

# Évaluer un modèle de simulation complexe en situation d'inter-disciplinarité

Version 2019-02-22

- N.B. : Commentaires de Lena (Mars et Avril 2018) sur les parties 3.2 et 3.3 non pris en compte
- Décembre 2018 / Février 2019 : Ajout de la partie 3.1
- Janvier/Février 2019 : Corrections après commentaires de Lena sur 3.1 à 3.1.4

## Sommaire

<b>3.1 Comment évaluer un modèle ?</b>	<b>2</b>
3.1.1 Évaluation, validation, vérification... : désambiguïsation	3
3.1.2 Les étapes de l'évaluation d'un modèle	6
3.1.3 Une évaluation de la plausibilité d'un modèle : la « <i>face validation</i> »	12
3.1.4 Vers une évaluation visuelle	20
<b>3.2 Des indicateurs pour SimFeodal</b>	<b>23</b>
3.2.1 Indices et indicateurs	23
3.2.2 Hiérarchiser et catégoriser les indicateurs	29
<b>3.3 Les indicateurs et dimensions de SimFeodal</b>	<b>33</b>
3.3.1 Évaluer la polarisation des foyers paysans	33
3.3.2 Évaluer la hiérarchisation du système de peuplement	42
3.3.3 Évaluer la fixation et la dissémination du peuplement	49

### 3.1 Comment évaluer un modèle ?

Depuis les travaux précurseurs en simulation informatique (NAYLOR et FINGER 1967 ; HERMANN 1967 ; SARGENT 1979) jusqu'aux recherches contemporaines (AMBLARD, ROUCHIER et BOMMEL 2006 ; BANOS 2013 ; AUGUSIAK, VAN DEN BRINK et GRIMM 2014 ; REY-COYREHOURCQ 2015), la plupart des chercheurs ont toujours mis en avant qu'un modèle de simulation non évalué n'avait ni utilité – pour NAYLOR et FINGER 1967 notamment –, ni validité. Sans caricaturer ces écrits, on peut noter que tous cantonnent les modèles non évalués à des « jeux » ou encore, pour les plus modérés, à des outils uniquement pédagogiques.

Comme indiqué dans le **chapitre 1**, nous considérons SimFeodal comme un modèle résolument pédagogique, et l'on pourrait dès lors se passer d'en mener une évaluation quelconque. Il nous semble pour autant que l'exercice intellectuel que constitue la (co-)construction d'un modèle de simulation perdrait de son intérêt intrinsèque s'il ne donnait lieu à des procédures, quelles qu'elles soient, ayant pour objectif d'assurer une certaine qualité au modèle, à défaut de lui garantir une validation stricte.

Nous sommes en effet convaincu que même pour des modèles visant à « assister la construction de théories »<sup>1</sup> pour reprendre les termes de LAKE 2014, p. 260, ou encore, selon la classification alternative de l'auteur, pour les modèles à utilité « de développement »<sup>2</sup>, les différents outils d'évaluation permettent d'acquérir une connaissance précieuse sur l'objet modélisé, ne serait-ce que par les effets collatéraux qu'entraîne l'évaluation d'un modèle. Qu'un modèle soit statistique ou à base d'agents, de type descriptif ou explicatif, à visée pédagogique ou prédictive, ou encore constitue un modèle « hybride » entre ces catégories, un modèle de simulation demeure un modèle qu'il convient d'évaluer pour être en mesure d'en tirer des connaissances (SARGENT et BALCI 2017, p. 299-300).

Sans entrer dans les spécificités conceptuelles de ce qu'est l'évaluation d'un modèle ou de l'histoire de ces méthodes<sup>3</sup>, nous nous contenterons dans la suite de cette partie de donner une vision aussi succincte que possible de ce qu'est l'évaluation, en particulier pour en dégager les méthodes employées usuellement. Cela nous permettra en particulier de défendre et de promouvoir l'une de ces méthodes, la validation visuelle, que nous jugeons très adaptée dans le cadre de co-constructions interdisciplinaires de modèles.

---

1. « [...] Simulation models to support theory building – so-called heuristic modelling – [...] »

2. « 'developmental' utility », c'est-à-dire les modèles dont le développement et l'implémentation bénéficient aux chercheurs qui y prennent part plutôt qu'à ceux qui se contentent de les utiliser a posteriori.

3. En particulier parce que ce sujet a été très largement traité dans un travail de thèse récent au sein de notre laboratoire de recherche (REY-COYREHOURCQ 2015, pp. 58-184), travail auquel nous renvoyons vivement pour plus d'approfondissements.

### 3.1.1 Évaluation, validation, vérification... : désambiguïsation

Il nous semble important de commencer cette partie par un point de définition et de clarification des concepts mobilisés, non pas par convention d'usage, mais parce que les usages en matière d'emploi des termes d'évaluation, de validation (méthodologique, formelle...) ou encore de vérification sont particulièrement diffus et trompeurs dans la littérature relative à la modélisation, y compris dans le champ plus restreint de la simulation à base d'agents en sciences humaines et sociales.

Depuis les travaux fondateurs, dans les années 1960, la logique qui consiste à éprouver un modèle – c'est-à-dire à (1) vérifier qu'il corresponde correctement d'une part au système qu'il décrit, et d'autre part à (2) la manière dont il est décrit – donne lieu à différentes terminologies. On notera en particulier que les deux articles considérés comme pionniers, tous deux parus en 1967, reposent pour l'un sur la notion de vérification (NAYLOR et FINGER 1967), et pour l'autre sur celle de validation (HERMANN 1967), sans pour autant que la distinction entre les deux approches puisse être vue comme consistante. Quelques décennies plus tard, une fois la pratique de simulation informatique plus développée et mûre, un consensus de pratique a été adopté autour de l'expression englobante de « Validation, Verification and Testing techniques (VV&T) », par l'entremise d'une proposition d'Osman BALCI (BALCI 1994) de clarification et de définition de chacun de ces composants. Pour reprendre ses mots en une distinction devenue courante en simulation à base d'agent, la *validation* consiste à concevoir le bon modèle<sup>4</sup> – sens (1) exposé plus haut –, alors que la *verification* permet de s'assurer que le modèle est bien construit<sup>5</sup> – sens (2). Le « *Testing* » correspond aux techniques mises en œuvre, et s'applique donc indistinctement à ces deux termes (*validation* et *verification*).

En dépit de cette définition stricte, les usages persistent dans une absence de distinction formelle entre vérification et validation, le plus souvent en englobant ces pratiques dans le terme plus large et moins défini d'« évaluation ». Il n'est d'ailleurs pas rare que ces trois termes soient employés de manière interchangeable, voir intervertie, comme un recensement rigoureux des usages le démontre (AUGUSIAK, VAN DEN BRINK et GRIMM 2014). Dans cet article, les auteurs mènent une méta-analyse de la littérature sur les usages de chacun des termes liés à l'évaluation<sup>6</sup>, et en particulier de celui de *validation*, et du sens signifié par leurs auteurs respectifs. Ils en tirent le constat qu'une large partie des termes analysés a été employé par plusieurs auteurs leur affectant des sens contradictoires. Par exemple : « the term 'validation' has been given virtually any possible meaning in this context » (*Ibid.*, p. 120).

Les auteurs de cette étude puisent dans ce constat les besoins d'une nouvelle terminologie, unique et explicite, permettant de dépasser notamment les successions d'attrait et de rejet que la notion de validation entraîne de par son

4. « Model validation deals with building the *right* model. »

5. « Model verification deals with building the model *right*. »

6. *Corroboration, Evaluation, Testing, Validation, Verification et Substantiation*

positivisme. Ils proposent ainsi un nouveau terme, l'« *evaludation* », assorti à une typologie de concepts – explicitement définis (*Ibid.*, Table 2, p. 125) – liés à l'évaluation permettant d'identifier l'objet et le sujet de chacune des phases du cycle d'« *evaludation* » (voir figure 3.1).

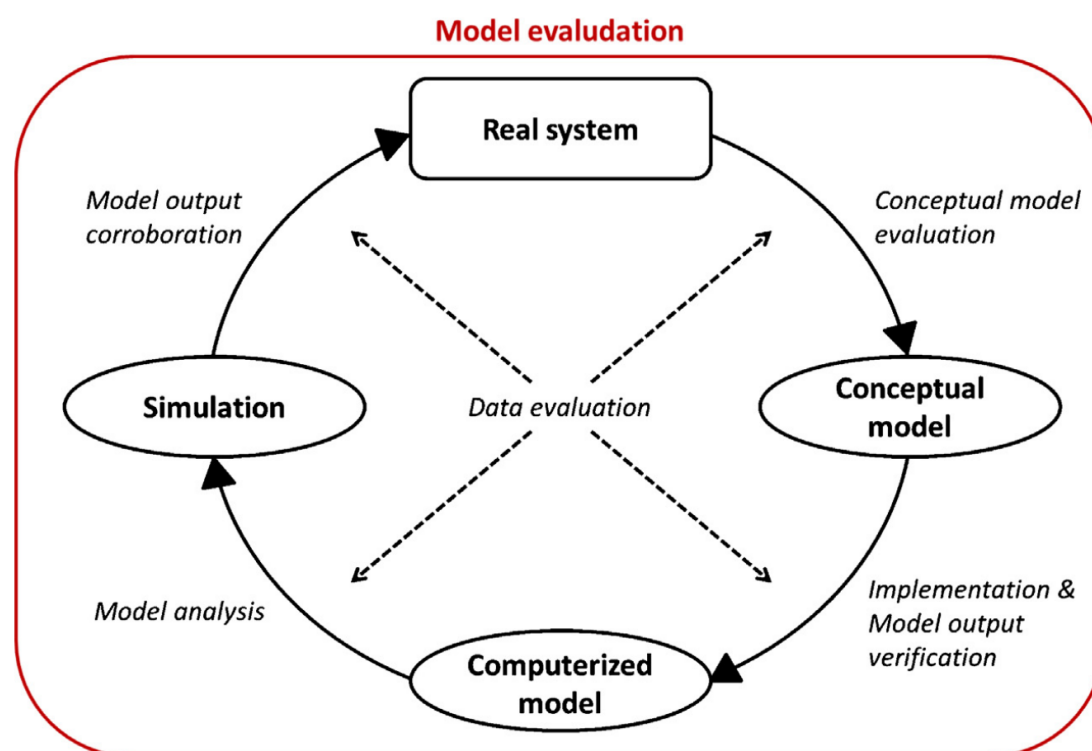


FIGURE 3.1 – Représentation schématique du cycle de modélisation et de la typologie des termes relatifs à l'évaluation de modèle, dans AUGUSIAK, VAN DEN BRINK et GRIMM 2014, Fig. 1, p. 121

Nous partageons absolument le besoin formel, identifié par les auteurs de cette étude, de définir un nouveau terme dans leur proposition, et nous souscrivons à leur approche de définition. Il nous semble toutefois, et nous le déplorons, que ce travail n'ait pas encore suffisamment percolé dans la communauté scientifique, et en particulier dans le monde francophone. Au moment de rédiger ces lignes, seuls deux auteurs en font un usage en français, au moins l'un des deux (REY-COYREHOURCQ 2015, p. 89, 436) ne le mentionnant que par recension. **Ajuster après/si retour de Sophie Plassin, interrogée par mail.**

Par soucis d'homogénéité et de compréhension par un plus grand nombre de ce travail, nous nous contenterons donc de nous inscrire dans les choix de AMBLARD, ROUCHIER et BOMMEL 2006 (voir Encadré 3.1), en particulier parce qu'ils nous semblent assez largement adoptés dans la communauté scientifique de modélisation en sciences humaines et sociales francophone, quand bien même ces concepts nous paraissent moins robustes que ceux présentés auparavant.

#### Encadré 3.1 : Les mots de l'évaluation

Pour AMBLARD, ROUCHIER et BOMMEL 2006, on emploie le concept d'**évaluation** pour définir l'approche d'ensemble, et on distingue alors

« **validation interne** » – correspondant à la *verification* définie par BALCI, c'est-à-dire s'assurer de la bonne conception du modèle, et « **validation externe** » – ce que BALCI nomme *validation*, soit l'assurance que le modèle est adapté à ce qu'il cherche à représenter.

« Il est classique de différencier deux étapes dans la validation : interne et externe.

- La phase de vérification ou **validation interne** comprend d'abord une vérification de conformité entre les spécifications et le programme implémenté et pose la question : est-ce que le modèle implémenté est bien celui que je voulais implémenter ? [...] Ensuite, la validation interne concerne la recherche et l'identification des propriétés du modèle. Dans le cas des simulations multi-agents, des preuves logiques ne peuvent être obtenues et se pose alors la question : est-ce que mon modèle possède les propriétés attendues ? Parmi ces bonnes propriétés, on considère par exemple la robustesse ou des études de sensibilité pour vérifier si les réponses sont bien différenciées sur l'espace des paramètres. Cette phase de validation interne concerne de fait une validation dans le contexte ou la logique propre du modèle.
- La deuxième phase de validation, la **validation externe**, correspond à l'évaluation de l'adéquation entre le modèle et le phénomène réel dont il est censé rendre compte. Pour cette dernière phase, la comparaison aux données empiriques ou le fait que le modèle soit capable d'exhiber des faits stylisés identifiés sur le système modélisé sont des critères clés.

Ainsi, ce qui est étudié au travers des simulations, ce sont tout d'abord les propriétés systémiques (structurelles et dynamiques) du modèle, les formes qui peuvent apparaître du fait des hypothèses posées (validation interne); ensuite est évaluée la pertinence du modèle vis-à-vis de situations que l'on souhaite représenter ou prévoir (validation externe). »

AMBLARD, ROUCHIER et BOMMEL 2006, p. 110-111

### 3.1.2 Les étapes de l'évaluation d'un modèle

Parmi les nombreuses techniques disponibles pour l'évaluation, il est courant de privilégier telle ou telle méthode en fonction de la phase d'avancement d'un modèle. Traditionnellement, l'usage veut ainsi que le modélisateur tende vers des méthodes de plus en plus formelles à mesure que l'évaluation progresse<sup>7</sup>. Les schémas des étapes d'évaluation de KLÜGL (figure 3.2) et de NGO et SEE (figure 3.3) constituent un bon résumé de cette progression – représentée de manière itérative quand bien même chaque auteur insiste sur le fait que ces étapes doivent être menées en multipliant les allers-retours entre elles – que l'on peut brièvement décrire plus avant.

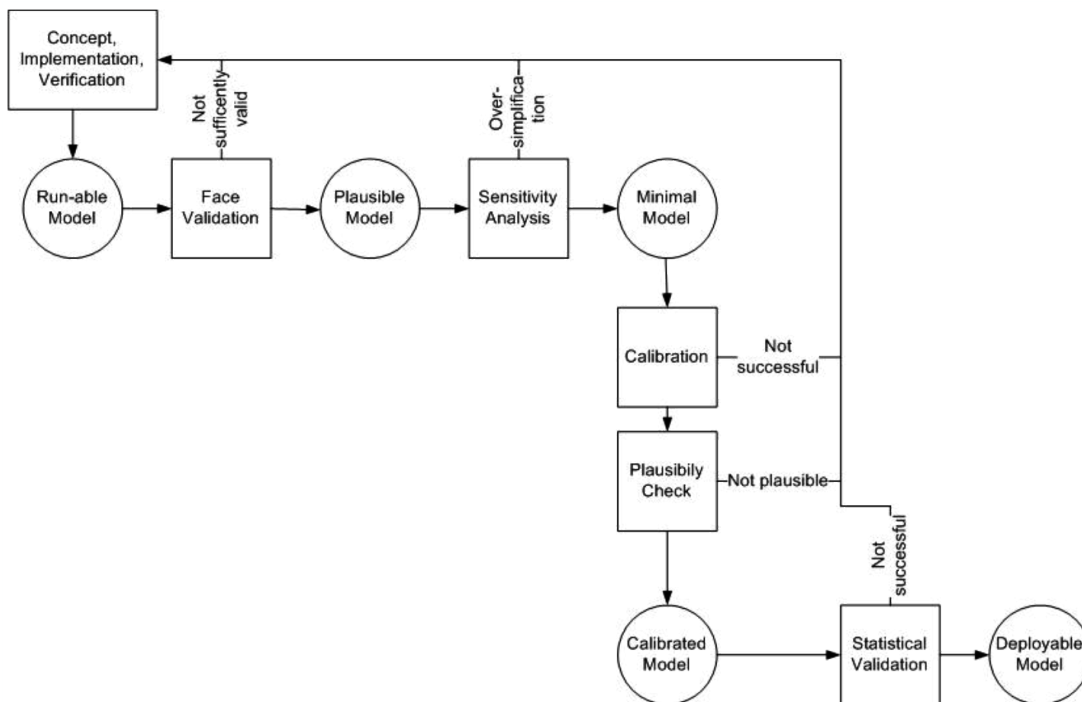


FIGURE 3.2 – « Sketch of a general procedure for validating an agent-based simulation », KLÜGL 2008, fig. 1 p. 42

7. Voir par exemple la typologie des méthodes d'évaluation, des plus « informelles » aux méthodes « formelles » chez BALCI 1994, figure 3, p. 131

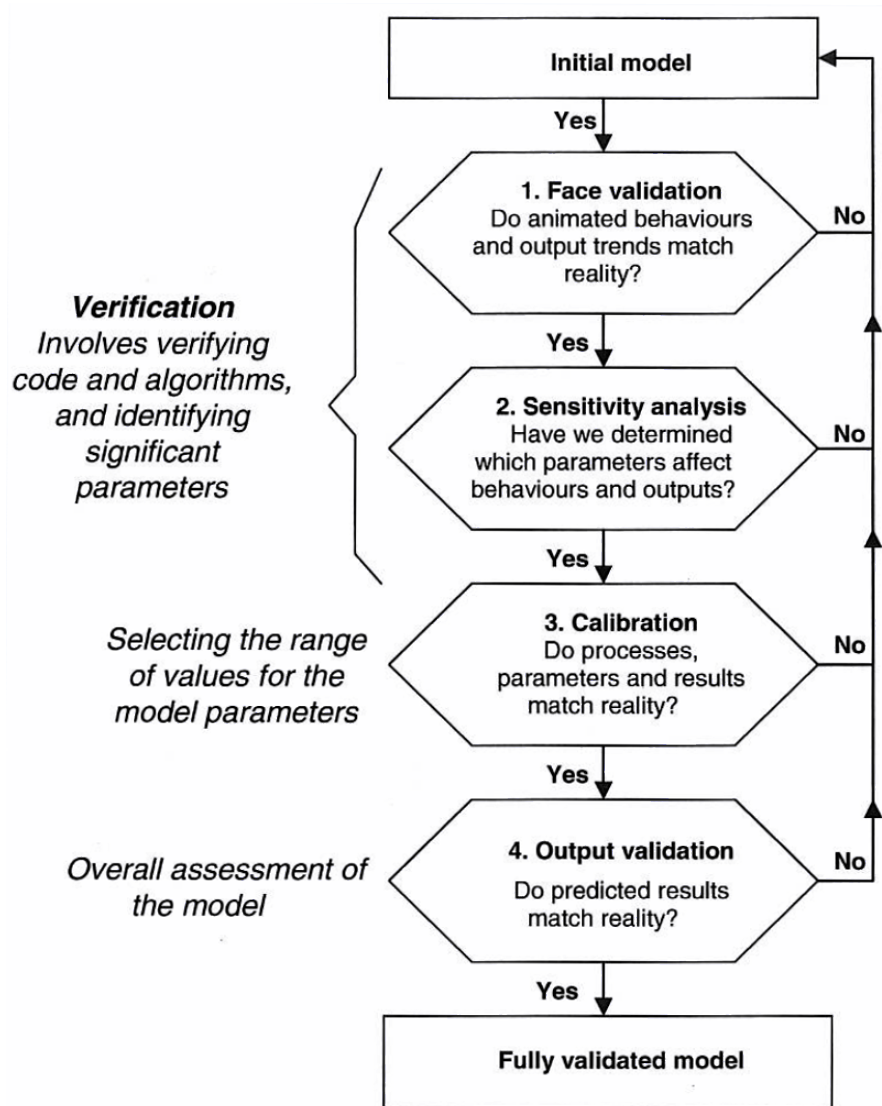


FIGURE 3.3 – « General validation process of an ABM », NGO et SEE 2012, fig. 10.1 p. 183

« **Face validation** » La première étape, de « *face validation* », consiste ainsi à vérifier visuellement, en se confortant à des intuitions sur le comportement attendu, la plausibilité du modèle. On entend par plausibilité la potentielle adéquation entre le déroulement (en termes de dynamiques observés) et l'issue (au travers des données produites) d'une simulation, et les connaissances expertes que l'on possède sur le système modélisé. Cette étape est souvent considérée comme une mesure préalable relevant du bon sens plus que de l'évaluation strictement dite, et relève autant de la validation interne que de la validation externe. Nous reviendrons plus longuement sur cette étape, essentielle d'après l'avis partagé dans la littérature mais néanmoins largement sous-exploité et méjugée à notre avis, dans les pages suivantes (voir section 3.1.3).

**Analyse de sensibilité** Quand une première version du modèle a été implémentée, il est recommandé de procéder à l'analyse de sa « sensibilité », entendue à l'égard des paramètres du modèle : en faisant varier, selon des méthodes plus ou moins complexes<sup>8</sup>, les valeurs des différents paramètres du

8. Un point plus précis y est consacré dans le chapitre 6

modèle, on peut observer l'influence de chaque paramètre sur le déroulement du modèle. Cette mesure, relative à la validation interne, intervient tôt, et doit être répétée lors de chaque modification majeure dans les mécanismes du modèle. L'analyse de sensibilité permet en effet de simplifier le modèle conceptuel et son pendant implémenté : si l'analyse révèle qu'un paramètre, quelles que soient les valeurs qui lui sont attribuées, n'a qu'un effet minime voir négligeable sur les sorties du modèles, alors il peut être judicieux de supprimer ce paramètre et/ou le mécanisme qui le mobilise. Réduire le nombre de paramètres et/ou de mécanismes d'un modèle peut sensiblement l'améliorer, selon le principe de parcimonie qui voudrait qu'un modèle plus simple soit meilleur<sup>9</sup>. Même sans aller jusqu'à ce type de découvertes sur l'inutilité de certains paramètres, l'analyse de sensibilité permet de gagner en connaissance sur le fonctionnement d'un modèle complexe et/ou non-déterministe, ne serait-ce que parce qu'elle aide souvent le modélisateur à trouver une « polarité » à l'effet des paramètres : si tel indicateur de sortie (voir section 3.2.1) croît quand on diminue les valeurs d'un paramètre et décroît quand on les augmente, alors on peut prévoir l'effet d'une modification de ce paramètre, ce qui peut éclairer le fonctionnement thématique du modèle.

