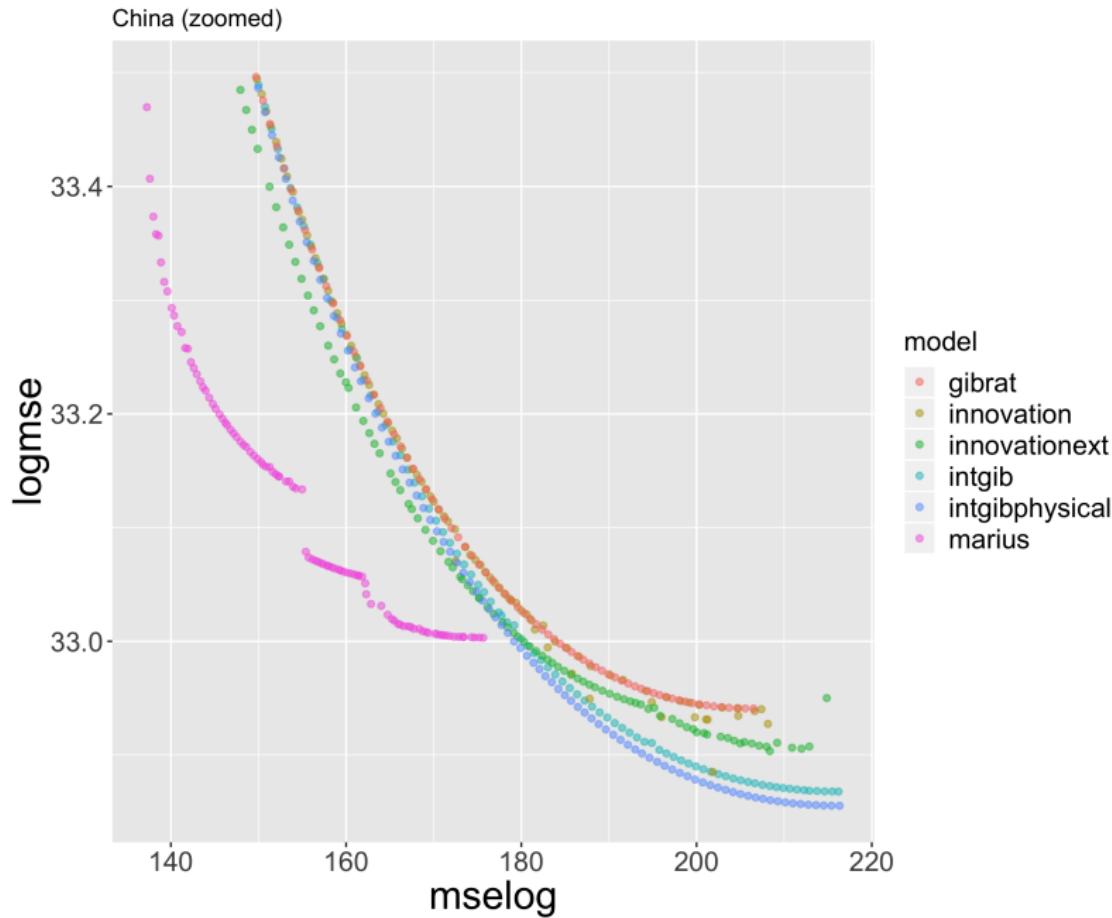
Contribution of mechanisms to the quality of simulation (closeness to data)

Models with different combination of mechanisms have been calibrated intensively against empirical data, using generic algorithms for more than 100000 generations. This plot shows the results of a regression explaining one measure of the quality of models (a small difference between simulated and empirical urban trajectories) by their mechanisms composition (the fact that any of the supplementary mechanisms is activated or not). Each bar represents the value of the estimated coefficient for each activated mechanism, in comparison with the same model structure without this mechanism, everything else being equal.

Statistical Significance (% of error)