**Calibration** Une fois le modèle mieux connu et surtout réduit à ses composantes nécessaires et suffisantes, on cherche à en améliorer la qualité de représentation, c'est-à-dire à faire en sorte, en jouant sur les valeurs de paramètres, que le modèle reproduise plus précisément le système qu'il décrit (validation externe) et qui correspond à l'observation empirique ou aux connaissances thématiques. Cette démarche, nommée calibration, soulève l'enjeu d'isoler, pour chaque paramètre, une étendue de valeurs acceptables et optimales. La complexité – au sens figuré – de cette étape réside dans la complexité – au sens propre – du modèle qu'il convient de calibrer : dans un modèle complexe, où chaque mécanisme peut influencer chacun des autres mécanismes de manière non linéaire, la modification des valeurs d'un paramètre doit certes modifier l'état du modèle en lui-même, mais a le plus souvent tendance à modifier par là même l'optimalité des valeurs des autres paramètres. Le problème ressemble à celui des vases communicants : pour que le modèle soit calibré, il faut que chaque valeur de paramètre soit optimale, mais la modification de chacun des paramètres peut dérégler l'effet des autres paramètres, et par la même les valeurs qu'ils doivent se voir attribuer. On ne peut donc procéder paramètre par paramètre, en les réglant un par un, au risque d'entrer dans une boucle infinie de calibration, mais au contraire, il est nécessaire de considérer l'ensemble – ou un sous-ensemble – des paramètres et de tester des valeurs qui iraient vers une optimisation du comportement du modèle. La calibration en elle-même n'est pas à proprement parler une véritable procédure d'évaluation d'un modèle : elle vise ainsi plutôt à « améliorer » le modèle plus qu'à juger de sa validité. Il s'agit ainsi plutôt d'une méthode et d'un problème d'optimisation que d'évaluation. Les différents auteurs mettent toutefois en avant son intérêt dans l'évaluation de modèle en ce qu'elle permet de garantir une meilleure validation externe du modèle puisqu'elle aboutit à l'isolation d'étendues op-

---

9. Les avis divergent nettement sur ce point, voir par exemple la définition de la *simplicité* dans AMBLARD, ROUCHIER et BOMMEL 2006, p. 120



timales de valeurs de paramètres : en menant de nouveaux tests d'évaluation (analyse de sensibilité, *face validation* etc.) (KLÜGL 2008, p. 43) sur les valeurs optimales identifiées, on peut évaluer si elles sont porteuses de sens d'un point de vue empirique ou au moins vis-à-vis de la connaissance experte du système modélisé.

**Validation statistique** La validation statistique (« *output validation* » dans la figure 3.3) est sans doute la méthode d'évaluation la plus évidente pour quiconque a été amené à concevoir un modèle. Il s'agit de confronter les données produites par le modèle – les *outputs* – aux données empiriques – ou observées – qu'ils cherchent à reproduire. Autrement dit, en termes statistiques, à s'assurer de la qualité de l'ajustement – la *goodness of fit* – des données simulées. On en mesure l'écart avec les données observées, quand de telles données sont disponibles, en cherchant à minimiser cet écart : plus l'écart est faible, alors plus le modèle parvient à reproduire les observations qui ont servi de support à sa conception et construction. La validation statistique est donc une méthode de validation externe. Les différents auteurs du champ de l'évaluation recommandent de ne mener cette étape qu'à la fin du processus d'évaluation, quand le réflexe en pratique est souvent de s'appuyer sur les données théoriques dès le début de la conception du modèle. A défaut de suivre cette recommandation, le modélisateur risque d'emmener le modèle vers du « sur-ajustement » (*overfitting*), et d'inscrire alors celui-ci dans une forme de tautologie, le modèle étant alors construit précisément pour produire ce qu'il devrait plutôt faire émerger. En conservant la validation statistique comme l'une des dernières étapes du cycle d'évaluation, c'est-à-dire en s'empêchant d'essayer de faire coller le modèle aux données qu'il doit reproduire, on s'assure de l'indépendance des mesures de l'ajustement, et on peut donc garantir une certaine objectivité quant à l'évaluation du modèle.

**Validations formelles** Absentes des deux figures (3.2 et 3.3), les méthodes de validation formelles sont toutefois porteuses d'un intérêt assez prégnant quand elles sont applicables. Ces méthodes visent à résoudre de manière analytique un modèle complexe, c'est-à-dire à mettre en équations les comportements du modèle, leurs effets d'interaction, et à résoudre ces équations pour en proposer les ensembles finis de solutions et/ou d'états. Cela requiert d'être en mesure de convertir un modèle exprimé dans un formalisme quelconque en un système d'équations dynamiques, et de parvenir en outre à résoudre l'ensemble de ce système. Dans l'évaluation de modèles au sens large, cette étape peut se révéler indispensable et assez directe, par exemple quand il apparaît nécessaire d'évaluer un modèle basé sur la théorie des jeux, que l'on traite alors sous forme d'analyse de graphes.

Dans le cas plus spécifique des modèles à base d'agent, cas dans lequel nous nous inscrivons ici, la situation est plus difficile. On emploie généralement la modélisation à base d'agents parce qu'elle encourage une approche anthropomorphique, plus aisément compréhensible et requérant moins de connaissances mathématiques que d'autres approches, mais aussi car il est terriblement complexe d'exprimer des systèmes dotés de multiples interactions, qui plus

est multi-scalaires, sous forme de réseaux d'équations. En un sens, on pourrait presque considérer qu'on fait appel à de la modélisation à base d'agents quand on ne peut mobiliser des modèles formels. Le processus qui tendrait à formaliser, mathématiquement, des modèles agents est alors intrinsèquement contre-intuitif et difficile, quand bien même certains pensent que ce n'est pas une fatalité mais une question de temps<sup>10</sup>. Si la nature même de cet exercice implique vraisemblablement que peu s'y essayent, on notera tout de même que quelques auteurs (ZHANG 2011 ; GRAUWIN, GOFFETTE-NAGOT et JENSEN 2012)<sup>11</sup> sont parvenus à résoudre de manière analytique un modèle foncièrement pensé comme un modèle agent – en automate cellulaire en l'occurrence –, le modèle de Schelling<sup>12</sup>. En dehors de l'intérêt que cela peut représenter pour la connaissance de ce modèle en particulier, rappelons tout de même que le modèle de Schelling a été énoncé à la fin des années 1960, que c'est un modèle particulièrement parcimonieux, et qu'il a tout de même fallu attendre le tournant des années 2010 afin d'y trouver une solution formelle.

Dans l'absolu, en ce qui concerne les modèles à base d'agents, il est donc difficile d'aller vers une validation formelle totale, mais on peut tout de même gagner en connaissances en mobilisant des méthodes qui cherchent à « résoudre » des modèles de simulation, de manière non formelle, par exemple autour des nombreux travaux actuels (CHÉREL, COTTINEAU et REUILLON 2015 par exemple) qui visent à l'exploration complète des comportements possible d'un modèle<sup>13</sup>.

**Quelle évaluation pour quels modèles ?** Les étapes d'évaluation énumérées ci-dessus consistent autant en une approche chronologique – relative aux phases successives de la construction d'un modèle – qu'en un gradient de qualité de l'évaluation, souvent considéré en fonction de la difficulté et du coût temporel nécessaire à chacune de ces méthodes<sup>14</sup>. Il est évident à la lecture des

10. Par exemple Alain FRANC, mathématicien : « L'une des difficultés de l'acceptation des SMA comme modèles est que ces comportements sont très mal compris mathématiquement. Il existe peu de résultats qui permettent de relier un type de règles avec un type de comportement, alors que de tels liens sont à la base du succès des systèmes dynamiques, où l'on connaît (parfois...) les gammes de paramètres qui mènent à un comportement d'équilibre, cyclique ou chaotique, et l'on sait qu'il ne peut y en avoir d'autres. [...] Il existe donc une tension entre, d'un côté, les systèmes dynamiques qui forment une théorie riche et solide de modélisation mathématique, mais pour un nombre assez restreint de situations (bien des difficultés apparaissent dans le cadre non linéaire, que l'on peut lire dans la richesse des travaux sur la modélisation de la turbulence par exemple) et, d'un autre côté, les SMA qui permettent des simulations à partir de règles plus riches et diversifiées, mais pour des résultats dont la compréhension mathématique très souvent nous échappe (il y a peu de théorèmes). On peut donc dire en résumé que les SMA sont « en avance » sur la compréhension mathématique des systèmes dynamiques et peuvent proposer des cas d'études aux mathématiciens. » (OURIACHI et al. 2018, Annexe 2, « Retour sur les SMA comme outil et cadre conceptuel de modélisation. », pp. 479-482 )

11. **Lena : regarder Axelrod et Axtell, dans leur article sur les 4 modes d'utilisation d'un SMA, article très maths.**

12. On reviendra et on présentera ce modèle de manière détaillée dans le **chapitre 4**. La description du modèle, en particulier, est effectuée dans **lien vers la description**.

13. C'est par exemple l'un des enjeux principaux, en matière de recherche, d'une plateforme telle qu'OpenMOLE (REUILLON, LECLAIRE et REY-COYREHOURCQ 2013)

14. « [One] should start with cheap tests that allow fast rejection of the model and continue investing more and more effort when the model becomes more and more valid. », KLÜGL 2008, p. 42, par exemple.

auteurs de références du champ (par exemple ceux référencés en section 3.1) que pour eux, « plus » le modèle est évalué, c'est-à-dire se confronte aux étapes d'évaluation de plus en plus formelles, plus il sera digne de confiance et donc capable d'apporter des connaissances sur les objets qu'il tend à représenter. Robert SARGENT par exemple différencie les méthodes d'évaluation selon que le système modélisé est observable ou non, c'est-à-dire « s'il est possible ou non de collecter des données sur le comportement opérationnel de l'entité »<sup>15</sup>. Pour autant, ces auteurs soulignent aussi que selon les choix de modélisation et les caractéristiques du système modélisé, toutes ces étapes ne sont pas nécessairement accessibles ou possibles.

Nous pensons qu'un autre facteur peut affecter plus fortement l'éventail des méthodes possibles d'évaluation : la parcimonie du modèle réalisé. Ainsi, avec un modèle très parcimonieux, qui s'inscrirait dans un certain purisme des méthodes « KISS », doté d'un nombre minime d'*outputs*, il nous semble que toutes les méthodes, y compris les plus formelles, sont assez simplement – si ce n'est pour la résolution analytique, on l'a vu plus haut – applicables. A contrario, un modèle très descriptif, ancré dans une approche « KIDS », fourmillant d'*inputs*, de paramètres et d'*outputs* sera bien plus complexe à évaluer de manière quantitative, ou « objective » selon les mots des pionniers de l'évaluation.

Pour illustrer l'écart entre ces approches en matière de possibilités de quantification de l'évaluation, prenons l'exemple d'une analyse de sensibilité : cette technique consiste à faire co-varier les valeurs des paramètres afin d'observer les effets que, chacun ou conjoints, ils produisent sur les sorties du modèle. Avec un modèle de Schelling, dans lequel on identifie en général trois paramètres (on y reviendra dans le chapitre 4), que l'on peut faire varier chacun selon une granularité de dix valeurs, et tenir compte de l'aléa en menant dix répliques, on peut mener une analyse de sensibilité basique au moyen de  $(10^3 \times 10)$  10 000 simulations. Dans le cas d'un modèle doté d'une dizaine de paramètres, et avec le même type d'analyse basique, le nombre de simulations nécessaire dépasserait déjà le milliard...

Pour de tels modèles, malgré tout assez peu complexes au regard de certains tenants du genre KIDS, une analyse de sensibilité rigoureuse ou une calibration fine ne sont en aucun cas envisageables selon les canons méthodologiques de l'évaluation. Dans ce cas, les théoriciens de l'évaluation recommandent, à défaut de mieux, de tout de même mener les premières étapes d'un cycle d'évaluation (PETTY 2010, p. 342) : « While moving beyond face validation to more objective and quantitative methods should always be a goal, face validation is clearly preferable to no validation at all. »

Nous ne partageons pas la réticence associée à cette recommandation, et au contraire, considérons que dans ce type de cas, des méthodes de « *face validation* » peuvent être, à condition d'être rigoureuses, extrêmement satisfaisantes et sans doute suffisantes pour évaluer un modèle de simulation, ce que nous

---

15. « The major attribute affecting operational validity is whether the problem entity (or system) is observable, where observable means it is possible to collect data on the operational behavior of the program entity. », SARGENT 2009, p. 6.

allons essayer de montrer par la suite.

### 3.1.3 Une évaluation de la plausibilité d'un modèle : la «*face validation*»

Avant d'aller plus avant dans la justification de l'utilité des méthodes de *face validation*, il convient de définir plus précisément ce à quoi la littérature réfère quand elle préconise cette méthode d'analyse de plausibilité d'un modèle.

#### 3.1.3.1 Définition

Le terme semble avoir émergé dans les années 1940, en particulier dans le champ scientifique de la psychologie et des études pédagogiques (NEVO 1985). Concept discuté et disputé dans ces domaines (MOSIER 1947), on y attribue un besoin pour les modèles, statistiques dans ce cas, de présenter à la fois une validité à l'épreuve des données, mais aussi de présenter une apparence de validité, c'est-à-dire de sembler plausibles<sup>16</sup>.

Pour illustrer ce besoin de « plausibilité », on peut prendre l'exemple des problèmes de corrélations fallacieuses (ou « *spurious correlations* ») et autres exemple de la non-équivalence de causalité et corrélation. À la suite d'un article (SHAW 2017) liant utilisation de glyphosate et nombre d'enfants diagnostiqués autistes (figure 3.4), plusieurs chercheurs ont renvoyé à l'analyse menée par un membre de l'espace de discussion *reddit* (JASONP55, voir [www.reddit.com/r/skeptic/comments/14qbn9/](http://www.reddit.com/r/skeptic/comments/14qbn9/)). Celui-ci qui proposait en effet – de manière ironique – une explication opposée, liant prévalence de l'autisme et vente de produits de l'agriculture biologique (figure 3.5).

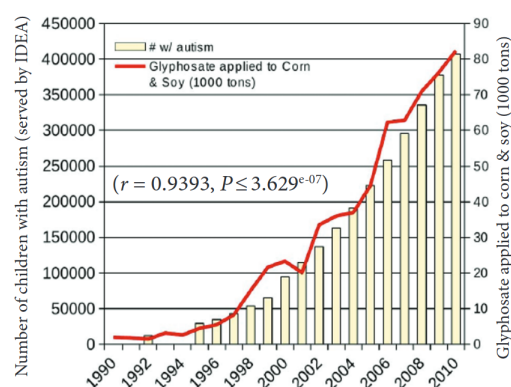


FIGURE 3.4 – Relation entre autisme et utilisation de glyphosate, d'après SHAW 2017, Figure 2, p. 51

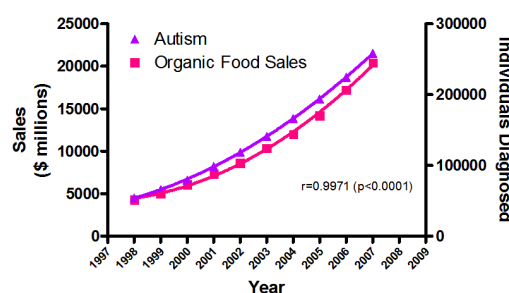


FIGURE 3.5 – Relation entre autisme et vente d'aliments « bio », d'après JASONP55, 2012, sur *reddit*

16. « In this usage, the term "face validity" implies that a test which is to be used in a practical situation should, in addition to having pragmatic or statistical validity, appear practical, pertinent and related to the purpose of the test as well; i.e., it should not only be valid but it should also appear valid. This usage of the term assumes that « face validity » is not validity in any usual sense of the word but merely an additional attribute of the test which is highly desirable in certain situations. » MOSIER 1947, p. 192

En termes de validation statistique, le second modèle est meilleur, donc plus valide : son coefficient de corrélation est de 0.997 quand celui du premier ne vaut que 0.939. En dehors du principe statistique qui veut que causalité et corrélation ne soient pas équivalents, cet exemple illustre la nécessité d'apporter un éclairage en terme de plausibilité dans l'évaluation : tout épidémiologiste pourra rejeter les conclusions du second graphique en se basant sur son intuition face à l'absence de relation empirique entre autisme et alimentation biologique. Cet exemple, trivial, illustre le besoin d'une évaluation basée sur la plausibilité, quand bien même une validation statistique aurait déjà été effectuée et jugée concluante.

L'utilisation très polysémique – et donc contradictoire – de « *face validity* », et les importants débats autour de l'usage de ce terme ayant poussé à sa désuétude<sup>17</sup>, on le retrouve pourtant au cœur de l'un des articles fondateurs de l'évaluation de modèles de simulation, où Charles HERMANN le définit ainsi :

« Face validity is a surface or initial impression of a simulation or game's realism. Probably no approach to model validity is reported more frequently than the subjective estimates of experimenters, observers, or human participants as to the correspondence between the model's operation and their perception of the actual phenomena which the game or simulation represent. [...] »

Face validity can be a significant part of a validity strategy. A quick impression that "things don't seem right" may be the only validity check possible during the actual operation of a game or simulation. Such validity judgments and their evaluation may also be part of the learning experience provided by operating models designed for instructional purposes. »

HERMANN 1967, p. 221

Quelques années plus tard, on en trouve une définition plus succincte chez un des fondateurs de l'évaluation de modèles : « Face validity is asking people knowledgeable about the system whether the model is reasonable. » (SARGENT 1979, p. 500). Cette définition introduit un aspect qui nous semble important en matière de *face validation* : il ne s'agit pas de faire évaluer la plausibilité d'un modèle par un quelconque examinateur, mais bel et bien par un expert du sujet modélisé. Ce type d'évaluation n'a donc pas uniquement vocation à démasquer des comportements contre-intuitifs, mais bel et bien à faire expertiser, par un thématicien, le déroulement et l'aboutissement d'un modèle de simulation.

Il a fallu attendre la relative démocratisation des plate-formes de modélisation à base d'agents pour qu'une auteur, Franziska KLÜGL, se penche véritablement sur l'identification et l'explicitation de la *face validity*, et en donne une définition plus précise, mais englobante car centrée sur les usages plus que sur la méthode en elle-même :

17. MOSIER 1947, p. 205 recommande même son abandon : « Since the term "face validity" has become overlaid with a high degree of emotional content and since its referents are not only highly ambiguous but lead to widely divergent conclusions, it is recommended that the term be abandoned. »

« Face validity can be seen as the result of face validation. Under this paradigm I want to subsume all methods that rely on natural human intelligence. Examples are structured walk-through, expert assessments of descriptions, animations or results. Thus, face validity shows that processes and outcomes are reasonable and plausible within the frame of theoretic basis and implicit knowledge of system experts or stake-holder. Face validation may be applied from the early phases of the simulation study under the umbrella of conceptual validations. It is often also called plausibility checking. [...]

Face validation usually plays an important role during model design. All tests based on reviews, audits, involving presentation and justification of assumptions and model structure are used for reaching this form of plausibility. »

KLÜGL 2008, p. 39, 41

La description des méthodes possibles menant à cette évaluation n'est pas en reste non plus dans cet article, puisque l'auteur identifie trois familles de cette *face validation*, chacune pouvant être menée par des experts différents<sup>18</sup> :

#### Composantes de la *face validation*

- **Évaluation du déroulement** Évaluation du déroulement d'une simulation dans son ensemble. Il s'agit ici de juger de la plausibilité des dynamiques (à l'échelle du système dans son ensemble, ou de composantes de celui-ci) reproduites dans la simulation, via une observation en direct de la simulation.
- **Évaluation des sorties** Cette approche consiste plutôt à une évaluation qualitative des sorties produites par la simulation. Cela peut prendre la forme de vérification des valeurs (approche que l'on retrouve dans les méthodes d'évaluation plus formelles, via une automatisation de ces types d'évaluation) par un expert, mais aussi d'analyse des co-variations et évolutions temporelles de différents indicateurs de sortie. L'évaluation des sorties peut être appliquée sur le système modélisé dans son ensemble, mais aussi au niveau des types d'agents mobilisés.
- **Évaluation « immersive »** Il s'agit ici d'évaluer le modèle au travers de la vraisemblance des actions et réactions individuelles des agents qui y interagissent. L'accent est donc mis sur la plausibilité du comportement des agents (niveau micro), plus que sur celle des dynamiques macroscopiques résultantes. Les experts de ces deux niveau d'observation peuvent être différents (un psychologue spécialiste des réactions individuelles en cas d'incident ne peut porter un jugement de même niveau qu'un physicien spécialisé dans les dynamiques de foules par exemple), et il faut donc, à chaque niveau d'observation du modèle, faire intervenir un expert adéquat.

18. L'énumération qui suit est une traduction libre et une reformulation partielle de KLÜGL 2008, p. 41-42

Pour l’auteur, ces trois approches d’évaluation sont complémentaires et s’inscrivent dans des temporalités différentes de la phase de vie du modèle. Elle encourage ainsi plutôt à mener l’évaluation des sorties après les deux autres, puisque ces dernières sont comparativement moins coûteuses en termes de calcul (KLÜGL 2008, p. 42).

Il nous semble que si les deux premières approches sont applicables à tout modèle, l’évaluation immersive comporte un postulat lourd sur la rationalité des comportements individuels qui n’est pas universelle dans les modèles à base d’agents. Cela se prête bien à de nombreux modèles où les agents représentent des humains dotés de comportements rationnels, ou encore des particules dont la trajectoire individuelle est prévisible en dehors des effets d’interaction. Toutefois, tout un pan de la modélisation en géographie repose sur des agents non anthropomorphiques, ou encore sur des entités primaires dont seules les interactions ont vocation à faire émerger un comportement d’ensemble. Dans le cas de SimFeodal par exemple (REF CHAP2), les comportements individuels des foyers paysans ne reposent pas sur des hypothèses de vraisemblance : le foyer paysan qui se déplace de villes en villes, parfois en faisant des allers-retours, au cours des 300 ans modélisé, ne s’appuie sur aucune connaissance empirique, et tendrait même à contrevenir aux connaissances expertes de la mobilité résidentielle des foyers paysans médiévaux. Le suivi d’un foyer paysan, isolé de ses co-agents, au cours du déroulement du modèle, par un expert thématique, n’est donc pas sujet à évaluation, au contraire du suivi des structures spatiales de niveau macroscopique engendrées par cette accumulation de déplacements.