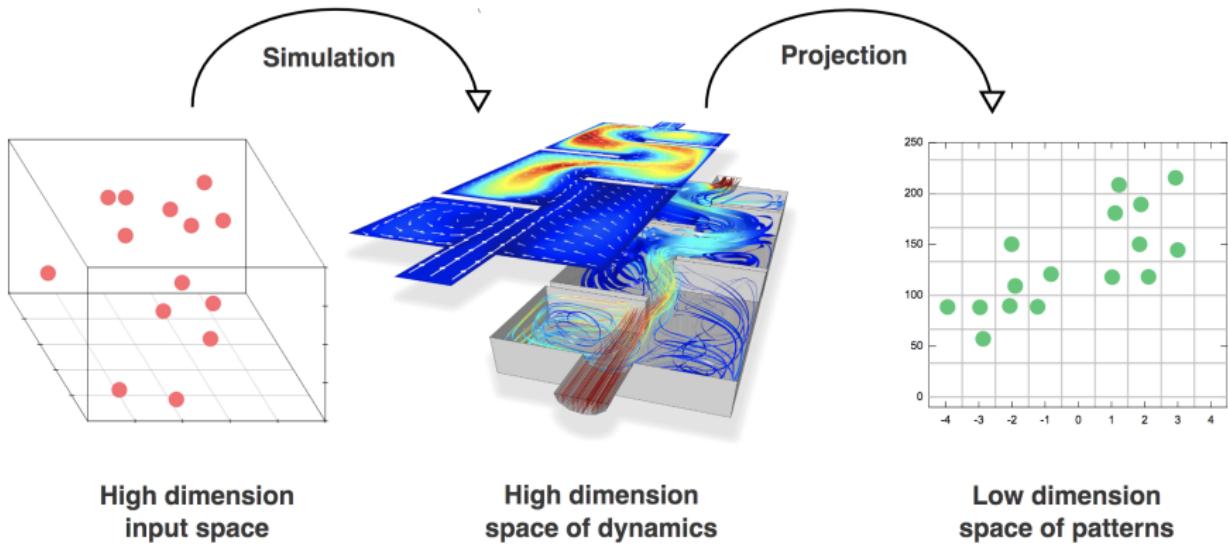


Autre exemple de multi-modélisation [Raimbault, 2018b]



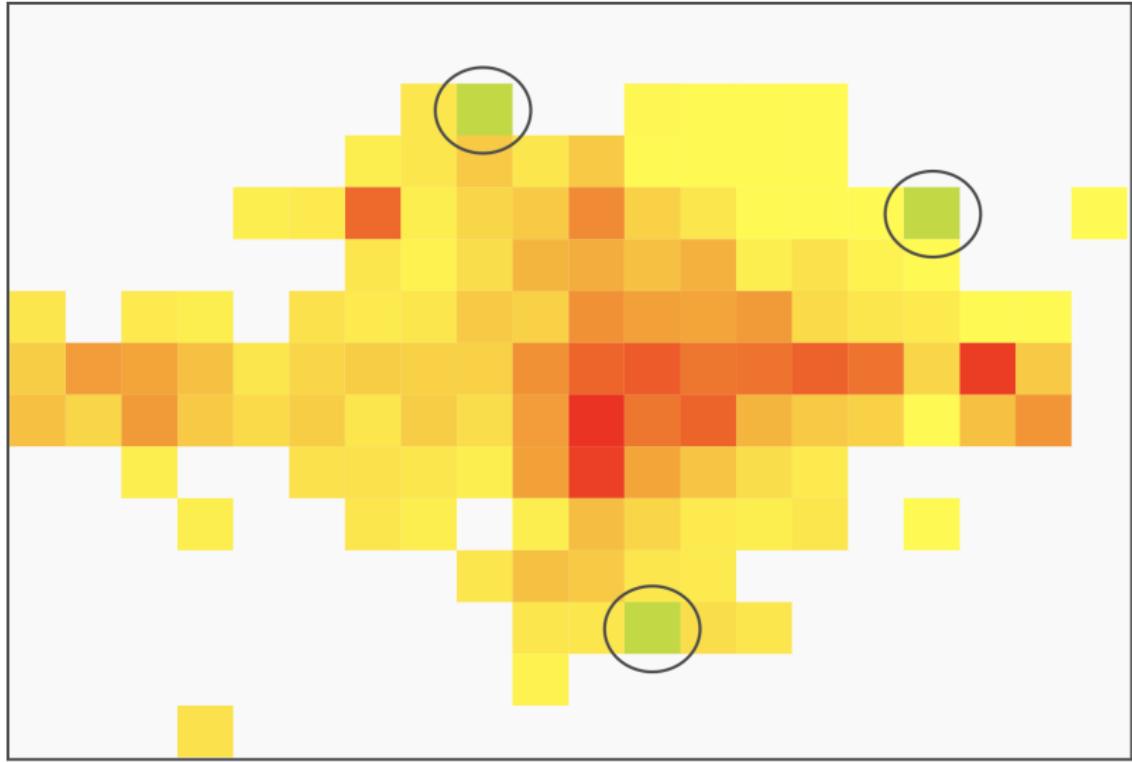
Approche alternative : recherche de motifs

[Chérel et al., 2015]