L’évaluation immersive nous paraît donc peu adaptée à l’évaluation de plausibilité de tous les modèles. Notons cependant que même quand elle n’a pas de sens thématique, cette méthode peut toutefois s’avérer utile en matière d’évaluation interne, dans un aspect de « débogage » : quand bien même les réactions et attributs des agents ne reproduisent pas une connaissance experte, l’observation de celles-ci peut toujours servir au modélisateur pour vérifier l’absence de valeurs aberrantes ou encore la juste activation de chacun des mécanismes.

### 3.1.3.2 Limites

Comme mentionné auparavant (3.1.2), pour de nombreux auteurs (HERMANN 1967 ; BALCI 1994 ; KENNEDY et al. 2006), la *face validation* ne peut qu’être une étape préalable à des méthodes d’évaluation plus quantitatives et formelles. Les raisons données sont souvent le manque d’objectivité d’une démarche fondamentalement basée sur l’expertise et l’impression. Parmi ces auteurs, Charles HERMANN est sans doute celui qui se montre le plus méfiant vis-à-vis de la pratique de la *face validation*, en en pointant plusieurs limites :

« Although face validity has value in the early stages of model building or for quick checks during actual operation, its severe limitations should be recognized. Sometimes the experimenter will not know what behaviors are "realistic" because of his limited expe-



rience observing the actual phenomena. Participants can become interested and highly motivated in an incorrect representation of the desired environment. If the simulation involves the substitution of one property for another, some features may appear quite unreal and yet replicate the performance of the reference system for which the simulation was designed. The acceptance of face validity as a rough, first approximation might be improved if the simulator explicitly stated in advance what observations would constitute indications that an aspect of the observable universe had been successfully captured. In summary, face validity in its usual form suffers from the lack of explicit validity criteria. »

HERMANN 1967, p. 222

Ces réserves nous semblent être autant de pistes pour justifier de l'intérêt d'une démarche scientifique de *face validation*. En reprenant les critiques dans l'ordre énoncé par l'auteur, on peut y répondre ainsi :

- **Manque de connaissance experte** Cette première remarque nous apparaît comme quelque peu biaisée : si l'on confie une évaluation experte à des non experts, naturellement, cela ne peut déboucher sur une évaluation correcte du modèle. Cela est d'ailleurs applicable quelque soit la méthode d'évaluation : une expertise ne vaut que par la qualité de l'expert. De manière plus nuancée, on notera d'ailleurs que cette phrase montre ici l'absence d'un élément de définition de la *face validation* partagé par les autres auteurs : HERMANN considère par là que c'est au modélisateur uniquement de mener cette phase d'évaluation, alors que la littérature s'entend quant au fait que ce rôle échoit à des experts. Ce faisant, HERMANN se positionne dans la logique de construction de modèles par des modélisateurs, sans apport des thématiciens, et donc dans l'approche classique de séparation forte entre ces deux acteurs indispensables du modèle (voir **chapitre 1, prestation vs co-construction**).
- **Invraisemblance de certains comportements** HERMANN met en avant que dans un modèle, tous les mécanismes n'ont pas vocation à être vraisemblables. Ainsi, en mentionnant ces « propriétés de substitutions », il rappelle un élément important d'une évaluation, quelle qu'en soit la méthode : on ne doit et ne peut juger de la plausibilité que des aspects du modèle qui cherchent à reproduire un comportement plausible. Il nous semble qu'ici aussi, la critique de l'auteur revient à ignorer l'importance du dialogue entre modélisateur et évaluateur, tout en assumant cette fois que l'évaluateur ne serait pas le modélisateur : le modélisateur connaît les « substitutions » opérées dans le modèle, et se gardera donc bien de juger de leur vraisemblance. A contrario, un expert thématicien pourrait être étonné par certains comportements micro, dans la mesure où il ne connaîtrait pas les correspondances entre éléments du modèle et éléments du système modélisé. Là encore, cette limite repose donc surtout sur le choix d'un mode de construction isolé, c'est-à-dire n'impliquant pas et le thématicien et le modélisateur.
- **Explicitation préalable des objectifs** La dernière remarque de cette ci-



tation nous semble, sans conteste, être la plus importante et la plus juste. L'auteur note ainsi que la *face validation* ne peut constituer une méthode d'évaluation adaptée si l'on ne spécifie pas, en amont, les critères qu'elle doit s'attacher à examiner. C'est là encore vrai de toutes les méthodes d'évaluation, mais nous souscrivons aux remarques de HERMANN quant à l'importance primordiale que cela revêt pour la *face validation*. En matière de plausibilité, on pourrait ainsi, comme cela nous semble souvent être le cas, se contenter d'évaluer « à chaud » les différentes dynamiques et sorties d'un modèle, sans s'encombrer d'une démarche, ou feuille de route, spécifique. Le risque est alors d'introduire encore plus de subjectivité dans cette analyse, et en particulier de briser la capacité de reproductibilité ou de justification d'une évaluation : une évaluation peut être subjective tout en étant justifiée, appuyée par des arguments, et dès lors, reproductible si tant est que chacun de ces éléments soit explicités. Quand un modèle est évalué par une seule personne, par exemple un expert thématique, la nécessité d'une telle démarche est peu visible, chacun étant en capacité d'estimer qu'il sera en mesure de justifier *a posteriori* son évaluation. A contrario, quand un modèle résulte d'un travail collaboratif, qui plus est quand il implique plusieurs évaluateurs, les évaluations d'un même résultat peuvent varier. Il est donc indispensable de les expliciter autant que possible, et pour se prémunir d'un travail gigantesque d'analyse postérieure des résultats tout autant que pour se doter d'un outil de discussion et de débat commun, il apparaît primordial de fixer une grille d'évaluation, ou, en d'autres termes, d'un ensemble de critères à observer. Cela ne limite aucunement la nécessaire subjectivité et complémentarité des évaluateurs experts, mais permet au contraire d'inscrire leurs discours dans un référentiel de comparabilité.

### 3.1.3.3 Intérêts de la *face validation*

En dépit des limites identifiées ci-dessus, qui nous semblent surpassables à condition de définir une grille d'évaluation précision en amont, la *face validation* présente de nombreux atouts en dehors de la facilité de sa mise en œuvre traditionnelle.

Là où HERMANN et les auteurs classiques cantonnent la *face validation* à une étape préalable à une véritable évaluation, KLÜGL justifie l'intérêt propre de cette démarche méthodologique, y compris dans les phases plus avancées de la démarche classique d'évaluation :

« One may argue why face validity is need, when statistical validation is successfully done? Face validation assures that the processes and structures are reasonable for a human expert. Especially, when there is (semi-)automatic calibration of a simulation that is used in combination with statistical validation, a careful check of plausibly is necessary. This is in general true for all kinds of simulation, but it is particularly important for agent-based simulations.[...]

Although face validation may be informal and inconsistent, but it at least results in plausibility of modeled processes. Our experience with modeling and simulation in many interdisciplinary projects showed that even the formulation of a plausible model supports theory building and future empirical research. »

KLÜGL 2008, p. 40, 43

Pour l’auteure, la face validation complète ainsi d’autres méthodes d’évaluations mieux considérées, et nous irons même plus loin en considérant que chacune de ces méthodes peut potentiellement être améliorée en la conjuguant à une analyse de plausibilité visuelle de ses résultats.

On notera d’ailleurs que chez BALCI, dans une analyse de l’applicabilité des méthodes de « VV&T » au différentes phases du cycle de vie d’un modèle, seules 5 techniques<sup>19</sup> (sur 77 analysées) présentent la caractéristique d’être mobilisables à chacune de ces phases (voir la Figure 3.6).

---

19. Il s’agit systématiquement de méthodes « informelles » : (1) La vérification de la documentation (*Documentation Checking*); (2) la *Face Validation*; (3) les inspections de code; (4) les « revues » de modèle (*Reviews*) et enfin (5) les « procédures pas à pas » d’évaluation (*walk-through*).

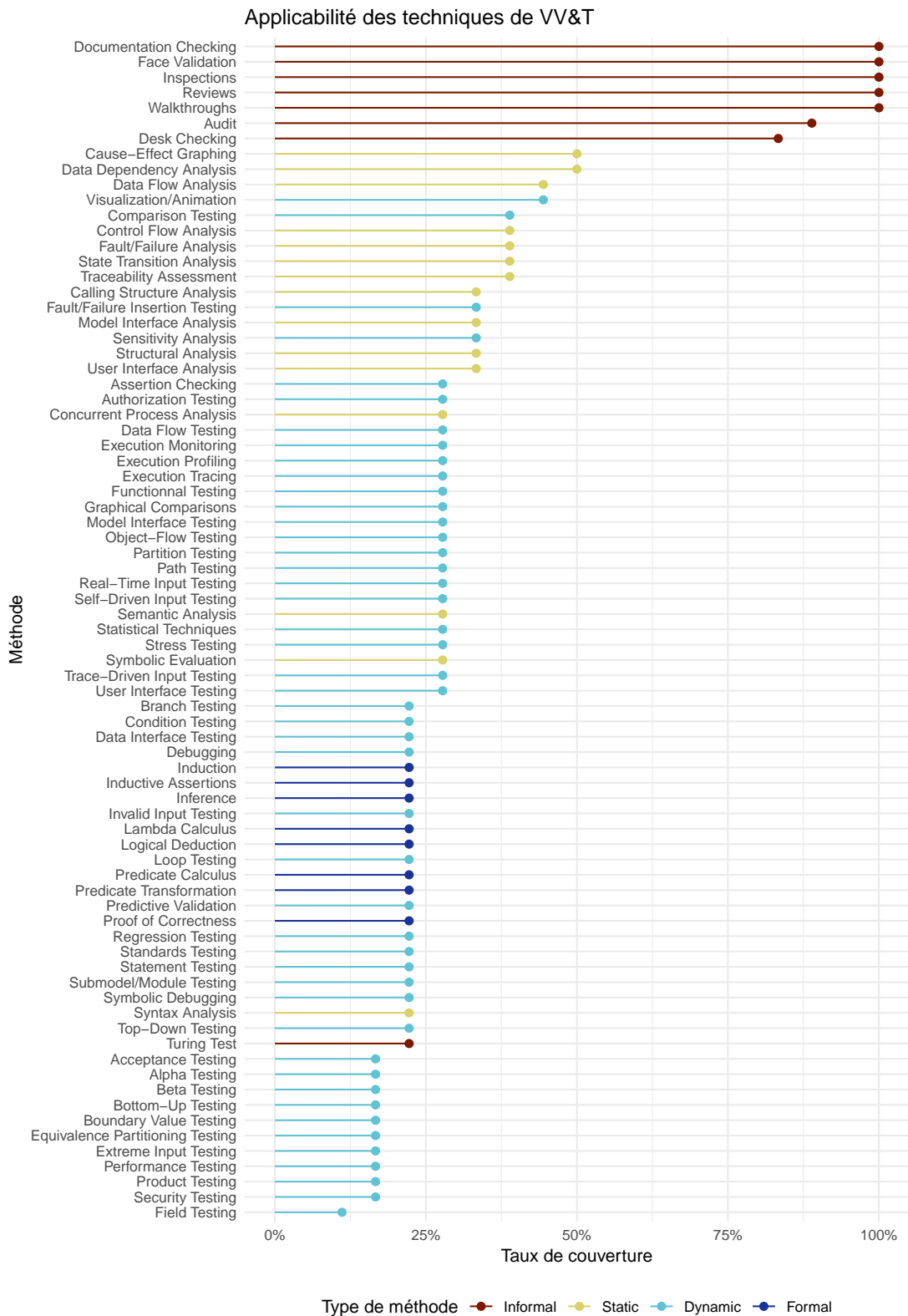


FIGURE 3.6 – Part des étapes de la cycle de vie d'un modèle pour lesquelles différentes méthodes de « VV&T » peuvent être mobilisées, selon le type de méthode (intitulés d'origine). D'après BALCI 1998, Table 3, pp. 45-47 (taux de couverture) et BALCI 1997, Figure 2, p. 139 (taxonomie des méthodes).

### 3.1.4 Vers une évaluation visuelle

Au regard de ces éléments, cantonnent-ils la *face validation* à une phase préalable mineure ou au contraire à une bonne pratique qui doit être menée lors de chacune des phases de construction et d'évaluation d'un modèle, il nous semble que l'utilité des approches de *face validation* est assez largement prouvée.

#### 3.1.4.1 Une démarche comparative

Comme identifié en limites et en intérêts, il nous semble de plus que la *face validation* souffre, en matière de réputation, d'un manque de clarification de la démarche qu'elle met en œuvre. Nous considérons que dotée d'un protocole d'évaluation rigoureux, cette méthode peut constituer une alternative crédible à des méthodes d'évaluation plus répandues, par exemple les méthodes statistiques. Ces dernières sont souvent basées sur l'analyse de l'écart entre des données empiriques et des données simulées, et cherchent à quantifier et à minimiser cet écart. La *face validation* procède à la même démarche comparative :

« Face validation is a validation method that compares simuland behavior to model results. [...] Based on their knowledge of the simuland, the observers subjectively compare the behavior of the simuland as reflected in the simulation results with their knowledge of the behavior of the actual simuland under the same conditions, and judge whether the former is acceptably accurate. Differences between the simulation results and the experts' expectations may indicate model accuracy issues. »

PETTY 2010, p. 341

Nous postulons que la démarche comparative que l'on retrouve dans l'évaluation statistique peut donc être appliquée sans recherche de quantification, c'est-à-dire en évaluant ces mêmes écarts de manière subjective et qualitative. Dans les modèles descriptifs dotés de nombreux indicateurs non résumables, tel que SimFeodal (REF chap 2), la comparaison terme à terme entre des valeurs numériques est potentiellement possible, mais non ordonnable et ne peut donc déboucher sur un indicateur unique (voir Quels types d'indicateurs pour SimFeodal ?, page 28). De plus, les données empiriques qui permettraient de mener une comparaison sont trop lacunaires et incertaines pour être jugées suffisamment fiables pour évaluer le modèle.

#### 3.1.4.2 Une démarche rigoureuse

Nous pensons toutefois que ces carences quantitatives peuvent être compensés par les connaissances expertes des différents thématiciens impliqués dans la construction et l'évaluation d'un modèle. Ainsi, dès lors que le processus d'évaluation est pensé en amont de son application, et qu'il est possible de parvenir à la création d'une grille d'analyse, c'est-à-dire à un protocole d'évaluation, il nous semble que la différence avec l'évaluation statistique est assez restreinte. Pour évaluer les modèles du type de SimFeodal, nous considérons dès lors qu'il est tout à fait possible de faire reposer cette démarche sur une

évaluation experte, s'inscrivant dans les logiques de *face validation*, et en particulier de sa composante d'évaluation des sorties (voir Composantes de la *face validation*, page 14).

### 3.1.4.3 Un nouveau terme ?

Dans la suite de cet ouvrage, nous nommerons cette approche « **évaluation visuelle** ». Ce terme n'est, à notre connaissance, que peu employé<sup>20</sup>, et le semble surtout en études environnementales, par exemple pour définir une méthode de comptage d'espèces animales et végétales (HARMELIN-VIVIEN et al. 1985 par exemple). Le pendant anglophone, la « *visual evaluation* », semble s'inscrire dans le même champ disciplinaire (HORST, ENGELKE et MEYERS 1984 par exemple), et paraît aussi assez faiblement utilisée dans un usage scientifique. Dans l'usage qui en sera fait dans ce manuscrit, ce terme est forgé au regard de la « *face validation* » naturellement, et pourrait être confondu avec. Il s'agit toutefois de s'éloigner de l'aspect « apparence » présent dans le terme – qui insiste donc sur une validité de façade –, pour embrasser au contraire la méthode visuelle. Cette dernière a fait ses preuves – dans de nombreux autres champs disciplinaires – et constitue un pan non négligeable des méthodes d'analyse, et nous pensons donc adéquat d'en faire un usage argumenté dans le domaine de l'évaluation de modèles de simulations. On notera un usage proche de cette dernière, appliquée aux modèles aussi, mais statistiques cette fois-ci, visant à l'évaluation visuelle de modèles de « *Data Science* ». Les auteurs (EILERS et al. 2017) partagent le constat d'une utilité réelle de l'évaluation visuelle (intitulée « Visual Model Evaluation » dans leur cas), et mettent une emphase particulière l'intérêt des intuitions que peuvent avoir les experts thématiques à la vue des résultats d'un modèle (p.2). Pour ce faire, ils insistent sur l'intégration d'experts dans le processus d'évaluation, et sur le besoin de communications, lors de cette phase de travail, entre les modélisateurs et ces experts<sup>21</sup>.

### 3.1.4.4 Définir l'évaluation visuelle

Pour définir cette évaluation visuelle, nous repartirons de la définition de la *face validation* sur laquelle cette méthode s'appuie.

**Définition** Il s'agit donc d'évaluer, visuellement, la plausibilité du comportement d'un modèle à partir des données qu'il produit. Cette plausibilité peut être entendue comme la correspondance entre le système modélisé et le

20. Voir par exemple le nombre d'occurrences dans Google Scholar, Isidore ou encore (Google Ngram Viewer par exemple). **Est-ce vraiment utile ?**

21. « Integrating these expert groups [data scientists and domain experts] to follow a common objective is still a major challenge today for a successful data science project in the industry and therefore a suitable field for information systems research. A collaborative analysis system addressing this issue should therefore focus on both aspects. It is important to most efficiently support human decision-makers with data-driven expert systems, and much research has been carried out in this area (Shim et al. 2002; Power 2008). But it is equally important that domain experts are also part of the system itself, e.g. by supporting data scientists with their domain knowledge when constructing the underlying models. A key success factor for this purpose is communication between different groups. » (EILERS et al. 2017, p. 2)

modèle du système, correspondance s'exprimant en comparant les données en sortie de simulation – et en les agrégeant au besoin pour tenir compte de la nécessaire réplication<sup>22</sup> – au comportement du système modélisé. Cette correspondance doit être qualifiée avant de mener cette phase d'évaluation, c'est-à-dire qu'il est nécessaire de spécifier les critères d'observation et les réponses attendues. Ces éléments, les critères d'évaluation, ne peuvent être formulés par n'importe qui : si le modélisateur autant qu'un expert externe peuvent les spécifier, il convient de s'assurer de l'expertise – thématique et de la connaissance du système tel que modélisé – de l'évaluateur. On obtient ainsi un système à évaluer au filtre d'une grille d'analyse qualitative et basée sur le visuel. Il devient alors possible d'apprécier l'écart le modèle et le système qu'il représente, sans chercher pour autant à quantifier ou à mesurer cet écart. Il s'agit en effet plutôt d'ordonner différentes versions ou paramétrages d'un modèle de simulation afin de juger de ceux qui semblent minimiser cet écart.

Cette méthode, contrairement à d'autres, plus quantitatives, permet donc au final de tirer avantage des méthodes qualitatives telles que la *face validation* – par exemple la capacité d'évaluer un modèle qui ne reposerait que sur peu de données empiriques ou encore sur des données incertaines –, tout en se confortant à une démarche d'évaluation rigoureuse, loin de l'estimation « au doigt mouillé » à laquelle peuvent donner lieu certaines méthodes reposant sur la plausibilité et l'estimation.

#### 3.1.4.5 Des critères pour l'évaluation : les indicateurs

Transition avec sous-partie suivante : à écrire après les corrections/reprises de 3.2 et 3.3

---

22. Cet aspect est discuté dans les chapitres 1, 2 et 7. En matière d'évaluation, tel que pointé par la majorité des auteurs cités dans cette sous-partie de chapitre, il est ainsi nécessaire de tenir compte de la variabilité d'un modèle, variabilité intrinsèque dans un modèle stochastique. Il n'est donc pas possible d'évaluer, visuellement ou non, un modèle stochastique sur la base d'une seule exécution. Au contraire, seule l'exécution d'un certain nombre de **réplications** (voir chapitre 1) permet de s'assurer que le comportement évalué correspond bien au comportement habituel, ou tendanciel, du modèle.

## 3.2 Des indicateurs pour évaluer le modèle SimFeodal

Le modèle SimFeodal présenté dans le chapitre 2 correspond à une « version 0 » du modèle souhaité, c'est-à-dire qu'il en constitue une première pré-version. L'ensemble des mécanismes figurant dans le modèle conceptuel ont été implémentés mais l'ensemble des liens, interactions et valeurs de paramètres ne sont pas encore stabilisés. De ce fait les résultats des simulations ne répondent pas nécessairement aux attentes définies au **paragraphe XX**. Si l'on a déjà décrit le principal objectif du modèle dans le chapitre précédent (celui de comprendre les mécanismes sous-jacents au processus de polarisation qui s'est déroulé entre 800 et 1100), il convient ici d'explicitier comment les résultats d'un tel processus peuvent être saisis. Ceux-ci sont en effet nombreux et hétérogènes, concernant aussi bien des concentrations de foyers paysans que l'émergence de pôles. Certains sont centraux, d'autres secondaires, et le modélisateur a des attentes relativement à l'ensemble des résultats obtenus en fin de simulation. La description précise de ces attentes se révèle importante dans le cadre du paramétrage – et de l'ensemble des étapes de la vie du modèle – de SimFeodal. Dans cette partie, on explicitera d'abord le sens que l'on prête à ces attentes, sous la forme « d'indices empiriques » et « d'indicateurs de sortie de simulation ». Ces indices et indicateurs sont nombreux, certains sont multivariés, et il s'agira donc de présenter des méthodes visant à réduire la complexité de ces indicateurs de sortie, en adoptant une démarche proche de ce qui se fait en statistiques : réduction de dimensionnalité et/ou catégorisation et hiérarchisation de ces indicateurs. En mobilisant ces méthodes, on pourra ensuite décrire et qualifier le comportement du modèle SimFeodal tel qu'il a été décrit, dans sa « version 0 », dans le chapitre précédent ( **chapitre 2** ).

### 3.2.1 Indices et indicateurs

On attend d'un modèle, sans entrer encore dans le détail, qu'il reproduise au moins les grands traits de l'élément empirique dont il cherche à rendre compte. Ces grands traits peuvent s'entendre de multiples manières, et se formaliser avec encore plus d'approches. Ici, nous avons souhaité proposer une dichotomie simple entre le domaine de l'empirique et celui de la simulation, en systématisant l'usage d'un vocabulaire qui est souvent employé de manière plurielle. Pour être en mesure d'évaluer la vraisemblance du comportement reproduit par le modèle sur le plan empirique, il est nécessaire de mettre en correspondance des éléments empiriques et des éléments issus de la simulation. Nous caractérisons ces éléments en deux grands ensembles : (1) **les indices empiriques**, éléments quantifiables ou au moins descriptibles émanant du domaine empirique, et (2) **les indicateurs de sortie**, variables informatiques produites par le modèle de simulation et devant pouvoir être comparés à chacun des indices empiriques. La figure 3.7 reprend, sous forme de schéma ontologique synthétique, ces deux ensembles de mesures, explicitant le vocabulaire mobilisé dans cette partie.

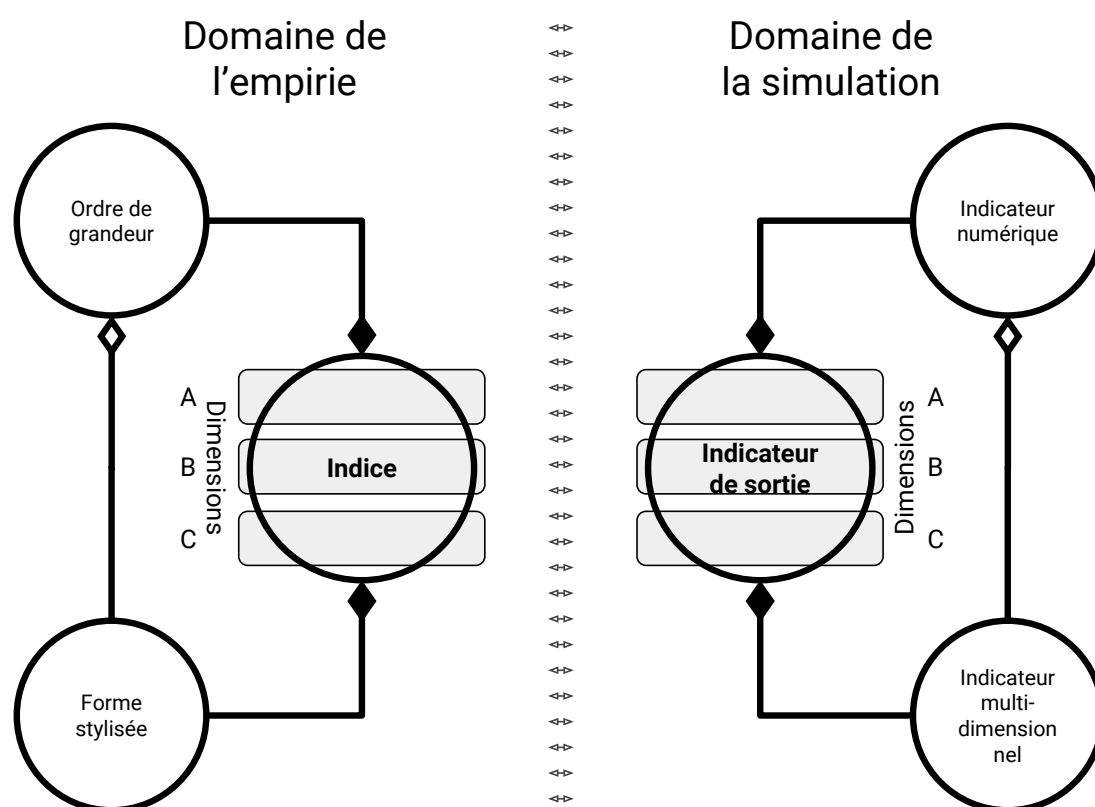


FIGURE 3.7 – Schéma de synthèse des correspondances entre mesures relevant du domaine empirique et mesures issues des simulations pour l'évaluation du modèle SimFeodal.

Lena : Paragraphe difficile

La

correspondance des éléments est représentée par une symétrie axiale, entre d'un côté des « indices empiriques » et de l'autre des « indicateurs de sortie de simulation ». Les losanges pleins désignent une relation de composition : un « indice empirique » est soit un ordre de grandeur, soit une forme stylisée. Les losanges vides indiquent une relation d'agrégation : une forme stylisée est une agrégation d'ordres de grandeurs. Les dimensions (A à C) regroupent des indices (et les indicateurs qui leur correspondent) qui peuvent être de plusieurs types, et sont elles aussi comparables et en correspondance entre les domaines.

### 3.2.1.1 Les indices empiriques

Afin d'évaluer la capacité du modèle à reproduire un phénomène observé, il est nécessaire de disposer dans le domaine empirique, de « points de repère ». Selon les modèles, ceux-ci peuvent revêtir de multiples formes et relever de l'ensemble des échelles spatiales et temporelles que l'on choisit de mettre en scène dans le modèle. Leur point commun est qu'ils doivent pouvoir être mesurés, au sens le plus large, c'est-à-dire être en capacité d'être reproduits et comparables avec d'autres mesures. Dans cette étude, on a décidé de qualifier ces points de repère d'« **indices empiriques** » et de les regrouper en deux catégories basées sur la précision avec laquelle ils peuvent être décrits<sup>23</sup>. La

23.

Lena : pas compris la note



figure 3.7 illustre cette catégorisation entre la première catégorie – les ordres de grandeur – et la seconde – les formes stylisées –.

**Ordres de grandeur** La première catégorie est constituée d'**ordres de grandeurs** empiriques estimés – avec une précision plus ou moins importante (cf. tableau 3 p. 317 du chap. TMD, à reproduire dans chap 2). Certaines valeurs empiriques sont ainsi connues, que ce soit d'après des sources primaires ou secondaires, et peuvent ainsi constituer des indices. Par exemple, on connaît avec quasi-certitude le nombre d'églises paroissiales de la région Touraine en 1100. D'autres valeurs empiriques sont en revanche issues d'estimations. Tel est le cas, par exemple, du taux de foyers paysans isolés en fin de période. Celui-ci ne peut être renseigné par des sources primaires et il a donc été nécessaire de l'estimer à partir de sources secondaires et en menant des extrapolations. Il est cependant possible de construire des indicateurs de sortie de simulation offrant une correspondance presque exacte de ces différents indices observés ou estimés (cf. 3.2.1.2, Correspondance entre indicateurs de sortie de simulation et indices empiriques, p. 26). Il est dès lors possible de mener une comparaison entre données observées/estimées et données simulées. Ces ordres de grandeur peuvent ainsi participer à l'évaluation du comportement du modèle simulé.

**Faits et formes stylisés** La seconde catégorie d'indice empirique est moins précise et ne repose pas sur une valeur observable ou estimable, mais plutôt sur la connaissance experte d'un phénomène. Il s'agit des « **faits stylisés** »<sup>24</sup>, rendant davantage compte d'une tendance dans la forme d'une relation ou d'une organisation que les indicateurs. On fait un large usage de ces faits stylisés en économie, mais aussi en géographie, par exemple quand on qualifie la tendance des systèmes de peuplement à se hiérarchiser. La valeur de la pente associée à la courbe rang-taille d'un système de villes tend ainsi vers 1 (Trouver ref, sans doute Pumain/Saint-Julien) à mesure que le système évolue et se hiérarchise. De la même manière, le modèle de transition démographique d'Adolphe Landry est un fait stylisé, énoncé à partir de l'observation de nombreuses récurrences de l'évolution des populations d'un pays en fonction de leurs taux de natalité et de mortalité. Ces exemples montrent qu'au sein des faits stylisés, il y a une certaine diversité quant à la précision de leurs énoncés : on peut quantifier précisément la courbe d'une relation rang-taille et l'allure de son évolution dans le temps, alors que la courbe logistique de la transition démographique nécessite davantage de mesures et est moins précise dans les paramètres de son énoncé. Dans notre cas d'étude, les faits stylisés sur lesquels on s'appuiera seront des d'une part des « allures » de courbes (par exemple l'évolution dans le temps d'un indicateur), d'autre part des formes de répartition spatiale, et enfin des « allures » de courbes résultant d'une composition d'ordres de grandeurs. Nous qualifierons le premier type de « **formes stylisées temporelles** » (courbe logistique estimée pour la polarisation des foyers paysans par exemple), le second type de « **formes stylisées spatiales** » (changement dans la forme d'occupation de l'espace par les agrégats entre le début et la fin de la période par exemple).

Et non sur la précision de leur connaissance, cf. section 3.2.2.1

Le troisième type, « **forme stylisée organisationnelle** », correspond à une forme repérable dans l'agrégation à l'échelle du système, comme dans l'observation des hiérarchies grâce aux courbes rang-tailles. Notons que ces formes stylisées relèvent le plus souvent d'une agrégation ou d'une composition d'ordres de grandeurs (comme figuré dans la figure 3.7) : l'évolution dans le temps de la population, par exemple, correspond à un vecteur d'ordres de grandeur, c'est-à-dire à une succession de mesures de la quantité de population pour chaque date étudiée. **Même dans le cas** d'une agrégation spatiale, par exemple quand on observe **la hiérarchie du système de peuplement**, il s'agit d'une agrégation d'ordres de grandeurs : cette forme stylisée est constituée d'un ensemble d'ordres de grandeurs, les populations de chaque agrégat de **population**.

Lena :  
pas spatial mais  
organisationnel, autre  
exemple.

Lena :  
Ajouter petites figures  
pour clarifier cette phrase

### 3.2.1.2 Les indicateurs de sortie de simulation

Les ordres de grandeur et formes stylisées évoqués relèvent du domaine empirique, c'est-à-dire qu'on dispose de données ou de connaissances d'experts à leur sujet. Afin de pouvoir les mobiliser pour évaluer la capacité du modèle à reproduire le phénomène d'intérêt, il est nécessaire de définir des **indicateurs de sortie** dans le modèle de simulation, c'est-à-dire des variables informatiques que l'on enregistrera durant l'exécution du modèle et que l'on pourra ensuite comparer aux indices empiriques définis.

**Définition** Comme pour les indices empiriques qui sont leurs équivalents dans le domaine empirique, on peut définir les indicateurs de sortie de simulation, en distinguant des formes numériques simples (des scalaires), et des indicateurs plus complexes, multidimensionnels. Ces derniers sont en effet nécessaires pour pouvoir confronter les sorties du modèle de simulation avec les formes stylisées identifiées dans le domaine empirique. Chaque indice empirique doit ainsi se voir correspondre, respectivement, un indicateur de sortie (figure 3.7).

**Correspondance entre indicateurs de sortie de simulation et indices empiriques** La correspondance entre indicateurs et indices ne correspond pas toujours à une équivalence exacte. En effet, si certains indicateurs peuvent trouver un équivalent strict dans le domaine empirique – le nombre de châteaux connus à chaque date a un sens strictement équivalent au nombre de châteaux simulés par le modèle –, d'autres correspondances sont moins directes.

Il peut s'agir de correspondances ayant trait aux mêmes éléments de base et le passage de l'indicateur à l'indice résulte alors d'une simple conversion. Par exemple, du point de vue empirique, on connaît à peu près les populations de

24. Définis ainsi par (LIVET, PHAN et SANDERS 2014) : « Un “fait stylisé” est une présentation simplifiée (i.e. taux, ratio ou écart, structure spatiale) d'une régularité empirique sur l'observation de laquelle il y a un large accord. Le terme a été popularisé en économie par Nicholas Kaldor (1961). [Les] faits stylisés peuvent être construits de la manière suivante : 1) en partant du domaine empirique, on identifie des relations saillantes ; 2) on opère quelques simplifications qui permettent d'inclure formellement ces relations dans des modèles ; 3) une fois admis que ces simplifications ne faussent pas trop les choses, on érige ces relations à la fois simplificatrices et formalisables au rang de “ faits stylisés”, dont les concepts théoriques doivent rendre compte. »

la région étudiée au début et à la fin de la période. Dans SimFeodal cependant, on ne modélise pas des individus en tant que tels, mais des foyers paysans. Le nombre de foyers paysans simulé n'est pas directement comparable à la population estimée, mais en supposant une moyenne de 4 ou 5 habitants par foyer paysan, il est possible d'en déduire un nombre d'habitants.

Dans d'autres cas enfin, le décalage entre indicateurs et indices est plus important. Il s'agit notamment de caractéristiques du système féodal que l'on sait importantes mais pour lesquelles on ne dispose pas de données facilement quantifiables. La puissance militaire des seigneurs, par exemple, est complexe à quantifier. On sait d'après connaissances expertes que la hiérarchie des puissances était forte à l'époque étudiée, majoritairement dominée par deux seigneurs (les comtes de Tours et de Blois) et assortie d'une grande quantité de petits chevaliers. On sait de plus qu'avec les liens de vassalité, les grands seigneurs disposaient des forces militaires des seigneurs qui leur étaient assujettis. Dans le domaine empirique on ne dispose pas d'éléments plus précis pour quantifier la puissance militaire des seigneurs. Dans le domaine du modèle, en revanche, on a défini un indicateur « proxy » de cette puissance à partir du nombre de foyers paysans s'acquittant de droits à chaque seigneur. De cette manière, on peut observer précisément en sortie de simulation la hiérarchie implicite entre les seigneurs reproduite par le modèle, avec une quantification de leurs puissances respectives. Ces éléments peuvent être comparés aux connaissances empiriques sur ces rapports de puissance entre les seigneurs à différents moments de l'époque féodale.

Les correspondances entre indicateurs de sortie et indices empiriques sont ainsi de nature multiple, reflétant différents niveaux de proximité entre le concept mobilisé dans le modèle et ce qui est observable dans le domaine empirique : les châteaux, entités d'intérêt dans le modèle, ont un équivalent direct dans le domaine empirique (il s'agit d'entités facilement observables et des données historiques les concernant sont disponibles) alors que la puissance militaire des seigneurs, élément moteur dans le modèle, a conduit à utiliser une variable dans le modèle pour laquelle on ne dispose pas d'observations empiriques.

La création d'indicateurs de sortie correspondant aux indices empirique permet donc de quantifier une information qui n'est pas forcément aisément quantifiable dans le domaine empirique.

**Indicateur composite** La forme numérique<sup>25</sup> (scalaire ou vectorielle) des indicateurs de sortie permet de trouver des manières plus simples d'éva-

25. **Lena :**

L'articulation des 2 § suivants ne me paraît pas évidente. Ils me semblent relever de 2 discussions différentes alors qu'ici ils paraissent liés :

D'un côté il y a les indicateurs composites/synthétiques qui sont issus d'une combinaison des indicateurs simples : ok ;

De l'autre il s'agit d'identifier une fonction objectif. Dans les modèles KISS il s'agit souvent d'une variable simple, par exemple la quantité de population.. Alors que « synthétique » dans ton texte semble beaucoup ressembler à « composite »

Est-ce que ce parag KISS fait sens ici ? Cette discussion là devrait peut-être figurer ailleurs ?

luer le modèle que d'observer l'ensemble des indicateurs. Chaque indicateur étant numérique, il devient en effet possible des les combiner au sein d'indicateurs composites, résultant en quelques indicateurs synthétiques permettant une évaluation plus rapide des résultats d'une simulation. Ces indicateurs composites sont très fréquemment utilisés en statistiques, permettant par exemple de résumer une information multidimensionnelle en un indicateur simple. L'Indice de Développement Humain (IDH), par exemple, est un indicateur composite dépendant de l'espérance de vie à la naissance, du niveau d'éducation et du niveau de revenu de chacun des pays caractérisés. On le trouve très souvent utilisé, parce qu'il permet de résumer le niveau de développement d'un pays en agrégeant trois dimensions majeures, l'aspect sanitaire, culturel et économique.

**Indicateur synthétique** En renforçant cette logique de synthèse de plusieurs dimensions, on peut aller plus loin dans la définition d'un unique indicateur, synthétique, permettant d'évaluer la qualité de représentation d'un modèle. Là aussi, c'est une pratique très fréquente, qui plus est dans le domaine de la simulation informatique en particulier sur des modèles de type « KISS » (ref. à chap 2 là ou ce sera abordé). Il s'agit alors de définir une « fonction objectif », ou « fonction de *fitness* », composée d'une pondération des quelques indicateurs composites qui auront été identifiés. Être en mesure d'évaluer un modèle à l'aide d'un unique indicateur a des avantages majeurs en pratique, puisque cela permet par exemple d'explorer et de paramétrer un modèle de simulation de manière entièrement automatique (trouver refs dans JASSS, dans Rey ou Schmitt) puisqu'on peut alors générer une cartographie simple des résultats du modèle en fonction des valeurs de paramètres utilisés.

Ces indicateurs composites et synthétiques résultent d'une quantification des autres indicateurs (excluant donc les formes stylisées qui sont plus libres d'interprétation), et apportent un grand confort dans le paramétrage d'un modèle de simulation.

**Quels types d'indicateurs pour SimFeodal ?** SimFeodal n'est pas adapté à de tels indicateurs : une large partie des faits stylisés et ordres de grandeur mobilisés proviennent de connaissances expertes, et les thématiciens qui les ont consolidées rechignent à créer de tels indicateurs composites, en ce que cela demande de pondérer précisément l'importance de chacun des indicateurs par rapport aux autres. Pour pouvoir pondérer cette importance, il faudrait de plus que les différents indicateurs mobilisés présentent le même niveau de certitude et de variabilité dans leurs résultats, ce qui est peu le cas des indices empiriques – et donc des indicateurs de sortie – choisis dans le cadre du modèle SimFeodal.

On aurait ainsi pu créer quelques indicateurs synthétiques<sup>26</sup>, mais ceux-ci ne prendraient en compte qu'une faible proportion du comportement attendu du modèle, résultant en une forte perte du pouvoir explicatif attendu du mo-

Lena :  
Pas forcément. En fait : La pondération concerne les indicateurs de sortie ; - l'incertitude est relative aux indices empiriques

26. Lena :  
Ou composites ?

dèle<sup>27</sup>. Par exemple, pour caractériser la polarisation du système de peuplement, il pourrait suffire de définir un indicateur composite fonction du niveau de concentration – le taux de foyers paysans dispersés –, du nombre de pôles et de l’espacement moyen entre les agrégats. Les valeurs de l’indicateur généré pourraient renseigner efficacement sur la capacité d’un ensemble de valeurs de paramètres à reproduire le phénomène de polarisation attendu. Cette information serait cependant grossière, dans la mesure où seraient agrégées dans le groupe des « simulations réussies » des configurations extrêmement diverses. L’information fournie risquerait alors d’être très éloignée des connaissances empiriques des thématiciens : une information multivariée ne peut pas toujours être résumée, en gardant tout son sens, par une seule variable (de manière univariée).

On a donc fait le choix d’évaluer SimFeodal en conservant des indicateurs de sortie non composites. Ce choix implique toutefois un problème majeur dans l’analyse des sorties de simulation auquel la réduction de dimensionnalité est une réponse : il est plus simple d’analyser quelques indicateurs plutôt qu’un grand nombre d’entre eux.

Lena :

On a l'impression que tu viens d'expliquer que ce n'est pas intéressant !

Revoir cette formulation donc 😊

### 3.2.2 Hiérarchiser et catégoriser les indicateurs

SimFeodal s’appuie sur une dizaine d’indicateurs numériques, ainsi que sur plus d’une trentaine d’indicateurs multidimensionnels. Tous ces indicateurs ne présentent pas le même degré de certitude, la même échelle d’observation, et surtout, le même niveau de précision sur les phénomènes modélisés. A chaque changement dans le modèle, pour une évaluation complète de la capacité de cette version à reproduire les indices empiriques, il faudrait donc observer et analyser chacun de ces nombreux et divers indicateurs. Dans le contexte du paramétrage d’un modèle s’appuyant sur une logique itérative et incrémentielle (voir Encadré 3.2), on imagine bien que cela n’est pas possible : le nombre d’indicateurs est bien trop élevé pour avoir rapidement une vision globale de la qualité de représentation du modèle. Il faut dès lors, comme pour toute analyse synthétique, concevoir une hiérarchie d’observation et d’utilisation des indicateurs : il ne sera pas nécessaire d’analyser chacun des indicateurs dans la plupart des cas, seuls les indicateurs jugés plus importants pourront être analysés. Les indicateurs de moindre importance ne seront mobilisés que pour départager des situations dont la différence ne serait pas suffisamment explicitée par l’usage des indicateurs principaux.

#### 3.2.2.1 Incertitude

Dans le modèle de simulation, les indicateurs de sortie sont à analyser en tenant compte de la précision des indices qu’ils représentent. Il ne faudra

27. Lena :

Pas sûre de comprendre...

En fait il est difficile de créer ces indicateurs si les thématiciens ne peuvent fournir une pondération qui fasse sens pour eux. La combinaison des variables ne fait simplement pas sens pour eux. C’est plutôt cela le pb non ?



ainsi pas étudier la croissance du nombre d'agrégats au cours de la simulation de manière fine, par exemple en étudiant le coefficient directeur de la courbe, quand l'empirie ne donne quasiment aucune information à ce sujet si ce n'est qu'il y a bien plus d'agrégats en fin de période qu'au début. On peut vouloir quantifier la précision de ces données, par exemple à l'aide des méthodes développées dans le champ des observations floues et/ou incertaines (voir par exemple le travail de Cyril de Runz sur les données « imparfaites » (DE RUNZ 2008)). Cette quantification de l'incertitude pourrait alors servir de base à l'établissement d'une hiérarchie des indicateurs : on analyserait en premier lieu l'écart entre les ordres de grandeurs empiriques bien connus (cf. **tableau du niveau de certitude des objectifs**) et les indicateurs calculés sur les données simulées. Les ordres de grandeur plus incertains seraient analysés dans un second temps (augmentation de la charge fiscale entre 800 et 1100 par exemple), et les formes stylisées viendraient enfin clore cette hiérarchie d'indicateurs. Toutefois, SimFeodal se caractérise d'une part par une très forte hétérogénéité dans les niveaux de connaissance des ordres de grandeurs et faits stylisés modélisés, et d'autre part, se voulant un modèle théorique (**A dire spécifiquement dans le chapitre 2; y faire une ref ici**), il n'y a pas d'obligation de « coller aux données » à tout prix : la vraisemblance d'ensemble du modèle compte bien plus que la précision de chacune de ses composantes.

### 3.2.2.2 Catégoriser les indicateurs : définir des dimensions d'analyse

En présence de plus d'une quarantaine d'indicateurs, il est toutefois nécessaire, a minima, d'organiser leur analyse. On a vu qu'il n'était pas justifié de mener cet ordonnancement à partir des propriétés intrinsèques des indicateurs du modèle. Au contraire, et cela porte bien plus de sens vis-à-vis du rôle d'un modèle, la hiérarchisation des sortie du modèle doit suivre la hiérarchie implicite qui structure les hypothèses et objectifs du modèle en lui-même. Ces hypothèses et objectifs sont multiples dans SimFeodal, et dès lors, une hiérarchie globale ne peut être définie. Il convient donc de catégoriser les indices empiriques – et les indicateurs de sortie de simulation leur correspondant –, avant d'organiser, au sein même de ces catégories, les indices les caractérisant. La hiérarchisation des indicateurs se fera donc relativement à chacune de ces catégories.