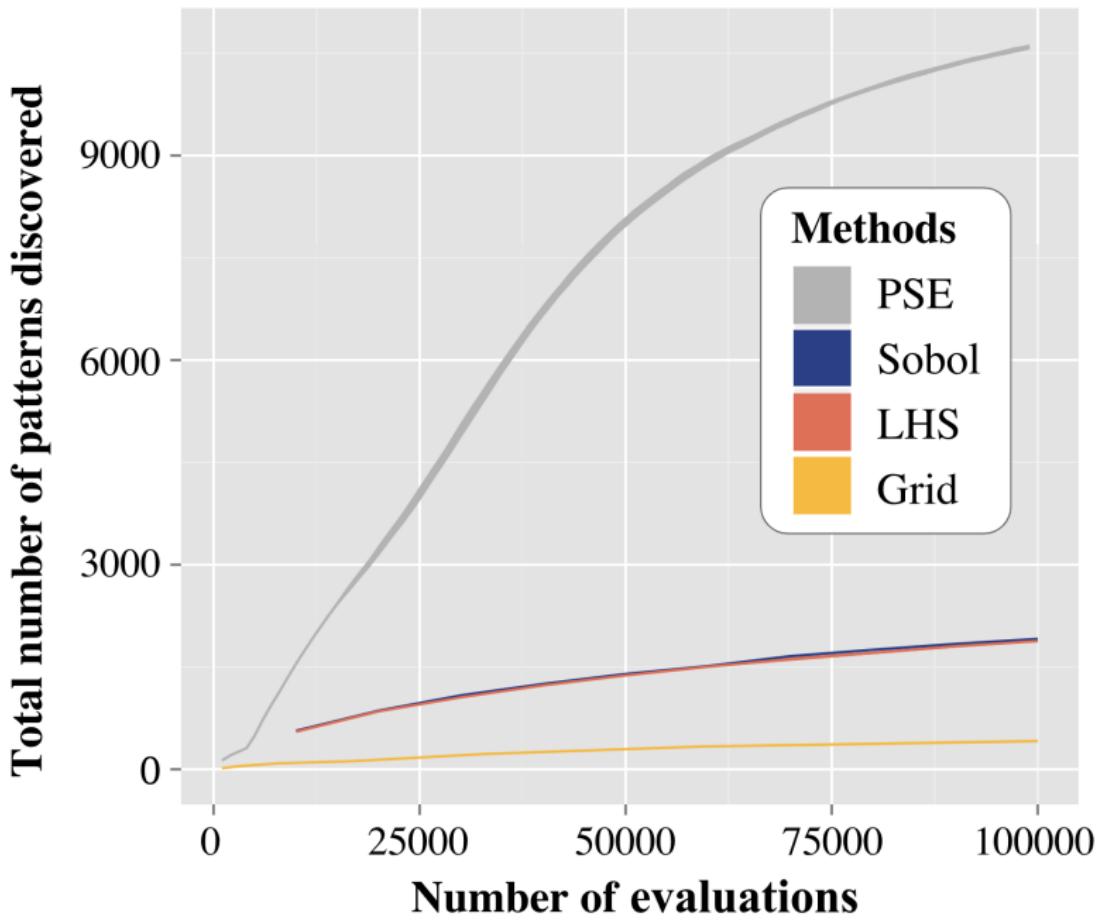


Recherche de nouveauté

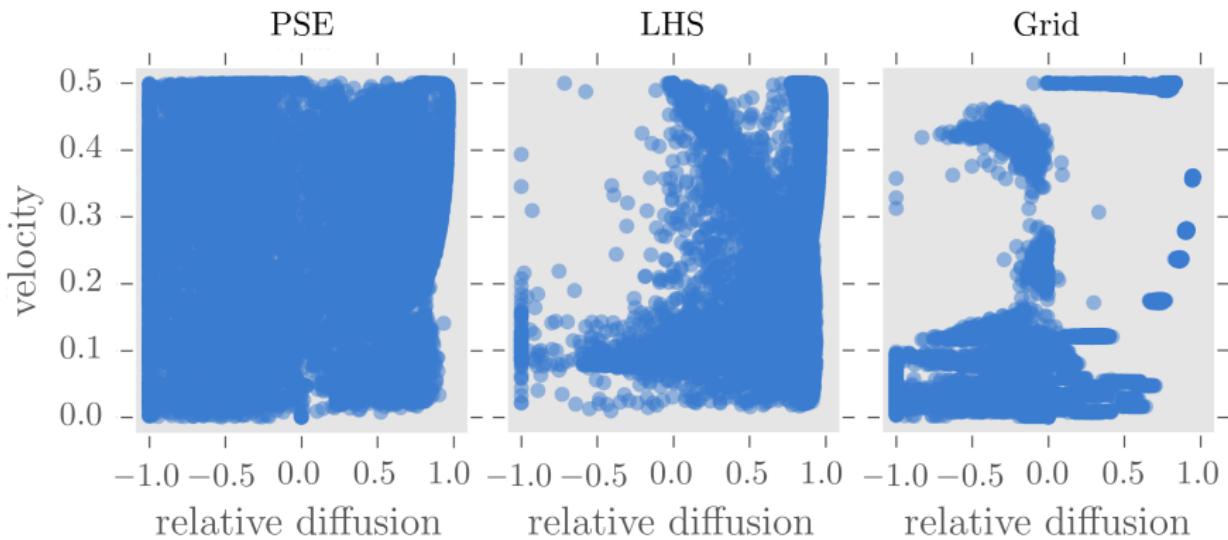
Entrées produisant des motifs rares ont une forte fitness