Dans le chapitre précédent (**ref chap 2**), nous présentions les principales dynamiques à reproduire avec le modèle SimFeodal : (1) polarisation, (2) hiérarchisation et (3) fixation des foyers paysans. En postulant que ces dynamiques sont caractéristiques du modèle, on peut s'appuyer sur cette triade pour caractériser les sorties du modèle, c'est-à-dire mener la confrontation entre indices empiriques et indicateurs de sortie. En reprenant ces catégories, que l'on nommera **dimensions** (voir figure 3.7), on va donc répartir chacun des indicateurs dans la dimension qu'il sera le mieux en mesure de décrire. Cette répartition n'a pas à être égale, chaque dimension pouvant s'appuyer sur un nombre différent d'indicateurs. De même, chaque dimension sera composée d'indicateurs dotés d'une qualité de représentation ou d'un niveau de certitude hétérogène. Le seul point commun des indicateurs de sortie de chaque

dimension doit être thématique. Les trois dimensions choisies – polarisation, hiérarchisation et fixation –, et les indicateurs qui les caractérisent dans le modèle, sont dès lors considérés comme les trois dimensions d’analyse des sorties de SimFeodal.

### 3.2.2.3 Hiérarchiser les indicateurs dans chaque dimension

Chacune de ces dimensions s’applique à plusieurs types d’agents du modèle. Pour définir la hiérarchie interne aux dimensions, on retiendra les agents les plus impactés par les dynamiques correspondant à ces dimensions : la polarisation, par exemple, peut être observé depuis le point de vue de ce qui polarise (les attracteurs) tout autant que de ce qui est polarisé (les foyers paysans). On aura alors tendance à examiner d’abord un indicateur de sortie caractéristique mono-dimensionnel, caractéristique de la structure dans son ensemble à son état final. Les indicateurs de sortie représentatifs des dynamiques, par exemple les indicateurs multi-dimensionnels, ayant mené à cette structure finale, seront étudiés dans un second temps. Dans cet exemple, on analysera donc d’abord le résultat effectif de la polarisation, c’est-à-dire la concentration des foyers paysans en agrégats, avant d’observer la répartition et les diversité des attracteurs ayant entraîné ce phénomène. On peut dès lors définir des « indicateurs principaux » pour chaque dimension, représentatifs des grands traits des structures auxquelles on souhaite aboutir en sortie de simulation, et des « indicateurs secondaires », permettant d’affiner l’évaluation de chacune de ces dimensions.

### 3.2.2.4 Une hiérarchie mouvante

Notons que l’analyse des indicateurs de sortie suit une hiérarchie parfois mouvante, et en tous les cas, assez peu quantifiable : si l’ordre d’observation est plutôt stable, l’importance que l’on portera à chacun des indicateurs peut varier. Les indicateurs principaux de chaque dynamique sont ainsi « incontournables », c’est-à-dire qu’un résultat trop loin de celui des indices empiriques est disqualifiant. Parmi les indicateurs secondaires, il n’est pas toujours possible, d’après les connaissances des experts sur le sujet, d’établir une priorité ou une pondération de chaque indicateur. L’évaluation de la polarisation par exemple (page 33), se définit principalement par rapport à un indicateur principal – le taux de foyers paysans dispersés –, mais selon les résultats des autres indicateurs de sortie, on leur attribuera une importance variable. L’étude de la dispersion des agrégats et pôles peut ainsi se révéler plus importante que celle de l’évolution du nombre d’agrégats selon les paramètres que l’on souhaite ajuster, ou se montrer tout au moins plus différenciante selon l’état du paramétrage.

#### *Encadré 3.2 : Incrémentalité des indicateurs*

De la même manière que les paramètres et mécanismes d’un modèle de simulation tendent à évoluer<sup>28</sup> au cours du temps de la construction, souvent afin d’affiner un comportement observé, les indicateurs de sortie

sont amenés à évoluer aussi.

Ainsi, en cas de modifications fines du modèle, il est fréquent que les indicateurs initialement choisis ne suffisent plus à départager des versions du modèle quant à un phénomène spécifique. Par exemple, quand on observe le phénomène de polarisation dans les sorties de SimFeodal, l'indicateur du nombre d'agrégats est extrêmement synthétique et informatif jusqu'à ce que l'objectif soit atteint ou que les modifications ne parviennent plus à le faire évoluer. À partir de ce moment, afin d'améliorer la vraisemblance de la situation simulée par le modèle, on peut se focaliser sur la distribution spatiale de ces agrégats, par exemple pour vérifier qu'ils sont bien répartis de manière homogène dans l'espace, et non trop concentrés.

L'observation de la répartition spatiale requiert certes de nouvelles analyses, mais surtout, par exemple, d'enregistrer les positions des agrégats au cours du temps. Si cet indicateur de sortie n'était pas utile avant cela, il n'y avait aucun intérêt à l'enregistrer. Il faut donc adapter l'implémentation du modèle pour générer, faire évoluer et enregistrer une nouvelle variable informatique correspondant à cet indicateur. Dès lors, on pourra composer un nouvel indicateur synthétique, qui, dans cet exemple, pourrait prendre la forme d'un indice de concentration spatiale).

Ce procédé incrémental dans la construction des indicateurs est très fréquent, mais pose toutefois un problème majeur : sauf à adapter chacune des anciennes versions du modèle implémenté pour y ajouter l'enregistrement des nouveaux indicateurs nécessaires, on ne pourra rendre strictement comparable les sorties de toutes les itérations du modèles informatique. Et même alors, il faudrait ré-exécuter des répliques de chaque version du modèle implémenté à chaque ajout d'indicateur, quand bien même les indicateurs présent initialement étaient jugés suffisants. Un dernier obstacle est plus gênant : certains indicateurs sont spécifiques à des mécanismes, et en cas de changement de ces derniers, ils peuvent ne plus être calculables ou simplement comparables. Par exemple, des versions antérieures du modèle enregistreraient les comportements individuels des foyers paysans quant à leur « choix » de déplacement, selon qu'ils étaient à l'origine localisés dans un agrégat ou dispersés. Une simplification du modèle a abouti à la modification des règles différenciant les possibilités de déplacement : on n'observe plus si le foyer paysan est dans un agrégat, mais plutôt s'il est dans un agrégat doté d'un pôle d'attraction. Dès lors, les analyses basées sur les choix de déplacement des foyers paysans selon leur origine ne sont plus comparables avec celles des versions antérieures au changement dans le modèle, quels que soient les détails d'implémentation de ce dernier.

Ces éléments expliquent que dans les résultats de chaque étape du paramétrage du modèle, on ne présente pas systématiquement l'ensemble des indicateurs, y compris quand ceux-ci pourraient être plus pertinents que les indicateurs présentés.

28. De manière incrémentielle et itérative, voir [dans le chapitre x ?](#) et THOMAS 2012, <http://itsadeliverything.com/revisiting-the-iterative-incremental-mona-lisa>



## 3.3 Les indicateurs et dimensions de SimFeodal

### 3.3.1 Évaluer la polarisation des foyers paysans

La polarisation des foyers paysans dans l'espace du modèle est sans doute la dimension principale des dynamiques spatiales que l'on cherche à reproduire. Rappelons ici que l'on estime, à partir des connaissances d'experts, que les foyers paysans sont très majoritairement dispersés en 800, et concentrés au sein de villages et petites villes en 1100. Le modèle cherche à reproduire cette polarisation, par le biais d'une concentration des foyers paysans, initialement localisés aléatoirement dans l'espace mais n'en parsemant qu'une faible part, en des agrégats de foyers paysans répartis dans une plus large partie de l'espace modélisé.

Mettre un schéma pour rendre compréhensible cette contradiction apparente.

Pour analyser la polarisation du système de peuplement, il est nécessaire de définir des indices permettant de caractériser ce phénomène. Ces indices doivent d'une part avoir une logique thématique, c'est-à-dire être appropriés à la description et à l'étude de la polarisation, mais doivent pouvoir être produits et enregistrés dans le modèle de simulation, formant des indicateurs.

Pour l'étude de la polarisation, il est nécessaire de faire appel à des indices hétérogènes, chacun devant être en mesure de décrire les différents aspects du phénomène de polarisation. En conséquence, on a choisi de faire appel à plusieurs indicateurs qui doivent permettre d'étudier aussi bien l'aspect structurel du système simulé en son état final que la forme et la tendance que prennent les changements qu'il subit.

L'indicateur principal est le taux de dispersion des foyers paysans. Si celui-ci est trop important (c'est-à-dire très supérieur aux valeurs estimées empiriquement), cela signifie que la polarisation générée par le modèle est insuffisante, et dès lors, obligatoirement insatisfaisante. A contrario, une valeur trop faible serait symptomatique d'un emballement des mécanismes simulés, figeant la situation dans une concentration absolue des foyers paysans, ne laissant dès lors plus de place à la diversification des situations locales et de la hiérarchisation d'ensemble.

Pour affiner ce constat, on fait appel à d'autres indicateurs : le nombre d'agrégats, de pôles, ou encore la dispersion spatiale de ces deux types d'entités. Ces indicateurs ne permettent pas, à eux seuls, de caractériser le succès de la dynamique de polarisation modélisée, mais ils aident à affiner l'analyse de cette dynamique telle que produite par le modèle de simulation. Ils éclairent ainsi le phénomène de polarisation sous des angles légèrement différents, ayant plus pour objet de diagnostiquer les problèmes potentiels qui mèneraient à une mauvaise polarisation plutôt que de qualifier celle-ci. Par exemple, la dispersion des agrégats et pôles peut renseigner, une fois le taux de foyers paysans dispersé jugé trop important, sur une des raisons probables de ce résultat non satisfaisant. Il s'agit donc d'indicateurs secondaires, permettant de préciser la

capacité du modèle à reproduire les faits stylisés, alors que l'évaluation de cette capacité est surtout le rôle de l'indicateur principal.

### 3.3.1.1 Taux de foyers paysans dispersés

Cet indicateur, et sa déclinaison temporelle, sont vraisemblablement les plus évidents : plus le taux de foyers paysans dispersés en fin de simulation est faible, plus le système de peuplement est polarisé. On a vu<sup>29</sup> qu'empiriquement, autour de 1160, on observe environ 20% de foyers paysans encore dispersés, alors qu'ils sont près de 95% au départ. Le modèle sera donc d'autant plus satisfaisant que cet indicateur s'approchera de 20% en fin de simulation.

La « déclinaison temporelle » mentionnée juste au dessus permet d'affiner légèrement la précision de l'information communiquée par cet indicateur : il faut certes atteindre un objectif quantifié (20%), mais les hypothèses empiriques permettent aussi de penser qu'il faut que l'évolution de cet indicateur au cours du temps présente une tendance stable à la diminution, diminuant ainsi plus ou moins, avec de faibles fluctuations à chaque pas de temps. Une configuration de paramètres présentant des valeurs d'indicateurs en sortie proches des objectifs en fin de simulation mais fluctuant fortement pendant son déroulement ne sera ainsi pas valide. Elle montrerait en effet un phénomène trop aléatoire au cours du temps, dont on ne peut penser qu'il s'est déroulé historiquement selon les dires d'experts qui considèrent que cette évolution s'est déroulée de manière continue dans le temps.

Dans la version 0 de SimFeodal, on atteint en moyenne (moyenne des répliques) 57% de foyers paysans isolés en fin de simulation. C'est un taux bien plus important que les 20% attendus, illustrant dès lors une polarisation trop faible. Dans la première moitié de la période (jusqu'en 1000, (figure 3.8)), le taux est en diminution constante et linéaire, ce qui est une bonne tendance, et ne présente de plus presque pas de variabilité, ce qui l'est aussi. Pourtant, à partir de 1000, la variabilité augmente, et la diminution du taux au cours du temps s'interrompt, avant de prendre une allure croissante. L'allure de la courbe sur la seconde moitié de la simulation est donc très éloignée de ce qui est attendu au regard des connaissances empiriques, aussi bien en matière de tendance qu'à la vue du taux décevant atteint au final.

29. Ne pas oublier de présenter les objectifs dans chapitre 2. Ou alors, on les présente ici au fur et à mesure qu'on mentionne les indicateurs

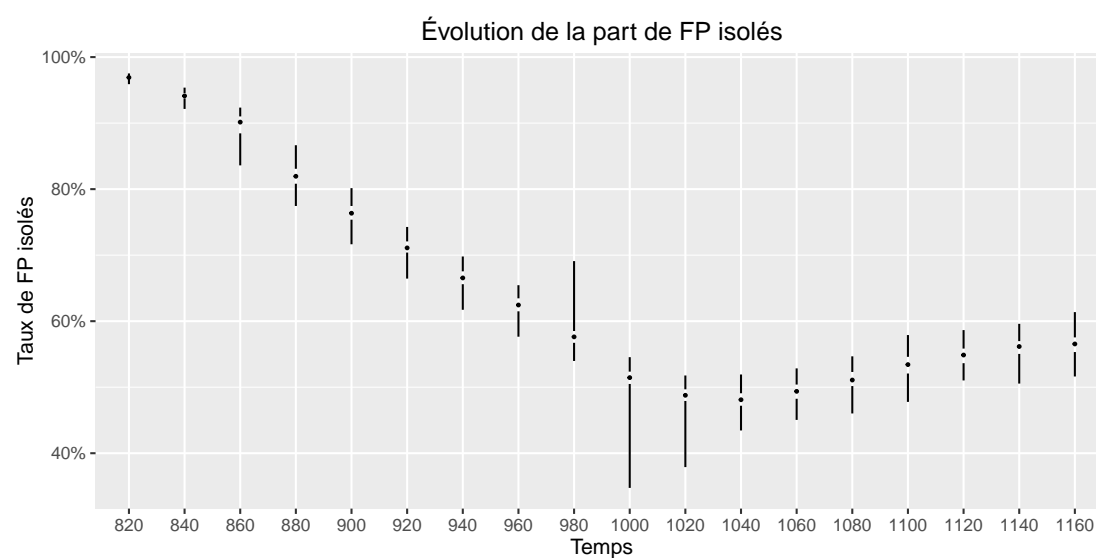


FIGURE 3.8 – Évolution de la part des foyers paysans isolés.

### 3.3.1.2 Nombre d'agrégats

Puisque les foyers paysans se concentrent au sein d'agrégats, il est logique d'observer l'évolution de ces derniers. Là aussi, (**objectif à définir**), on peut considérer qu'un nombre d'agrégats en fin de simulation proche de l'objectif, 200, permet de caractériser une polarisation réussie. Cet indicateur ne peut être lu seul, et c'est pour cela qu'il vient dans un second temps (cf. 3.2.2.3, Hiérarchiser les indicateurs dans chaque dimension, p.31) : en effet, un faible nombre d'agrégats peut aussi bien être révélateur d'une très faible polarisation des foyers paysans (ceux-ci restant dispersés) que d'une trop importante (un unique agrégat concentrant l'ensemble des foyers paysans par exemple). Une fois le taux de foyers paysans dispersés connu et ces potentiels biais pris en compte, le nombre d'agrégats et son évolution nous renseigne cependant sur les dynamiques de polarisation. D'après les connaissances expertes, on s'attend à ce que le nombre d'agrégats, très faible au départ (24 dans la version 0), suive trois phases : une première phase de croissance lente, le temps que les mécanismes agissent sur la polarisation, suivie d'une période de croissance plus rapide, une fois que tous les foyers paysans commenceront à être suffisamment attirés par les pôles pour y former des agrégats, et enfin, une nouvelle phase de croissance plus lente, une fois les foyers paysans répartis dans les agrégats existants et qui se déplaceront vers des agrégats plus importants, hiérarchisant le système de peuplement. Cette allure d'évolution rappelle les fonctions logistiques connues par exemple pour les cycles de diffusion/adoption des innovations **ref Pumain ou MN Comin**, et résulte des connaissances expertes des archéologues spécialistes de la période.

Le nombre d'agrégats est assez satisfaisant dans la version 0 du modèle. Il s'élève ainsi à **187** en moyenne, cette dernière étant d'ailleurs stable au regard des répliques. L'écart à l'objectif (200) est donc assez minime. On a toutefois vu que le taux de foyers paysans isolé était bien trop important, et dès lors, ces agrégats sont logiquement composés de trop peu de foyers paysans.

L'observation de cet indicateur au cours du temps (figure 3.9) est quant à elle assez satisfaisante : on retrouve bien les trois phases attendues, quand bien même le début de la croissance plus importante (vers 1000 ici) est un peu trop tardive. Ce moment coïncide de plus avec la stagnation puis l'inversion de la courbe d'évolution de la part de foyers paysans isolés.

On peut quand même considérer que si la polarisation n'est pas assez importante, les structures qui en résultent semblent correspondre à l'empirie.

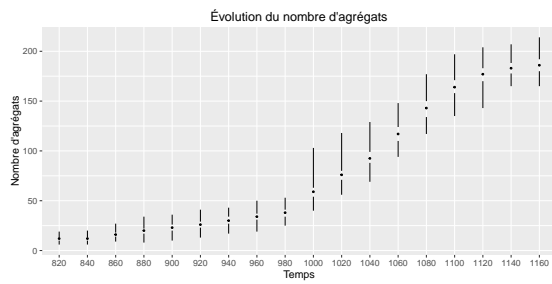


FIGURE 3.9 – Évolution du nombre d'agrégats.

Il y a un problème ici par rapport à ce qui a été mis dans le chapitre : dans le chapitre TMD, on mentionne 83 agrégats en moyenne, alors que dans l'analyse de Base, il y en a 187...

### 3.3.1.3 Nombre de pôles

Dans le modèle SimFeodal, les foyers paysans sont polarisés par des pôles d'attraction. Pour une polarisation efficace, il est donc nécessaire que les pôles soient suffisamment nombreux, c'est-à-dire, *a minima*, autant que d'agréats attendus (200). Contrairement aux agrégats, un nombre trop important de pôles ne constitue pas un problème : en considérant que 20% des foyers paysans doivent demeurer isolés en fin de simulation, il est vraisemblable qu'une partie des pôles, par exemple composés d'une église paroissiale, n'aient pas vocation à voir la constitution d'un agrégat autour d'eux.

Par ailleurs, afin de renforcer la polarisation par l'action de l'attraction différenciée – selon le nombre de foyers paysans contenus dans chaque agrégat<sup>30</sup> –, il faut que le taux de pôles contenant un agrégat soit important, et surtout croissant au cours du temps : comme dans l'empirie, cela est alors le marqueur que de petits pôles d'attractions, comme les églises paroissiales, parviennent à polariser suffisamment de foyers paysans de leur voisinage pour aboutir à la création d'un petit agrégat, un village par exemple.

Pour un résultat satisfaisant, il faut donc que le nombre de pôles augmente régulièrement au cours de la durée de la simulation, et que le taux de pôles contenant un agrégat augmente lui aussi de manière continue.

L'évolution du nombre de pôles est assez représentative de l'empirie dans la version 0 de SimFeodal. On aboutit ainsi sur **190** pôles en moyenne en fin de simulation, ce qui est dans le bon ordre de grandeur, quoi qu'un peu trop faible. Ce nombre est en croissance constante à partir de 940 (figure 3.10). Ce départ quelque peu tardif peut expliquer le relatif manque de pôles en fin de simulation.

L'analyse de la constitution de ces pôles en matière d'agréat est toutefois plus décevante : la croissance n'est pas linéaire, stagnant entre 940 et 1000, puis à partir de 1040. On constate ici un mauvais comportement du modèle par rapport à l'empirie : il y a création de nombreux pôles, mais ceux-ci ne suffisent pas à polariser les foyers paysans. On peut donc penser qu'il s'agit de nombreux pôles ruraux, par exemple composés d'une unique église paroissiale, ne suffisant dès lors pas à concentrer les foyers paysans alentours.

Dans cette dimension, comme pour l'indicateur principal qu'en est le taux de foyers paysans dispersé, l'effet de la polarisation apparaît nettement trop faible.

30. L'attraction qu'exercent les pôles contenant des agrégats de population est ainsi plus forte lorsque ces derniers sont constitués d'un nombre de foyers paysans important, et moindre quand ce nombre est faible. Cela place ce mécanisme dans une logique proche de celle de l'attachement préférentiel identifié par YULE 1925 et SIMON 1955, cités par schmittModélisationDynamiqueSystemes2014)

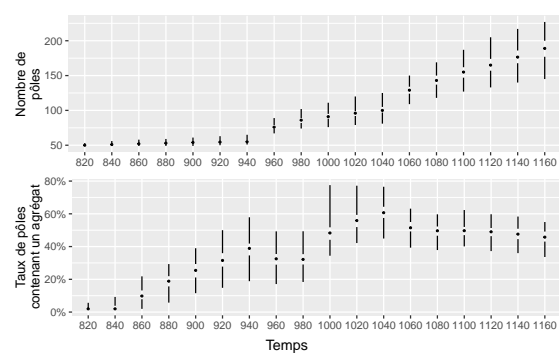


FIGURE 3.10 – Évolution du nombre de pôles et du taux de pôles contenant un agrégat.

#### 3.3.1.4 Dispersion des agrégats et pôles

Les indicateurs présentés ci-dessus avaient en commun d'être des indicateurs de sortie, générés automatiquement, et présentés sous forme d'agrégations des répliques. Pour préciser l'analyse de la polarisation, il est toutefois nécessaire d'aborder une dimension qui n'est fondamentalement pas agrégable dans les indicateurs de sortie : l'espace du modèle. Comme celui-ci est théorique et aléatoire, il n'y a aucun sens à agréger des entités différentes, par exemple des agrégats, sachant que ceux-ci occupent des localisations aléatoires et ne sont pas identifiables en tant que tels.

On ne peut cependant pas se passer d'une analyse spatiale de la répartition des pôles et agrégats afin de comprendre les dynamiques effectivement simulées par le modèle. La distribution spatiale des agrégats et des pôles est en effet un facteur majeur de la polarisation : s'ils sont très concentrés, les foyers paysans non présents alentours ne trouveront pas d'attracteurs à proximité, et ne seront de plus pas particulièrement affectés par l'augmentation des droits (banaux etc.) et des contraintes spatiales (proximité à une église, à un château etc.). A l'inverse, des agrégats entièrement dispersés ne favoriseraient pas la structure spatiale hiérarchisée que l'on cherche à faire émerger.

Afin que le comportement du modèle soit satisfaisant, il faut donc que les pôles et agrégats occupent l'ensemble de l'espace du modèle, tout en présentant des zones de concentration relatives plus importantes. Comme on ne peut agréger les représentations spatiales, il convient, pour cette analyse, de regarder individuellement un échantillon de configurations spatiales générées.

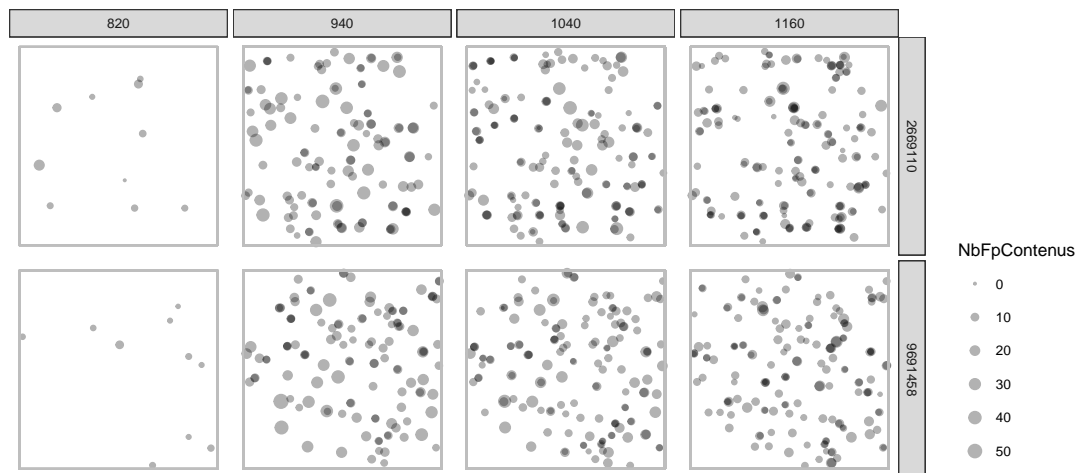
La lecture des cartes de répartition des agrégats et des pôles de deux répliques de la version 0 (figure 3.11) va, dans l'ensemble, dans le sens de l'empirie. On peut ainsi remarquer que ces entités ont bien tendance, au cours du temps, à se disperser dans l'espace modélisé. En fin de simulation, tout l'espace est occupé, et on remarque même que certaines zones voient une forte concentration en agrégats, reproduisant les faits stylisés connus. La répartition spatiale des agrégats – et des pôles – est donc bonne, et ne semble pas opposer d'obstacle à la polarisation attendue du système de peuplement.

Ce résultat peut cependant être nuancé par l'observation de la dynamique de cette dispersion : on constate ainsi que la dispersion s'effectue rapidement (entre 800 et 940), et n'évolue plus vraiment après. Ce phénomène est donc plus rapide dans cette version 0 du modèle que dans les connaissances empiriques de la région Touraine.

On peut enfin remarquer, en préalable aux dynamiques de hiérarchisation du système que l'on s'apprête à étudier, que ces sorties illustrent un manque criant de hiérarchie quant à la composition des agrégats : on ne remarque, visuellement, que peu de différence entre les agrégats, et surtout, cette hiérarchie semble faiblir entre 1040 et 1160 (les agrégats les plus importants ont vu leur nombre de foyers paysans diminuer).



Composition des agrégats



Pôles et attractivité

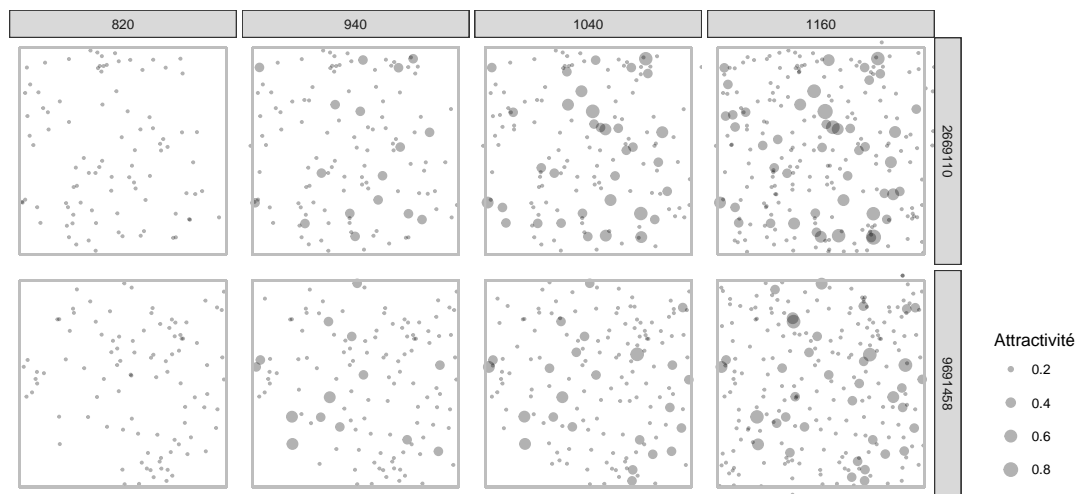


FIGURE 3.11 – Évolution de la répartition spatiale des agrégats et pôles, pour deux répliques.

### 3.3.2 Évaluer la hiérarchisation du système de peuplement

La seconde dimension de l'évaluation du modèle SimFeodal correspond à l'étude de la hiérarchisation du système de peuplement. On déduit en effet des connaissances empiriques une forte hiérarchisation du système de peuplement sur la période, et plus généralement, des entités présentes. On passe ainsi, en 800, d'un habitat dispersé dans lequel coexistent quelques agrégats de taille uniforme, à un habitat concentré dans des agrégats de taille très hétérogènes à la fin du XII<sup>ème</sup> siècle. La distribution des tailles des agrégats est estimée par les connaissances expertes, toutes proportions gardées, comme assez proche des distributions observées aujourd'hui dans les systèmes de peuplement. On souhaite ainsi que les agrégats modélisés suivent une distribution approchant la distribution log-normale.

Par extension, et là encore d'après les connaissances thématiques, l'ensemble des entités doit aussi suivre le même type de forme. Par exemple, les pôles, tant en terme d'attractivité que de composition, doivent aussi montrer une hiérarchie du même ordre, ainsi que les seigneurs – à travers leur puissance, au moins pour les petits seigneurs –, ou encore les paroisses, par le nombre de paroissiens qu'elles desservent.

Comme pour l'étude de la polarisation, on peut définir un indicateur principal de cette hiérarchisation du système de peuplement : la forme de la distribution de la composition en foyers paysans des agrégats.

De la même manière que pour la polarisation, les indicateurs secondaires ont aussi pour but de préciser cet indicateur principal, et en particulier d'analyser les moteurs de cette hiérarchisation du peuplement. On a en effet choisi d'observer plutôt la hiérarchisation des autres types d'entités – pôles, seigneurs, paroisses –, pour vérifier qu'elles accompagnent et/ou entraînent bien la hiérarchisation des agrégats. La hiérarchie des pôles, par exemple, a une influence directe sur l'attraction effectuée sur les foyers paysans (polarisation) et sur la hiérarchisation des agrégats : par effet d'attraction différenciée (voir la note de bas de page 30 page 38), des agrégats plus importants se constituent autour des pôles les plus importants.

Comme pour la polarisation, l'analyse de la capacité du modèle à reproduire la hiérarchisation du système de peuplement se fait donc en deux temps : en premier lieu, on évalue cette capacité à l'aide de l'indicateur principal, puis on précise cette qualification et on essaie de l'expliquer à l'aide des indicateurs secondaires.

#### 3.3.2.1 Hiérarchie des agrégats

L'indicateur principal est un indicateur agrégé, correspondant à la forme de la distribution des agrégats mesurés par le nombre de foyers paysans qui les composent. Cet indicateur est classique dans l'analyse des systèmes de peuplement, et il est courant de l'observer par le biais d'un indicateur agrégé simple, correspondant à la loi rang-taille. On observe pour cela le modèle statistique, ou sa représentation graphique tout du moins, mettant en relation le

logarithme de la taille des individus (le nombre de foyers paysans composant chaque agrégat ici) et le logarithme du rang de cet individu. Comme pour toute régression linéaire, on peut alors quantifier l'ajustement du modèle grâce au coefficient de détermination ( $R^2$ ), et spécifier la pente de la courbe, représentant le degré de hiérarchie, à travers le coefficient directeur ( $a$  dans la formule  $y = ax + b$ ).

Dans le cas de SimFeodal, le faible nombre d'agrégats ainsi que la variabilité de leurs tailles rend difficile cette analyse quantifiée, le coefficient directeur, par exemple, étant très sensible aux faibles effectifs. On utilise toutefois la représentation graphique décrite comme un indicateur majeur de la hiérarchie des agrégats.

Du point de vue des connaissances empiriques, la courbe doit ainsi voir sa pente augmenter avec le temps, tout en devenant plus convexe, ce qui représente la « longue traîne » des petits agrégats, empiriquement observée dans toutes les distributions de systèmes de peuplement.

Avec une autre représentation graphique du même phénomène, en discrétisant les agrégats selon leur taille et en dénombrant le nombre d'agrégats de chaque classe, on peut aussi avoir une vision plus synthétique (car moins exhaustive) de la forme de la distribution. On doit alors obtenir un nombre décroissant d'agrégats à mesure que la classe représente un nombre élevé de foyers paysans.

Ça vaudrait peut-être le coup, pour chaque indicateur présenté sous forme graphique (évolution ou forme), de faire un graphique théorique (une sorte de courbe parfaite) de ce que l'on souhaite observer.

La version 0 de SimFeodal présente une hiérarchisation des agrégats nettement trop faible. La courbe rang-taille (figure 3.12) est ainsi trop faiblement pentue, et présente une forme trop linéaire : la convexité due à la longue traîne n'est pas assez visible, en raison sans doute de trop faibles valeurs en haut de la hiérarchie, qui ne « tire » alors pas assez la distribution. Ce résultat de simulation est d'autant plus perturbant que son évolution est éloignée de l'empirie : on remarque en effet que la distribution se hiérarchise bien entre 800 et 1020, présentant à cette date une allure très satisfaisante. Pourtant, après cette période, les agrégats voient leur hiérarchie diminuer nettement et les agrégats les plus peuplés diminuer en taille.

On constate le même décrochage dans la discrétisation des agrégats (figure 3.13), où on remarque de plus que la hiérarchie la plus proche de l'attendu, où la position des classes suivrait une ligne droite, semble se dessiner entre 940 et 1040. En 1040, et plus encore en 1160, la proportion d'agrégats de taille moyenne et haute (de 30 à 100, et de plus de 100 foyers paysans) est trop faible, et pas assez hiérarchisée.

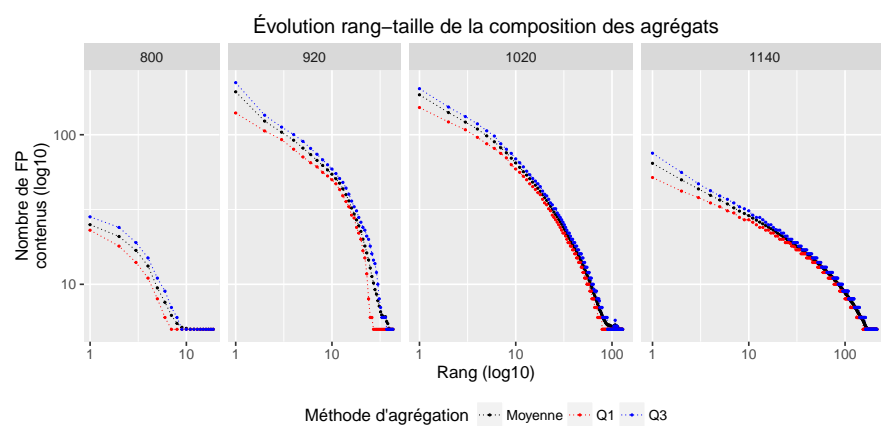


FIGURE 3.12 – Évolution de la courbe rang-taille des agrégats.

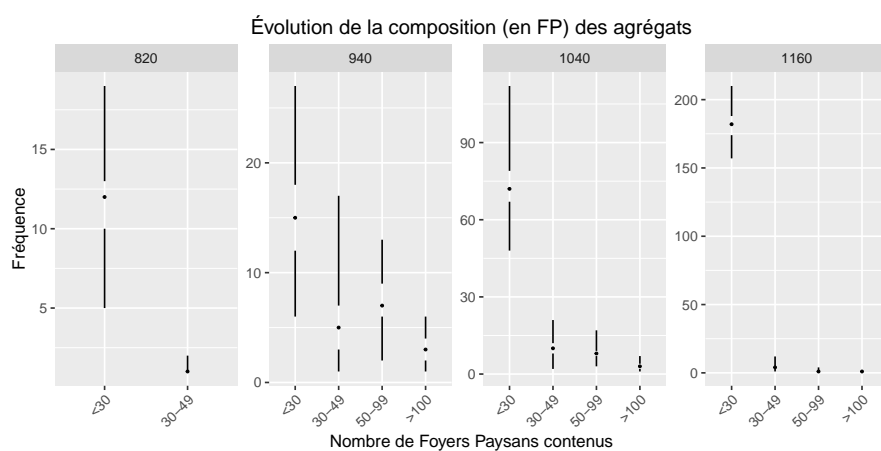


FIGURE 3.13 – Évolution de la composition des agrégats.

### 3.3.2.2 Hiérarchie des pôles

La hiérarchie des agrégats donne une bonne vision agrégée de la hiérarchisation du système de peuplement dans son ensemble. Pour autant, afin d'appréhender la dynamique de cette dimension, il est là encore nécessaire d'observer le comportement des composantes qui provoquent cette hiérarchisation. En effet, une forte hiérarchie des pôles entraînera une attractivité des foyers paysans très inégale<sup>31</sup>. De plus, comme indiqué plus haut, on cherche à obtenir une forte hiérarchie pour les différents types d'agents du modèle. L'observation de la hiérarchie des pôles est donc importante pour évaluer le modèle SimFeodal. Pour déterminer cette hiérarchie, on peut se fier à deux indicateurs complémentaires : le nombre d'attracteurs composant chaque pôle et l'attractivité de ces derniers. Ces indicateurs de sortie sont proches, mais apportent pourtant une vision légèrement différente : étant donné que chaque attracteur influe différemment, selon son type, sur l'attractivité globale d'un pôle, l'information sur l'attractivité et sur la composition ne sont pas redondantes, bien que fortement corrélées. On aurait pu présenter une information plus détaillée quant à cette composition, par exemple en différenciant le nombre de chacun des types d'attracteurs de chaque pôle, mais le nombre de combinaisons possible aurait rendu cette information confuse.

À partir des connaissances expertes qui guident l'évaluation de SimFeodal, on cherche à obtenir, pour ces deux indicateurs, une courbe d'allure similaire, c'est-à-dire une courbe décroissante, avec bien plus de pôles mineurs (faible attractivité ou nombre d'attracteurs) que de pôles plus importants. On cherche de plus à ce que cette courbe présente une allure log-normale, et donc que la proportion de pôles décroisse fortement à mesure que leur importance augmente.

Dans l'ensemble, on peut remarquer sur les deux indicateurs de la figure 3.14 que les pôles, dans cette version 0, sont très hiérarchisés. La tendance « évolutive va dans le bon sens : depuis une quasi-uniformité en 820, des pôles plus importants apparaissent et semblent se renforcer au cours du temps. On retrouve aussi, en observant les axes des ordonnées, la croissance du nombre de pôles identifiée plus haut (3.3.1.3, Nombre de pôles, p. 38) : de nouveaux pôles apparaissent tout au long de la simulation, et restent majoritairement peu importants, quand les pôles ayant commencé à se renforcer tôt continuent dans cette tendance.

En fin de période, les pôles sont très hiérarchisés, aussi bien en matière de composition que d'attractivité. Ils le sont même plus que ce que les connaissances expertes laissent entendre : au delà de 0.66 en attractivité ou de 4 attracteurs, on ne dénombre plus qu'un unique pôle de chaque importance, là où on attendrait que cette courbe continue plutôt à décroître. De plus, ces courbes montrent la variation des répliques de la version 0. Dès lors, on peut considérer que ces pôles majeurs ne sont en fait qu'un unique pôle (le plus souvent) d'importance variable selon les répliques,

31. En raison des logiques d'attachement préférentiel, voir la note de bas de page 30 page 38.

mais dans tous les cas d'importance bien supérieure aux pôles qui arrivent juste après dans le classement. Cette macrocéphalie ne correspond pas aux observations empiriques, d'autant que rappelons-le, la ville de Tours n'est pas modélisée dans SimFeodal.

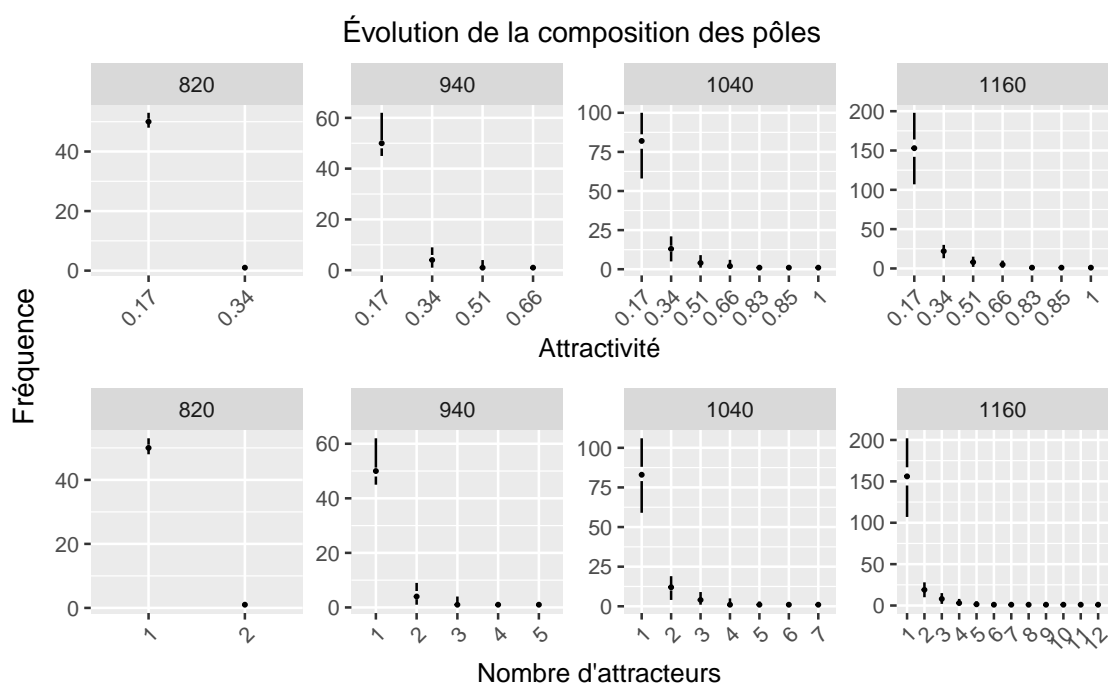


FIGURE 3.14 – Évolution de la composition et de l'attractivité des pôles<sup>32</sup>.

32. Les figures ne sont pas en log-log (ou log-lin) et on ne voit donc pas vraiment le haut de la hiérarchie. Peut-être faudrait-il changer ces axes, mais ça complexifiera la lecture pour les archéos/lecteurs non habitués aux axes log

### 3.3.2.3 Hiérarchie des paroisses

A l'instar des pôles, on attend aussi des ressorts paroissiaux d'être hiérarchisés. La période modélisée voit ainsi apparaître ces paroisses qui auront un rôle majeur dans la fixation du peuplement, et, pré-figurant le maillage communal, ont un double rôle de desserte efficace<sup>33</sup> et équitable<sup>34</sup>. En effet, avec la volonté d'encadrement et de prélèvement de l'Église sur la population, de nouvelles paroisses apparaissent pour desservir au mieux leurs potentielles ouailles. Une structure double en résulte : dans les zones les moins denses, le maillage est régulier mais lâche, de manière à minimiser le nombre d'églises paroissiales tout en s'assurant que chacun puisse y accéder dans un temps raisonnable<sup>35</sup>. Il y a donc un certain nombre d'églises paroissiales desservant peu de paroissiens. Au contraire, dans les zones les plus denses, et en particulier au sein des petites villes naissantes, l'objectif est d'être au plus près des résidents tout en garantissant à chacun de pouvoir assister aux différents offices : il y donc une croissance du nombre d'églises paroissiales proches les unes des autres, visant à accompagner un encadrement maximum de la population, ainsi, avec une logique concurrentielle de ce clergé féodal, qu'à capter l'importante source de revenus qu'assure la collecte de la dîme. Cette logique concurrentielle doit aussi permettre de restreindre le nombre de foyers paysans desservis par une unique paroisse.

On s'attend donc à avoir une courbe hiérarchisée dans la lignée d'une courbe log-normale, mais avec toutefois un double seuil minimal (l'effet de « longue traîne »), autour des paroisses « rurales » peu peuplées (moins de 10 foyers paysans desservis) et des paroisses du bas de la hiérarchie classique, peuplées de 10 à 40 foyers paysans<sup>36</sup>).

ici, impérativement, il faudra mettre un graphique schématique de ce qu'on attend.

En fin de simulation de cette version 0, la hiérarchie des paroisses est plutôt satisfaisante. On y retrouve en effet une forte hiérarchie, peut-être même légèrement trop importante. Il ne devrait ainsi par y avoir de paroisse composée de plus de 100 foyers paysans en fin de simulation, et la courbe présente une pente trop importante (celles de 940 ou 1040 sont ainsi plus conformes aux attentes). Pour autant, on remarque bien un effet de hiérarchisation au cours du temps, ainsi qu'un accroissement du nombre de petites paroisses, montrant que la dynamique, bien que non ajustée, s'inscrit bien dans la dynamique observée empiriquement. (figure 3.15)

33. C'est-à-dire desservir de manière optimale la plupart de la population.

34. C'est-à-dire faire en sorte que même les populations les plus isolées aient un accès aussi rapide que possible à une église paroissiale.

35. Cette distance-temps évolue au cours du temps, en fonction de l'accroissement de la fréquence de l'obligation de fréquentation des églises paroissiales.