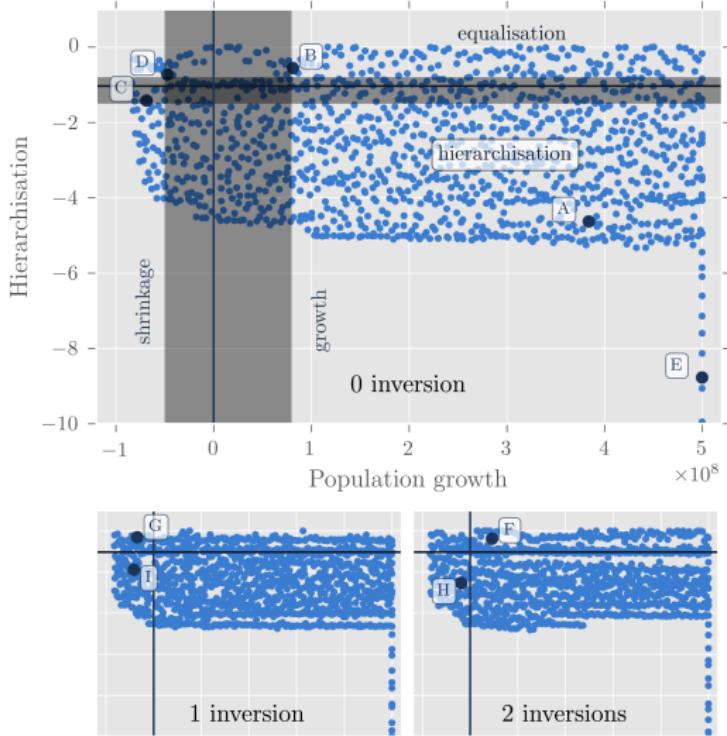
Résultats



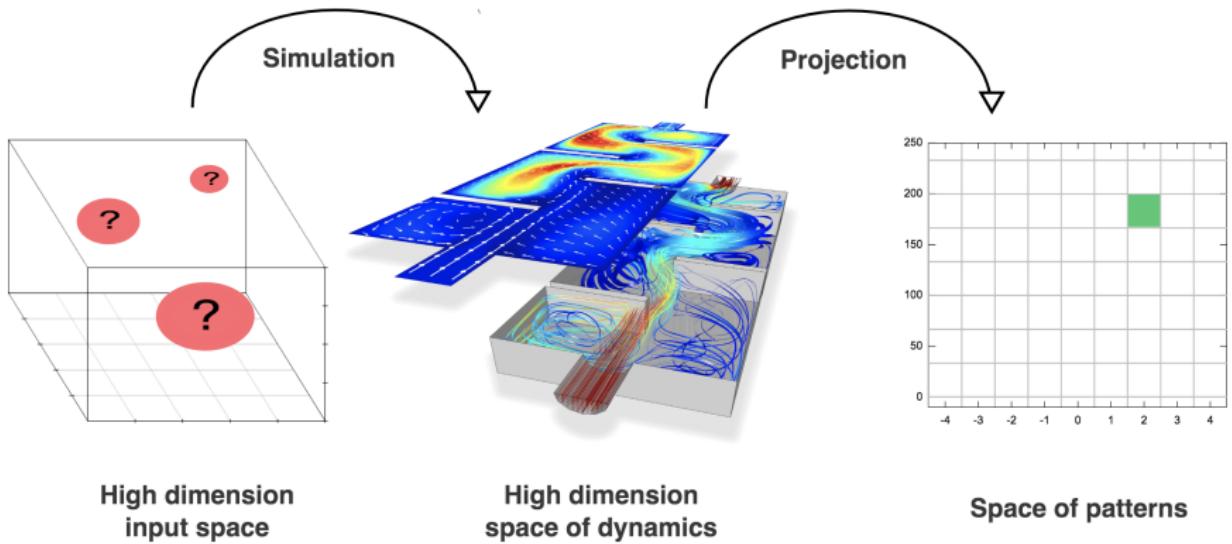
Résultats



Résultats

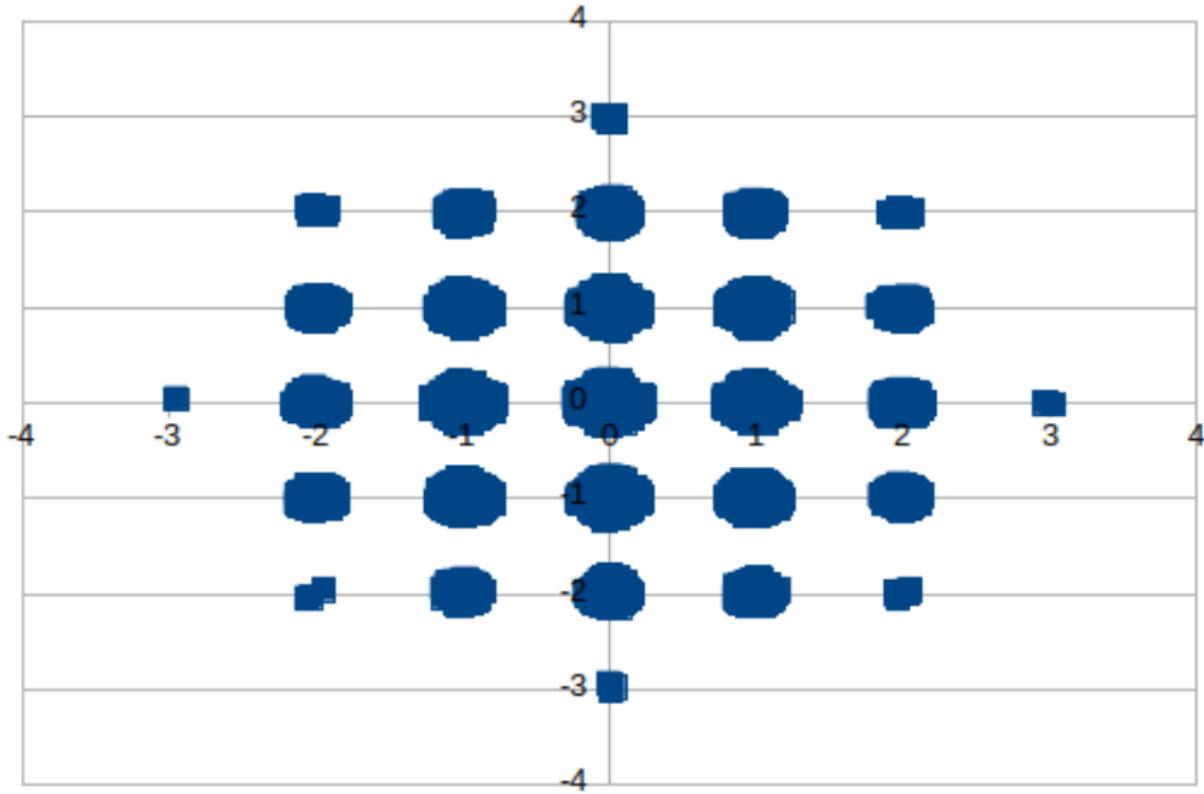


Problème inverse



Premiers résultats

Formulation: Δ pattern $< \varepsilon$



Résumé du positionnement d'OpenMOLE

Un saut qualitatif dans la connaissance extraite d'un modèle de simulation par les méthodes d'exploration des modèles

Réussites: SimpopLocal [Schmitt et al., 2015], Marius [Chérel et al., 2015], modélisation en écologie [Lavallée et al., 2018], en épidémiologie [Arduin, 2018], etc.

Caractéristiques principales:

- Role unique des axes complémentaires de l'accès aux environnements, de la disponibilité des méthodes, et de l'embarquement des modèles.
- Construction itérative et intégrée des modèles et théories, en utilisant l'ensemble des dimensions de la connaissance favorisées par la simulation et le calcul (modélisation, théorie, empirique, données, méthodes, outils [Raimbault, 2017]).
- Couplage des modèles et reproductibilité au coeur de l'approche par workflow [Passerat-Palmbach et al., 2017].