36. D'après les valeurs des paramètres `nb_min_paroissiens` et `nb_max_paroissiens`

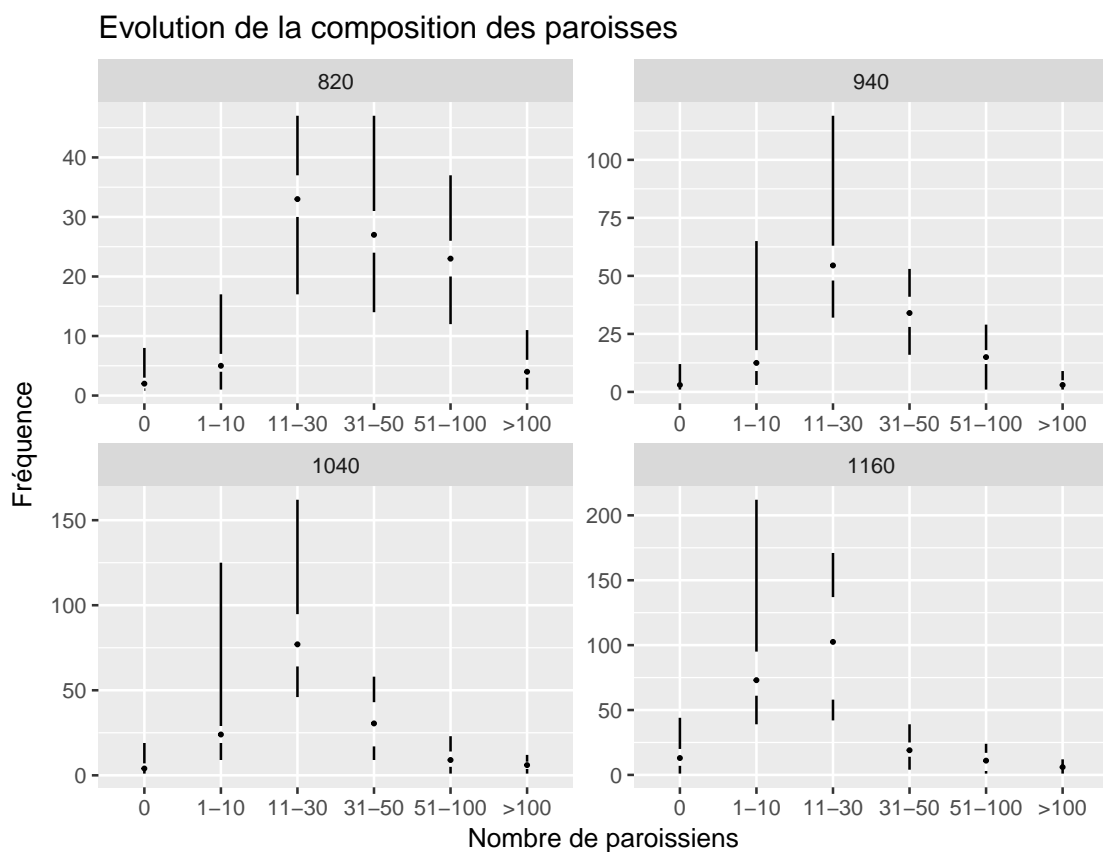


FIGURE 3.15 – Évolution de la composition des paroisses.



### 3.3.3 Évaluer la fixation et la dissémination du peuplement

La dernière dimension étudiée par SimFeodal est moins définie que les précédentes. On y observe ainsi la fixation du système de peuplement au sein du maillage naissant que constituent les paroisses. Cette notion de fixation pose problème par rapport à l'ensemble des indicateurs de sortie déployés jusqu'ici. En effet, SimFeodal est un modèle fondé sur un temps discret : on y observe, à chaque itération, le résultat des processus modélisés. On peut y observer une fixation d'un pas de temps par rapport à l'autre, par exemple en constatant qu'un agrégat constitué en 820 semble être toujours présent en 840. On peut de plus considérer que cet agrégat présente les mêmes caractéristiques au deux dates, en matière de localisation et de population. L'agrégat est donc stable dans le temps. L'échelle d'analyse mobilisée dans les indicateurs précédents – au niveau agrégé des agrégats, voir du système d'agrégats – pourrait donc s'appliquer ici aussi. Pourtant, quand on parle de fixation des foyers paysans, ce qui est observé empiriquement, il n'est pas question d'agrégats, mais des entités les composant, c'est-à-dire les foyers paysans. Ceux-ci ne sont pas observés directement quand on remarque que les agrégats semblent stable : on approche là à la différence entre un état stable et un état stationnaire. La relative stationnarité des agrégats dans le temps n'est ainsi pas garante de stabilité de leurs composantes : deux foyers paysans oscillant d'un agrégat à un autre, dans un mouvement opposé, produiraient ainsi une stationnarité de ces agrégats, mais, en se déplaçant, ils ne satisferaient pas au critère de fixation. Il y a donc un problème de changement d'échelle pour observer la fixation des foyers paysans : il n'est plus possible de raisonner à l'échelle des agrégats, et il faut se concentrer sur celle des foyers paysans.

Cette échelle pose un autre problème : les foyers paysans sont très nombreux (4000 dans la version 0) et se déplacent. Ils se déplacent de plus selon des modalités très différentes (**Faire ref aux mécanismes de déplacement du chapitre 2**), rendant complexe la caractérisation des mouvements de chacun et plus encore celle d'une aggrégation de ces catégories.

Pour ces raisons, la production d'indicateurs synthétiques spatiaux – une ou plusieurs cartes – ne suffirait pas à communiquer une information intelligible sur l'éventuelle fixation des foyers paysans. Il nous faut donc faire appel à des « proxys », non spatiaux, pour évaluer la fixation des foyers paysans.

A cet effet, on a retenu des indicateurs relatifs aux déplacements des foyers paysans et à leur raison : combien de foyers paysans se déplacent à chaque pas de temps (moins il y a de déplacement, plus la fixation est importante) ? Quelles sont les modalités de ces déplacements (un déplacement entre deux agrégats lointains n'a pas les mêmes conséquences en terme de stabilité qu'un déplacement minime au sein d'un même agrégat) ? Ou encore, comment évolue la satisfaction des foyers paysans, et avec celle-ci, la probabilité de se déplacer ?

Ces indicateurs permettent d'évaluer la capacité du modèle à reproduire la fixation des foyers paysans. Pour autant, la contrainte est ici double : on recherche une fixation, mais celle-ci est, empiriquement, supposée se dérouler et se voir renforcer par la mise en place du maillage paroissiale qui doit servir

de support à la nouvelle configuration spatiale émergente.

On s'appuiera donc aussi sur des indicateurs relatifs à cet espace support constitué par les paroisses : leur nombre, leur dispersion dans l'espace et l'efficacité de la desserte qu'elles assurent.

Notons que contrairement aux deux précédentes dimensions d'analyse, nous n'établissons ici pas de hiérarchie nette entre les indicateurs. Cette étude de la fixation est moins facilement appréhendable que celle de la polarisation ou de la hiérarchisation et les indicateurs qui la caractérisent apportent une complémentarité de points de vue plus qu'un affinement de l'évaluation de cette dynamique. Dès lors, les indicateurs présentés ci-après ne peuvent être catégorisés en indicateurs principaux et secondaires. On retrouvera cependant cette hiérarchie d'évaluation au sein des indicateurs, par exemple en suivant l'ordre des graphiques présentés. Par exemple, pour évaluer la fixation des foyers paysans, on observera d'abord le résultat produit (nombre de déplacements) avant d'entrer dans le détail de sa composition (types des déplacements).

### 3.3.3.1 Déplacement des foyers paysans

Le déplacement des foyers paysans est l'élément moteur de SimFeodal : c'est par le déplacement individuel de chacun des foyers paysans que la configuration spatiale évolue. Les déplacements affectent donc chacune des dimensions d'analyse – polarisation, hiérarchisation et fixation –, mais c'est au sein de cette dernière qu'il est le plus intéressant de les observer. Il est en effet attendu que de nombreux déplacements surviennent, afin que le système de peuplement puisse se structurer, mais pour autant, il est aussi nécessaire que ces déplacements tendent à diminuer au cours du temps, une fois le système en voie de stabilisation.

On pourrait donc attendre que les déplacements suivent une courbe négative (linéaire ou non) tendant vers 0, impliquant une absence de déplacements en fin de période. Pour autant, le mécanisme de déplacement est sans doute l'un des plus complexes du modèle, et on ne peut l'appréhender aussi simplement.

En premier lieu, les mécanismes de SimFeodal différencient deux types de déplacements (**Faire ref à chap2**) : les déplacements locaux (dans un rayon de 2500m dans la version 0) et les déplacements lointains.

- Les déplacements locaux visent à faire s'agréger des foyers paysans dispersés autour de pôles présents à proximités. Ce mécanisme peut être présent sur toute la durée de la simulation, sur un effectif faible. En effet, cet effet d'agrégation locale permet « d'optimiser » la répartition spatiale des agrégats, en renforçant leur hiérarchie et en faisant fusionner des agrégats qui seraient très proches les uns des autres. Pour autant, dans un objectif de fixation du peuplement, il est nécessaire de veiller à ce que les foyers paysans ne soient pas amenés à se déplacer localement de manière continue, par exemple en faisant des allers-retours entre des pôles ou agrégats spatialement proches.

- Les déplacements lointains servent un autre rôle : un foyer paysan qui ne serait pas en mesure d'augmenter sa satisfaction localement – faute de pôles suffisamment attractifs dans le voisinage – a une probabilité de se déplacer vers un agrégat situé n'importe où dans l'espace modélisé. Le modèle est très sensible à ce mécanisme qui agit comme une perturbation forte dans la structure spatiale du peuplement. Ce mécanisme « de dernier recours » ne doit être employé que rarement, en cas de situations où des foyers paysans seraient trop isolés pour pouvoir s'agréger localement. C'est notamment le cas pour les foyers paysans nouveaux arrivants, via le mécanisme de renouvellement (ref dans chap 2), qui peuvent se voir localisés n'importe où dans l'espace du modèle.

Pour ajouter à la complexité du mécanisme, et donc de l'évaluation de cet indicateur qu'est le nombre de déplacements au cours du temps, rappelons que différentes contraintes temporelles viennent bouleverser, à dessein, le comportement des foyers paysans. En particulier, entre 950 et 1050, les modalités d'évaluation de la satisfaction deviennent plus strictes (Voir dans chap2, frise). Cela engendre nécessairement une plus forte propension des foyers à se déplacer.

Si les connaissances empiriques d'un tel niveau de finesse ne sont pas disponibles, on peut tout de même avoir des attentes quant au comportement attendu du modèle. Au regard des éléments décrits plus haut, on peut ainsi chercher à ce que l'évolution des déplacements suive plusieurs rythmes au cours du temps :

1. Dans une première phase, du début de la simulation jusqu'aux perturbations débutant en 950 : quelques déplacements lointains marginaux ( $\approx 5\%$  à  $10\%$ ), stables au cours du temps ; et de plus nombreux (au moins  $\approx 30\%$ ) déplacements locaux menant à la constitution de petits agrégats locaux. Les déplacements locaux doivent diminuer au cours du temps, une fois les agrégats constitués.
2. Pendant la deuxième phase, entre 950 et 1050, les nombreuses perturbations devraient voir une nette augmentation des déplacements locaux, et dans une moindre mesure lointains, prémices à la constitution d'agrégats plus hiérarchisés.
3. Après ces perturbations, on devrait retrouver le niveau de déplacement de la seconde période, et là aussi, tendre vers une diminution des déplacements locaux, le système se stabilisant à l'approche de la fin de la période.

Tout au long de cette période, on cherche de plus à ce que les foyers paysans soient polarisés, c'est-à-dire ici, qu'ils se regroupent dans des agrégats de population. Parmi les modalités de déplacement, un autre indicateur utile est ainsi l'observation des provenances et destinations des foyers paysans qui se déplacent : plus les foyers paysans originellement dispersés auront tendance à rejoindre des agrégats, plus la simulation sera satisfaisante au regard des hypothèses empiriquement émises.

Dans cette version 0 de SimFeodal, il apparaît en premier lieu que les déplacements sont trop peu nombreux (figure 3.16). Les déplacements lointains sont à peu près dans les proportions attendues, mais les déplacements locaux sont trop peu nombreux. Surtout, les trois phases attendues ne se retrouvent pas sur ces sorties. On y remarque bien l'impact des perturbations, mais celles-ci ne font qu'augmenter la part de déplacements, laquelle reste stable avant et après ces perturbations. La version 0 de SimFeodal n'est donc pas satisfaisante en termes de stabilisation et de fixation des foyers paysans. L'observation des modalités de déplacement (figure 3.17) montre certes une part croissante de déplacements « Isolé -> Agrégé », mais dans des proportions là aussi trop faibles passé le tout début de la simulation.

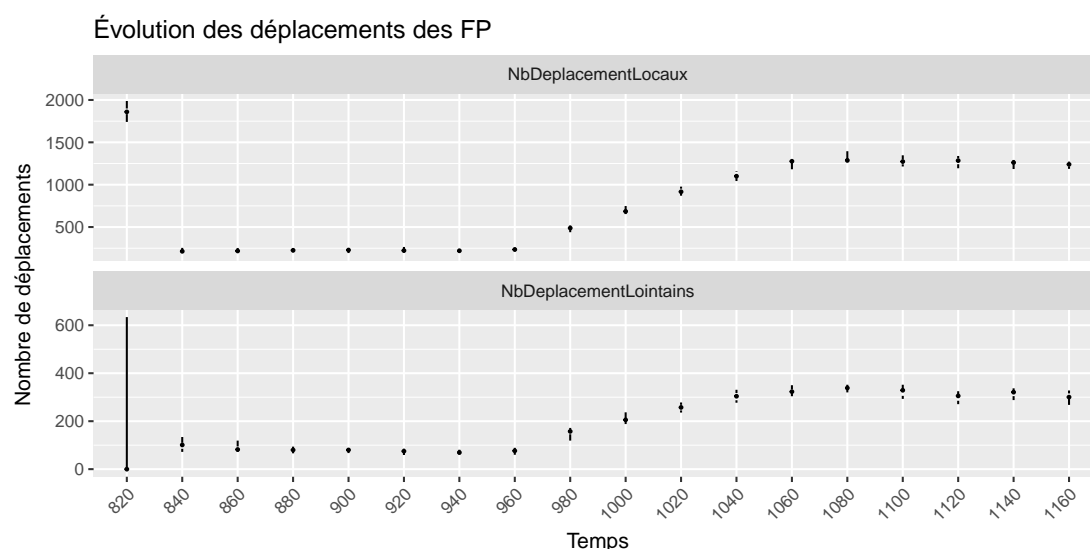


FIGURE 3.16 – Évolution du nombre de déplacement des foyers paysans.

Mettre le graphique en relatif

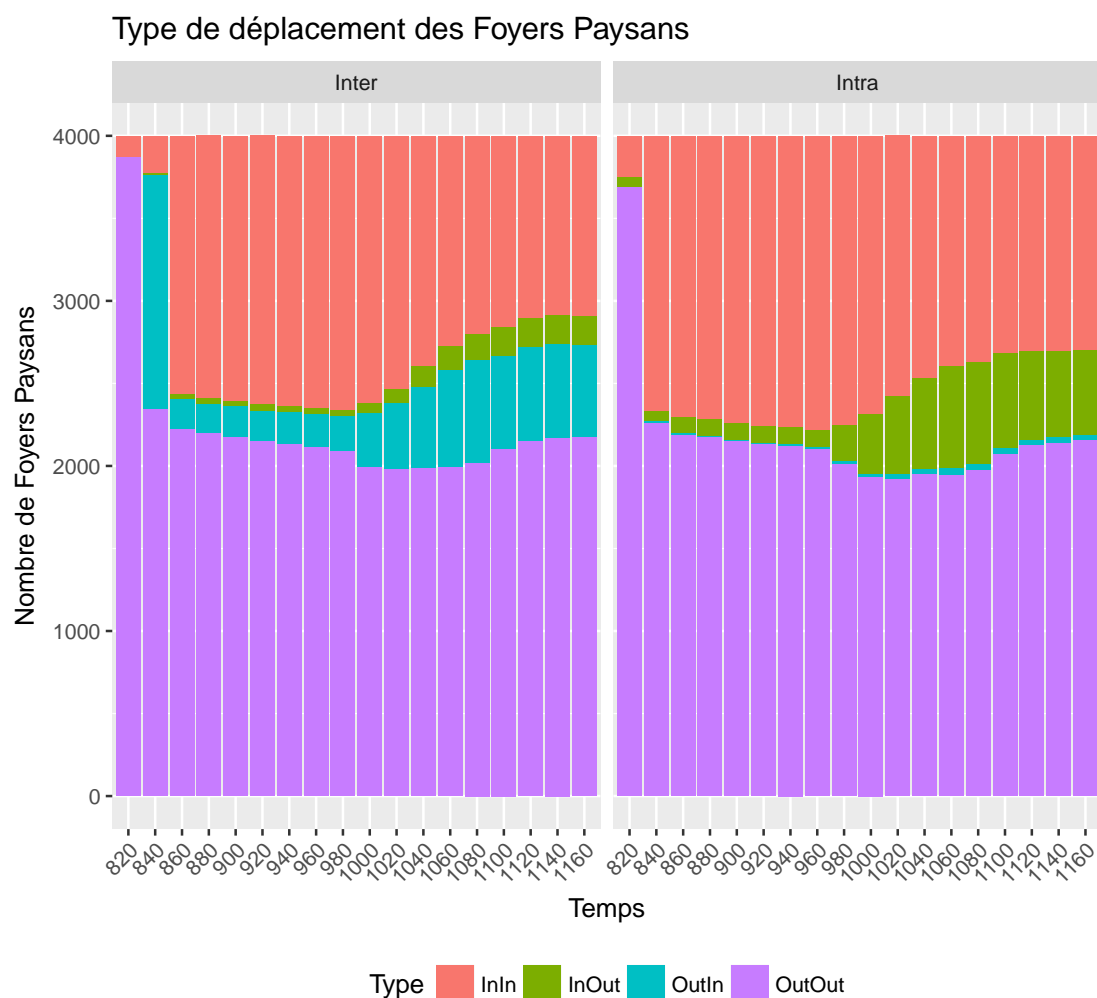


FIGURE 3.17 – Évolution des types de déplacement des foyers paysans.  
 Inter : Déplacement depuis le pas de temps précédent vs Intra : déplacement au sein d'un même pas de temps.

Ne garder que l'inter et mettre le graphique en relatif

### 3.3.3.2 Satisfaction des foyers paysans

On a vu dans les deux indicateurs précédents que les déplacements des foyers paysans sont très affectés par leur niveau de satisfaction. Pour comprendre ces déplacements, il est donc utile d'observer en détail l'évolution de la satisfaction qui les provoquent.

La satisfaction ne saurait être résumée en un simple indicateur de fixation, tant son rôle est prépondérant dans une large partie des mécanismes du modèle. Pourtant, mobilisé ici, cet indicateur apporte un éclairage différent. Il permet ainsi de préciser les indicateurs précédents en donnant une explication à leur éventuelle mauvaise réponse aux attentes.

Ainsi, une satisfaction globalement trop élevée ne serait pas assez motrice à des déplacements, résultant en une polarisation faible. Au contraire, une satisfaction globalement faible engendrerait une très forte mobilité, par exemple sous forme de mouvements pendulaires, d'où une absence de fixation du peuplement. Comme pour les déplacements, on attend, depuis les connaissances empiriques, qu'il y ait trois phases dans l'évolution de cette satisfaction : (1) une première phase, jusqu'en 950, où les foyers paysans sont globalement satisfaits, et le sont de plus en plus à mesure qu'ils s'agrègent; (2) une seconde phase, entre 950 et 1050, où les restrictions fortes (distance à un château, à une église paroissiale) auront pour effet de violemment abaisser le niveau de satisfaction; et enfin, (3) une dernière phase où, passées les perturbations, le niveau de satisfaction tend à remonter doucement, sous l'effet de l'agrégation et de la constitution généralisées de communautés paysannes et de la construction de châteaux et de nouvelles églises paroissiales.

Au vu des résultats de la version 0 de SimFeodal illustrés dans la figure 3.18, on constate que les trois phases attendues sont bien présentes. La perturbation en milieu de période est très forte, mais pourtant, le niveau de satisfaction général reste très élevé (à toute date, plus de 80% des foyers paysans ont une satisfaction supérieure à 0.5). Avec ce niveau de satisfaction et les mécanismes de déplacement de cette version 0, seuls les foyers paysans à proximité de très gros pôles (au moins un gros château et plusieurs églises paroissiales) pourront se déplacer localement. Et pour peu qu'il y ait plusieurs pôles importants à proximité, ils oscilleront entre les deux, gonflant artificiellement le nombre de déplacements locaux. L'observation de cet indicateur confirme que le niveau de déplacement est trop faible et ne présente pas l'allure attendue, tout en donnant une explication à la mauvaise fixation du peuplement observée.

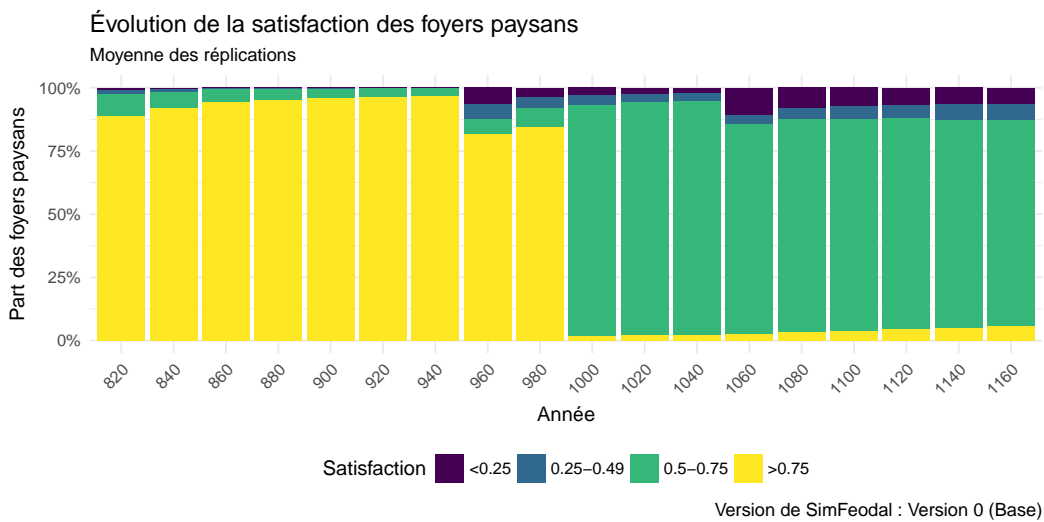


FIGURE 3.18 – Satisfaction des foyers paysans

### 3.3.3.3 Nombre et dispersion des paroisses

Tout au long de la période, de nouvelles églises paroissiales sont créées (cf. mécanisme dans chap. 2) et viennent renforcer l'encadrement des foyers paysans. L'évolution du maillage constitué par les ressorts paroissiaux représente donc l'évolution de la structure spatiale des foyers paysans. Pour que les foyers paysans soient satisfaits, il faut, à partir de 950, qu'ils soient suffisamment proches d'une église paroissiale. Celles-ci constituent donc, à mesure de l'avancement de la période, des pôles qui vont aider à la fixation des foyers paysans. Il est donc légitime d'observer la croissance du nombre et la répartition des paroisses telles que simulées dans le modèle SimFeodal.