Tendances et défis actuels en (géo-)simulation

Domaines principaux: quantification de la croissance et de la forme urbaine, fouille de données spatio-temporelles, géosimulation, approches multi-scalaires [Behnisch and Meinel, 2018].

Défis:

- Cas des ABM pour la planification [Perez et al., 2016] : complexité simplifiée, multi-dimensionnalité, trajectoires faisables, planification participative
- Modèles de simulation [Banos, 2013]: interdisciplinarité, modèles ancrés dans les données, exploration des modèles, problèmes multi-objectifs, reproductibilité et réutilisation des modèles, couplage des modèles.

Futur ? [Banos, 2017] connaissance plus profonde et intégrée; [Arribas-Bel and Reades, 2018] vers une nouvelle science géographique des données ?

Question générales:

- Surajustement des modèles de simulation
- Couplage des modèles
- Comportement direct et indirect
- Stochasticité

Spécifiques aux systèmes spatio-temporels:

- Non-stationnarité spatio-temporelle
- Données synthétiques spatio-temporelles

Dans quel mesure l'ajout de paramètres améliore effectivement l'information extraite ?

→ Fondamental pour des modèles parcimonieux, qui peuvent ensuite être utilisés comme briques pour des modèles plus complexes.

- Pouvoir explicatif brutal “boîte-noire” ?
- Extension de mesures de type AIC ?
- Optimisation multi-objectif incluant degrés de liberté ?

Couplage de modèles

Définition/théorie/quantification du couplage de modèles ?

→ Fondamental pour l'interdisciplinarité, la reproductibilité et la réutilisation des modèles; pour les approches multi-scalaires (rôle de la causalité descendante); pour le comparaison de modèles.

- Notion de “force de couplage” indépendante des modèles ?
- Structures de covariance ?
- Graphes causaux ?

Comportement direct et indirect

Comment obtenir un aperçu relativement exhaustif du comportement d'un modèle de simulation fortement non-linéaire ?

→ Espace faisable et motifs inattendus (algorithme PSE [Chérel et al., 2015]); question de l'équifinalité.

- Heuristiques de problème inverse en cours d'expérimentation
- Approche itérative pour extraire les motifs principaux de comportement

Quelle place et quelle gestion de la stochasticité dans la construction et l'exploration des modèles ?

→ Fondamental pour des modèles basé sur les donnés et appliqués; pour une connaissance robuste extraite des modèles de simulation.

- Prise en compte de bruits “réels”
- Une approche par augmentation de la dimension en cours de test
- De même, une approche bayesienne de la calibration (ABC)

Modélisation spatio-temporelle: non-stationnarité

Comment inclure la non-stationnarité spatio-temporelle dans les analyses empiriques, les simulations et calibrations de modèles ?

→ Complexité intrinsèque des systèmes spatiaux; fondamental pour les approches multi-scalaires.

- Lien entre non-stationnarité et non-ergodicité
- Lien entre approches basées sur les agents et systèmes dynamiques, vers des approches hybrides [Banos et al., 2015]

Données synthétiques spatio-temporelles

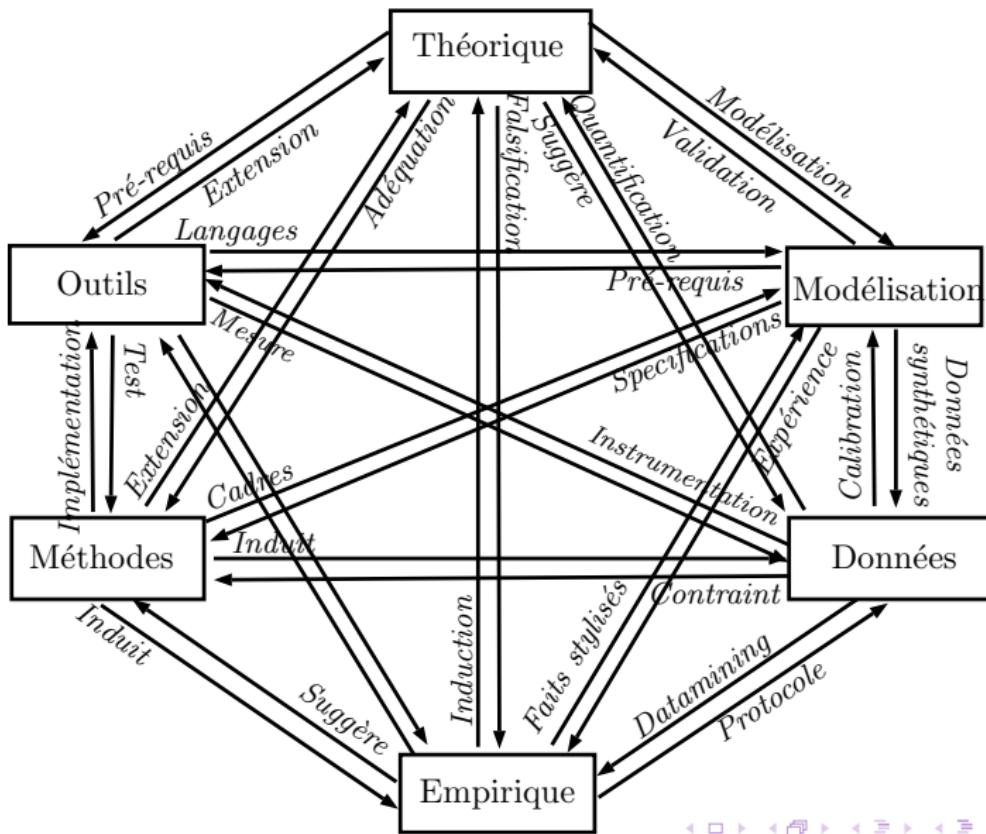
Comment étendre les analyses de sensibilité aux conditions spatiales initiales ? Comment générer des données spatiales synthétiques ?

→ Fondamental pour la robustesse des simulations spatiales

- Première approche dans le cas de grilles de densité de population [Cottineau et al., 2017]
- Extension à d'autres disciplines: écologie, géosciences
- Vers une bibliothèque générique intégrée à OpenMOLE

Cadre: domaines de connaissance

Cadre de connaissance appliqu  [Raimbault, 2017]



Un perspectivisme appliqué pour coupler les approches

Perspectivisme de Giere [Giere, 2010]: une “troisième voie” au delà de la dichotomie constructivisme-réalisme. *Tout processus de construction d'une connaissance scientifique comme une perspective par un agent pour répondre à un but par l'intermédiaire d'un medium (le modèle).*
[Compatible avec l'approche fonctionnelle de [Varenne, 2017]]

Agents cognitifs et les domaines de connaissance sont en co-évolution, par l'intermédiaire des domaines de connaissance.

Principes du perspectivisme appliqué:

- Favoriser la cohérence des perspectives et leur communication (spirale vertueuse de Banos entre disciplinarité et interdisciplinarité)
- Importance de la réflexivité pour faciliter le couplage des perspectives
- Nouvelles méthodes d'exploration de modèles accroissent l'intégration entre les domaines de connaissance
- Couplage de modèle comme un medium possible pour coupler les perspectives (hypothèse de transfert)

Encore à formaliser, spécifier des implémentations possibles, et expérimenter [in silico et in vivo].

Conclusion

Accomplissements significatifs au-delà des disciplines, construction de nouvelles pratiques de validation des modèles

→ encore beaucoup à faire : comment atteindre un réelle intégration ?
comment faire percoler les pratiques dans d'autres disciplines ? Quel
changement institutionnel ? Epistémologique ? [Pumain, 2018b]

École d'été OpenMOLE <https://exmodelo.org/>

Travaux connexes

Raimbault, J. (2017, December). An Applied Knowledge Framework to Study Complex Systems. In Complex Systems Design & Management (pp. 31-45). arXiv:1706.09244.

Raimbault, J. (2018). Caractérisation et modélisation de la co-évolution des réseaux de transport et des territoires (Doctoral dissertation, Université Paris 7 Denis Diderot). <https://halshs.archives-ouvertes.fr/tel-01857741>

Reserve Slides

References I



Arduin, H. (2018).

Modélisation mathématique des interactions entre pathogènes chez l'hôte humain: Application aux virus de la grippe et au pneumocoque.

PhD thesis, Université Paris-Saclay.



Arribas-Bel, D. and Reades, J. (2018).

Geography and computers: Past, present, and future.

Geography Compass, page e12403.



Banos, A. (2013).

Pour des pratiques de modélisation et de simulation libérées en géographies et shs.

HDR. Université Paris, 1.

References II



Banos, A. (2017).

Knowledge accelerator' in geography and social sciences: Further and faster, but also deeper and wider.

In *Urban Dynamics and Simulation Models*, pages 119–123. Springer.



Banos, A., Corson, N., Gaudou, B., Laperrière, V., and Coyrehourcq, S. R. (2015).

The importance of being hybrid for spatial epidemic models: a multi-scale approach.

Systems, 3(4):309–329.



Behnisch, M. and Meinel, G. (2018).