Sur un plan purement numéraire, plus les paroisses seront nombreuses, mieux la population sera desservie, et moins les foyers paysans se déplaceront : la fixation sera donc plus forte à mesure que le nombre de paroisses augmente. Sur un plan spatial, l'accumulation de paroisses en zones denses (dans des agrégats de populations) doit renforcer la polarisation de ces zones, et avec elle, accroître les chances de fixation des foyers paysans.

D'après les mécanismes mis en places dans cette version 0 de SimFeodal, on s'attend donc à ce que le nombre de paroisses augmente régulièrement au cours du temps, depuis un nombre initial évalué à 50 en 800, jusqu'à atteindre un objectif numérique fixé à 200 d'après les connaissances empiriques de la région modélisée.

Concernant la répartition spatiale, on cherche à atteindre le double phénomène décrit dans la partie 3.3.2.3 (Hiérarchie des paroisses, p. 47). Spatialement, cela devrait mener à une diminution de la superficie des paroisses les plus larges dans les zones peu denses. Dans les zones plus denses, concentrant les agrégats, cela devrait aussi mener à une diminution de la superficie, bien plus drastique cependant : avec la création de nouvelles paroisses au sein des agrégats, on devrait voir apparaître de nombreuses paroisses se partageant un espace très réduit.

Notons que la dispersion des agrégats et des pôles, vue précédemment (3.3.1.4, Dispersion des agrégats et pôles, p. 40), constituerait ici aussi un bon indicateur de fixation, en observant non plus l'évolution de la couverture spatiale, mais plutôt la fixation et le renforcement des dynamiques locales de polarisation.

Dans la version 0 de SimFeodal, on peut en premier lieu constater que la tendance – à la croissance – de l'évolution du nombre de paroisses est bonne (figure 3.19). Ainsi, de nouvelles églises paroissiales sont bien créées ou promues régulièrement au cours du temps. La quantité atteinte (240 en moyenne) dépasse un peu trop fortement l'objectif empirique, mais surtout, on remarque dans la figure la très forte variabilité de ces résultats (l'intervalle interquartile est de 100). De plus, cette variabilité augmente fortement avec le temps.

La dynamique est donc plutôt satisfaisante, mais le nombre atteint



autant que la variabilité sont améliorables et rendent clairement visible la nécessité d'un ajustement dans le modèle.

Sur le plan spatial, la figure 3.20 laisse bien apparaître la double évolution attendue : les paroisses les plus étendues, en périphérie, demeurent parmi les plus grandes mais voient leur superficie réduite à mesure que de nouvelles églises paroissiales y apparaissent. Dans le même temps, on remarque aussi les effets de fractionnement et de subdivision des paroisses les moins étendues. Cela résulte, dans la figure, en de nombreuses concentrations locales de paroisses que l'échelle des cartes nous permet plus de percevoir que de détailler précisément<sup>37</sup>.

Là aussi, courbe et nombre du nombre de paroisses est incohérent entre chapitre TMD (154) et résultats JIAP (240), + variabilité très forte dans JIAP et pas dans TMD.

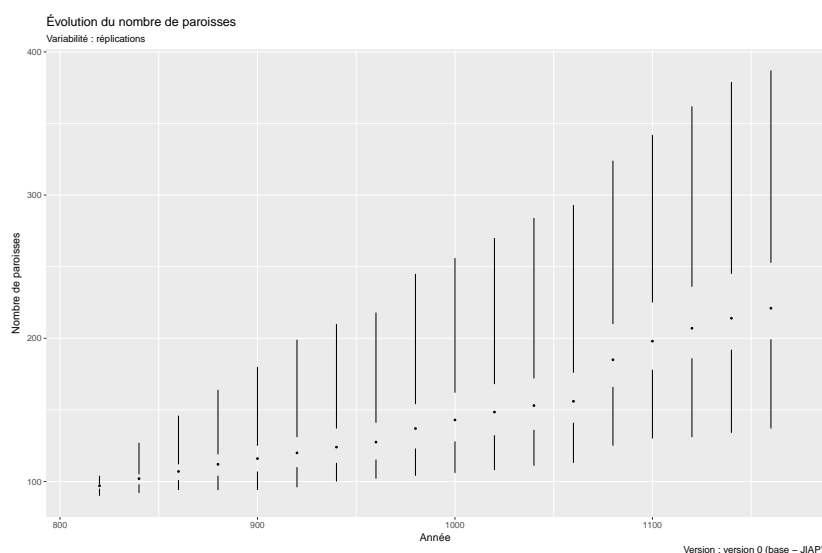


FIGURE 3.19 – Évolution du nombre de paroisses

37. La trop forte hétérogénéité dans les superficies des paroisses rend difficile la lecture de l'espace. On pourrait y remédier par exemple avec des cartogrammes, mais alors, on ne verrait plus que ces zones de forte densité.

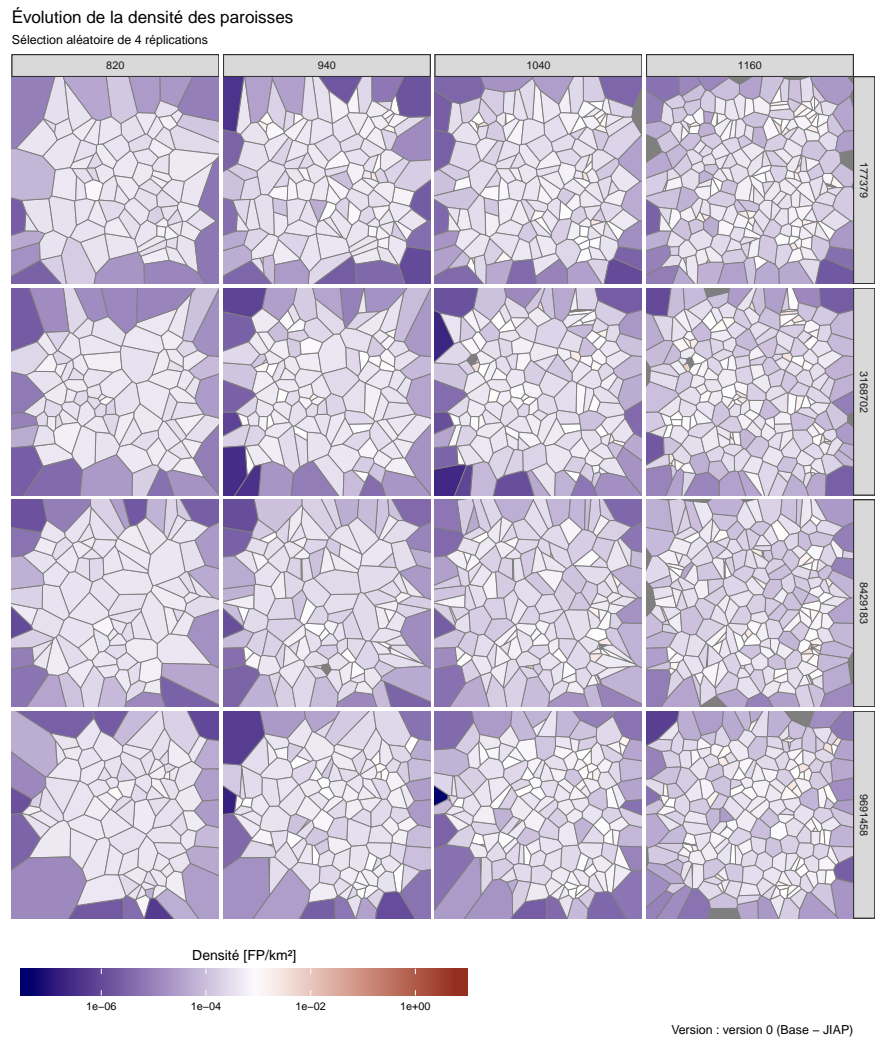


FIGURE 3.20 – Densité des paroisses

### 3.3.3.4 Efficacité et équité de la desserte paroissiale

On a vu précédemment (3.3.2.3 , Hiérarchie des paroisses, p. 47) que les paroisses devaient assurer conjointement une desserte efficace et équitable. Un indicateur possible pour évaluer l'efficacité du maillage simulé est d'en observer la couverture spatiale. Par définition, les paroisses couvrent ainsi l'ensemble du territoire simulé. Pour autant, certaines paroisses peuvent être très étendues, comme vu dans l'indicateur précédent. On peut toutefois quantifier la dispersion de ces paroisses en analysant leur répartition spatiale, ou plutôt, celle des églises paroissiales qui en constituent le cœur. Pour cela, on peut faire appel à une méthode d'analyse spatiale assez classique, similaire à la méthode des quadrats, en découpant l'espace en un maillage orthogonal régulier et en comptant le nombre d'églises paroissiales de chacune des mailles. On a vu que dans SimFeodal, les églises paroissiales sont réparties de manière à la fois dispersée (zones de faible densité) et concentrées (zones denses à proximité des agrégats importants). Plutôt que de faire un décompte des églises paroissiales et d'en tirer une hiérarchie, un indicateur plus simple est alors de faire un compte des mailles ne contenant pas de paroisse. Ces mailles étant régulières, l'indicateur résultant permettra alors d'appréhender simplement la part de l'espace couvert par des églises paroissiales. On a choisi ici de discrétiser l'espace modélisé (de dimensions  $100 \times 100$  km) en mailles de  $10 \times 10$  km, pour un total de 100 mailles de  $100\text{km}^2$ . Ces valeurs permettent d'approcher, à grands traits, la configuration théorique qui verrait chacune des 300 églises paroissiales être réparties dans une maille différente<sup>38</sup>.

Selon les mécanismes de SimFeodal, on s'attend à ce que l'indicateur ainsi produit diminue au cours du temps, à mesure que de nouvelles églises paroissiales viennent desservir le territoire, de manière régulière (comme l'évolution du nombre d'églises paroissiales). Cette diminution devrait toutefois être moins rapide que celle du nombre de paroisses, puisque les églises paroissiales créées au sein des agrégats devraient logiquement ne pas desservir de nouvelles mailles, mais plutôt se situer dans des mailles déjà occupées.

Du point de vue de l'équité de la desserte, on peut recourir à un indicateur plus partiel : il s'agit ici de décrire non pas l'évolution de la desserte des paroisses pour la plupart des foyers paysans, mais plutôt pour ceux d'entre eux qui seraient le plus isolés. Ainsi, on peut observer l'évolution de la proximité entre les foyers paysans et les églises paroissiales, et ce, pour les paysans les plus éloignés des dites églises paroissiales à chaque pas de temps. Une augmentation de l'équité devrait ainsi, par exemple pour les 10% de foyers paysans les moins bien desservis, montrer une diminution de la distance moyenne aux églises paroissiales.

On s'attend ici aussi à voir une augmentation de l'équité, et donc une diminution au cours du temps de cette distance. Empiriquement, on estime à 5km la distance maximale à laquelle un foyer paysan se situait d'une église

---

38. Il faudrait alors 300 mailles de  $\approx 6$  km de largeur, mais cette valeur ne constitue pas un multiple de 100. De plus, les églises paroissiales étant générées en partie au sein d'agréats existants pour renforcer la desserte du plus grand nombre, cette valeur est très loin d'être accessible.

paroissiale en fin de période (**Voir les seuils de distance dans chapitre 2**). Un paramétrage de SimFeodal sera donc considérée comme satisfaisant, sur ce plan de l'équité de desserte, s'il voit cette distance converger vers le seuil, voire passer en dessous, de 5km.

Il manque un indicateur d'équité. A compléter une fois qu'il sera mis en place. Par exemple :

Évolution de la distance des 10% de FP les plus éloignés d'une/à une église paroissiale ? == > Sans doute très intéressant, mais pas « productible » pour l'instant. De toute façon, à voir les résultats JIAP, il faudra sans doute faire retourner toutes les itérations du modèle avec des sorties semblables à celles de la v8 (4\_4\_A)

== > Peut être sous forme de boxplot ou d'histogramme (ridgeplot) ou de moyenne (plus influencée par grandes distances que médiane, donc mieux ici)

En matière d'efficacité, la version 0 de SimFeodal est assez peu satisfaisante : malgré un nombre d'églises paroissiales trop élevé (voir l'indicateur précédent), la couverture du territoire est assez faible. En effet, la figure 3.21 montre qu'en fin de simulation, près de 90% de mailles restent inoccupées par des églises paroissiales. La tendance est toutefois à la diminution de ces mailles vides, et cette diminution est assez stable et régulière dans le temps.

Comme dans le cas de l'évolution du nombre de paroisses, la variabilité est ici préoccupante et très insatisfaisante : l'écart interquartile est modéré (environ 15%), mais les minimums et maximums présentent un écart énorme (de 95% à 55%).

Pas possible d'analyser les résultats de la version 0 en matière d'équité sans les mesures correspondant. A faire une fois que les graphiques auront été produits.

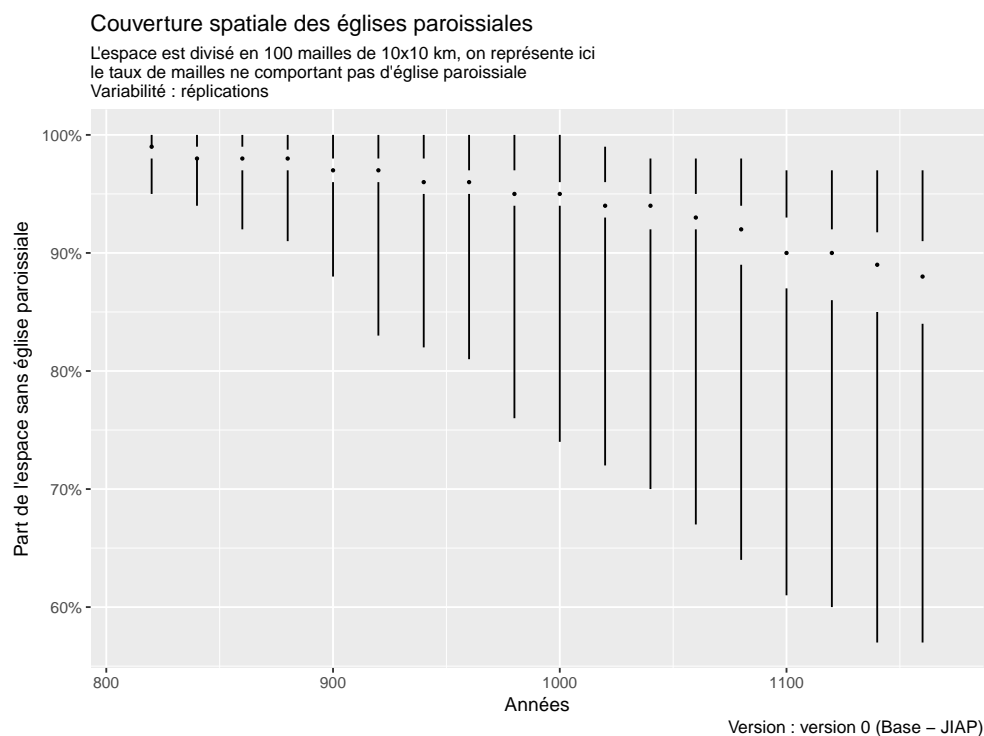


FIGURE 3.21 – Évolution de la couverture du territoire par les églises paroissiales.

Attention, résultats JIAP = pas fiables

## Références

- AMBLARD, Frédéric, Juliette ROUCHIER et Pierre BOMMEL (2006). « Evaluation et Validation de Modèles Multi-Agents ». In : *Modélisation et simulation multi-agents. Applications pour les Sciences de l'Homme et de la Société*, Hermès, p. 103-140.
- AUGUSIAK, Jacqueline, Paul J. VAN DEN BRINK et Volker GRIMM (2014). « Merging Validation and Evaluation of Ecological Models to 'Evaluation' : A Review of Terminology and a Practical Approach ». In : *Ecological Modelling. Population Models for Ecological Risk Assessment of Chemicals* 280, p. 117-128. DOI : 10/f52827.
- BALCI, Osman (1994). « Validation, Verification, and Testing Techniques throughout the Life Cycle of a Simulation Study ». In : *Annals of operations research* 53.1, p. 121-173. DOI : 10/ctgxcd.
- (1997). « Verification, Validation and Accreditation of Simulation Models ». en. In : *Proceedings of the 1997 WSC*. Atlanta, Georgia, p. 135-141.
- (1998). « Verification, Validation, and Accreditation ». In : *Proceedings of the 1998 WSC*. T. 1. IEEE, p. 41-48.
- BANOS, Arnaud (2013). « Pour des pratiques de modélisation et de simulation libérées en Géographie et SHS ». fr. Habilitation à Diriger des Recherches. Paris : Université Paris 1 - Panthéon-Sorbonne.
- CHÉREL, Guillaume, Clémentine COTTINEAU et Romain REUILLON (2015). « Beyond Corroboration : Strengthening Model Validation by Looking for Unexpected Patterns ». en. In : *PLOS ONE* 10.9, e0138212. DOI : 10/gddwr3.
- DE RUNZ, Cyril (2008). « Imperfection, Temps et Espace : Modélisation, Analyse et Visualisation Dans Un SIG Archéologique ». Thèse de doct. Université de Reims-Champagne Ardenne.
- EILERS, Dennis et al. (2017). « It's Not a Bug, It's a Feature : How Visual Model Evaluation Can Help to Incorporate Human Domain Knowledge in Data Science ». In :
- GRAUWIN, Sébastien, Florence GOFFETTE-NAGOT et Pablo JENSEN (2012). « Dynamic Models of Residential Segregation : An Analytical Solution ». In : *Journal of Public Economics* 96.1, p. 124-141. DOI : 10/fhrx73.
- HARMELIN-VIVIEN, M. L. et al. (1985). « Evaluation Visuelle Des Peuplements et Populations de Poissons Méthodes et Problèmes ». In :

- HERMANN, Charles (1967). « Validation Problems in Games and Simulations with Special Reference to Models of International Politics ». English. In : *Behavioral Science* 12.3. DOI : 10/bgxt27.
- HORST, G. L., M. C. ENGELKE et W. MEYERS (1984). « Assessment of Visual Evaluation Techniques 1 ». en. In : *Agronomy Journal* 76.4, p. 619-622. DOI : 10/br3zkb.
- KENNEDY, Ryan C. et al. (2006). « Verification and Validation of Agent-Based and Equation-Based Simulations : A Comparison ». In : *Simulation series* 38.2, p. 95.
- KLÜGL, Franziska (2008). « A Validation Methodology for Agent-Based Simulations ». In : *Proceedings of the 2008 ACM Symposium on Applied Computing*. SAC '08. New York, NY, USA : ACM, p. 39-43. DOI : 10/dpkh42.
- LAKE, Mark W. (2014). « Trends in Archaeological Simulation ». en. In : *Journal of Archaeological Method and Theory* 21.2, p. 258-287. DOI : 10/gd6b57.
- LIVET, Pierre, Denis PHAN et Lena SANDERS (2014). « Diversité et Complémentarité Des Modèles Multi-Agents En Sciences Sociales ». In : *Revue française de sociologie* 55.4, p. 689-729. DOI : 10/gfw9pr.
- MOSIER, Charles I. (1947). « A Critical Examination of the Concepts of Face Validity ». In : *Educational and Psychological Measurement* 7.2, p. 191-205. DOI : 10/ck5pmh.
- NAYLOR, Thomas H. et J. M. FINGER (1967). « Verification of Computer Simulation Models ». In : *Management Science* 14.2, B92-B106. DOI : 10/fmcsv4.
- NEVO, Baruch (1985). « Face Validity Revisited ». en. In : *Journal of Educational Measurement* 22.4, p. 287-293. DOI : 10/btf77x.
- NGO, The An et Linda SEE (2012). « Calibration and Validation of Agent-Based Models of Land Cover Change ». en. In : *Agent-Based Models of Geographical Systems*. Sous la dir. d'Alison J. HEPPENSTALL et al. Dordrecht : Springer Netherlands, p. 181-197. DOI : 10.1007/978-90-481-8927-4\_10.
- OURIACHI, Marie-Jeanne et al. (2018). « De l'élaboration de La Transition à Sa Modélisation : Retours Interdisciplinaires Sur l'expérience TransMonDyn ». In : *Peupler La Terre : De La Préhistoire à l'ère Des Métropoles*. Sous la dir. de Lena SANDERS. Perspectives Villes et Territoires. Tours : Presses universitaires François-Rabelais, p. 461-485.
- PETTY, Mikel D (2010). « Verification, Validation, and Accreditation ». In : *Modeling and simulation fundamentals : Theoretical underpinnings and practical domains*, p. 325-372. DOI : 10/bvhqd6.
- REUILLON, Romain, Mathieu LECLAIRE et Sébastien REY-COYREHOURCQ (2013). « OpenMOLE, a Workflow Engine Specifically Tailored for the Distributed Exploration of Simulation Models ». In : *Future Generation Computer Systems* 29.8, p. 1981-1990. DOI : 10.1016/j.future.2013.05.003.
- REY-COYREHOURCQ, Sébastien (2015). « Une plateforme intégrée pour la construction et l'évaluation de modèles de simulation en géographie ». Français. Thèse de doctorat en Géographie. Paris : Université Paris I - Panthéon-Sorbonne.
- SARGENT, Robert G. (1979). « Validation of Simulation Models ». In : *Proceedings of the 11th WSC - Volume 2*. WSC '79. Piscataway, NJ, USA : IEEE Press, p. 497-503.

- SARGENT, Robert G. (2009). « Verification and Validation of Simulation Models ». In : *Proceedings of the 2009 WSC*. IEEE, p. 162-176.
- SARGENT, Robert G. et O. BALCI (2017). « History of Verification and Validation of Simulation Models ». In : *Proceedings of the 2017 WSC*, p. 292-307. DOI : 10/gfw9pf.
- SHAW, William (2017). « Elevated Urinary Glyphosate and Clostridia Metabolites With Altered Dopamine Metabolism in Triplets With Autistic Spectrum Disorder or Suspected Seizure Disorder : A Case Study ». In : *Integrative Medicine : A Clinician's Journal* 16.1, p. 50-57.
- SIMON, Herbert A (1955). « On a Class of Skew Distribution Functions ». In : *Biometrika* 42.3/4, p. 425-440. DOI : 10/dm868s.
- THOMAS, Steven (2012). *Revisiting the Iterative Incremental Mona Lisa*.
- YULE, G Udny (1925). « II.—A Mathematical Theory of Evolution, Based on the Conclusions of Dr. JC Willis, FR S ». In : *Phil. Trans. R. Soc. Lond. B* 213.402-410, p. 21-87. DOI : 10.1098/rstb.1925.0002.
- ZHANG, Junfu (2011). « Tipping and Residential Segregation : A Unified Schelling Model ». en. In : *Journal of Regional Science* 51.1, p. 167-193. DOI : 10/fqc4mz.