Trends in Spatial Analysis and Modelling.

Springer.

References III

-  Chérel, G., Cottineau, C., and Reuillon, R. (2015).
Beyond corroboration: Strengthening model validation by looking for unexpected patterns.
PLoS ONE, 10(9):e0138212.
-  Cottineau, C. (2014).
L'évolution des villes dans l'espace post-soviétique. Observation et modélisations.
PhD thesis, Université Paris 1 Panthéon-Sorbonne.
-  Cottineau, C., Raimbault, J., Le Texier, M., Le Néchet, F., and Reuillon, R. (2017).
Initial spatial conditions in simulation models: the missing leg of sensitivity analyses?
In *2017 International Conference on GeoComputation: Celebrating 21 Years of GeoComputation*.

References IV

-  Giere, R. N. (2010).
Scientific perspectivism.
University of Chicago Press.
-  Lavallée, F., Alvarez, I., Dommangelet, F., Martin, S., Reineking, B., and Smadi, C. (2018).
A dynamical model for the growth of a stand of japanese knotweed including mowing as a management technique.
In *Conference on Complex Systems 2018*.
-  Openshaw, S., Charlton, M., Wymer, C., and Craft, A. (1987).
A mark 1 geographical analysis machine for the automated analysis of point data sets.
International Journal of Geographical Information System,
1(4):335–358.

References V

-  Passerat-Palmbach, J., Reuillon, R., Leclaire, M., Makropoulos, A., Robinson, E. C., Parisot, S., and Rueckert, D. (2017). Reproducible large-scale neuroimaging studies with the openmole workflow management system. *Frontiers in neuroinformatics*, 11:21.
-  Perez, P., Banos, A., and Pettit, C. (2016). Agent-based modelling for urban planning current limitations and future trends. In *International Workshop on Agent Based Modelling of Urban Systems*, pages 60–69. Springer.
-  Pumain, D. (1997). Pour une théorie évolutive des villes. *Espace géographique*, 26(2):119–134.

References VI

-  Pumain, D. (2003).
Une approche de la complexité en géographie.
Géocarrefour, 78(1):25–31.
-  Pumain, D. (2012).
Urban systems dynamics, urban growth and scaling laws: The question of ergodicity.
In *Complexity Theories of Cities Have Come of Age*, pages 91–103.
Springer.
-  Pumain, D. (2018a).
An evolutionary theory of urban systems.
In *International and Transnational Perspectives on Urban Systems*,
pages 3–18. Springer.

References VII

-  Pumain, D. (2018b).
From erc geodiversity : Knowledge accelerator and pending questions.
In *Conference on Complex Systems 2018*.
-  Rimbault, J. (2017).
An applied knowledge framework to study complex systems.
In *Complex Systems Design & Management*, pages 31–45.
-  Rimbault, J. (2018a).
Extracting knowledge from simulation models: trends and perspectives from the viewpoint of quantitative geography.
In *Conference on Complex Systems 2018*.
-  Rimbault, J. (2018b).
A systematic comparison of interaction models for systems of cities.
In *Conference on Complex Systems 2018*.

References VIII

-  Rimbault, J., Banos, A., and Doursat, R. (2014).
A hybrid network/grid model of urban morphogenesis and optimization.
In *4th International Conference on Complex Systems and Applications*, pages 51–60.
-  Reuillon, R., Schmitt, C., De Aldama, R., and Mouret, J.-B. (2015).
A new method to evaluate simulation models: The calibration profile (cp) algorithm.
Journal of Artificial Societies and Social Simulation, 18(1):12.
-  Rey-Coyrehourcq, S. (2015).
Une plateforme intégrée pour la construction et l'évaluation de modèles de simulation en géographie.
PhD thesis, Université Paris 1 Panthéon-Sorbonne.

References IX

-  Sanders, L., Pumain, D., Mathian, H., Guérin-Pace, F., and Bura, S. (1997).
Simpop: a multiagent system for the study of urbanism.
Environment and Planning B, 24:287–306.
-  Schmitt, C. (2014).
Modélisation de la dynamique des systèmes de peuplement: de SimpopLocal à SimpopNet.
PhD thesis, Paris 1.
-  Schmitt, C., Rey-Coyrehourcq, S., Reuillon, R., and Pumain, D. (2015).
Half a billion simulations: Evolutionary algorithms and distributed computing for calibrating the simpoplocal geographical model.
Environment and Planning B: Planning and Design, 42(2):300–315.



Varenne, F. (2017).

Théories et modèles en sciences humaines. Le cas de la géographie.
Editions Matériologiques.